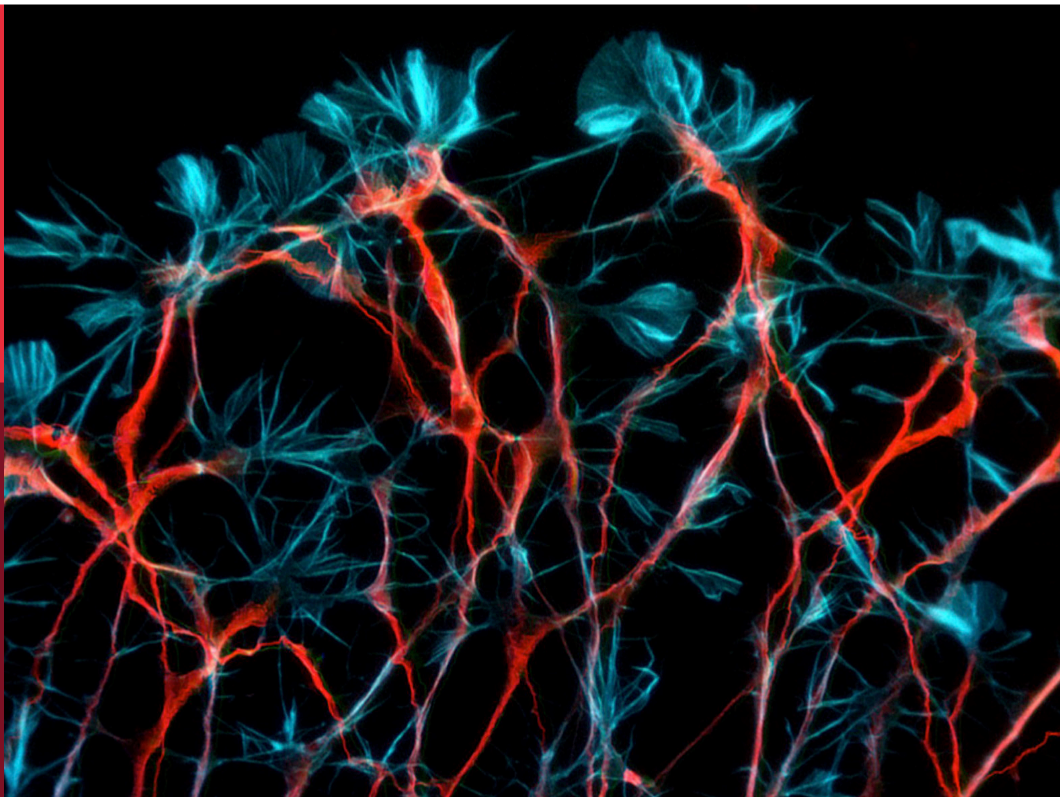
Research Topics

Neuro-detection: advancements in pattern detection and segmentation techniques in neuroscience

Edited by
Najib Ben Aoun, Ziliang Ren, Sadique Ahmad
and Ridha Ejbali

Published in
Frontiers in Computational Neuroscience
Frontiers in Neuroinformatics



ISSN 1664-8714
ISBN 978-2-8325-6894-1
DOI 10.3389/978-2-8325-6894-1

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Neuro-detection: advancements in pattern detection and segmentation techniques in neuroscience

Topic editors
Najib Ben Aoun — Al Baha University, Saudi Arabia
Ziliang Ren — Dongguan University of Technology, China
Sadique Ahmad — Prince Sultan University, Saudi Arabia
Ridha Ejbali — Gabes University, Tunisia

Citation
Ben Aoun, N., Ren, Z., Ahmad, S., Ejbali, R., eds. (2025). *Neuro-detection: advancements in pattern detection and segmentation techniques in neuroscience*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6894-1

Table of contents

Frontiers in Computational Neuroscience

TYPE Editorial
PUBLISHED 02 September 2025
DOI 10.3389/fncom.2025.1685174

OPEN ACCESS

EDITED AND REVIEWED BY
Si Wu,
Peking University, China

*CORRESPONDENCE
Najib Ben Aoun
✉ najib.benaoun@ieee.org

RECEIVED 13 August 2025
ACCEPTED 22 August 2025
PUBLISHED 02 September 2025

CITATION
Ben Aoun N, Ahmad S and Ejbali R (2025) Editorial: Neuro-detection: advancements in pattern detection and segmentation techniques in neuroscience.
Front. Comput. Neurosci. 19:1685174.
doi: 10.3389/fncom.2025.1685174

Editorial: Neuro-detection: advancements in pattern detection and segmentation techniques in neuroscience

Najib Ben Aoun[1]*, Sadique Ahmad[2] and Ridha Ejbali[3]

[1]Faculty of Computing and Information, Al-Baha University, Al Baha, Saudi Arabia, [2]EIAS: Data Science and Blockchain Laboratory, College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia, [3]Research Team in Intelligent Machines (RTIM), Engineering School of Gabès (ENIG), University of Gabès, Gabès, Tunisia

KEYWORDS

neuroscience, artificial intelligence, image segmentation–deep learning, brain tumors classification, neuroimaging, medical imaging

Editorial on the Research Topic
Neuro-detection: advancements in pattern detection and segmentation techniques in neuroscience

In recent years, the intersection of artificial intelligence (AI), deep learning, and neuroimaging has created transformative opportunities in the early diagnosis, classification, and treatment planning of neurological disorders and brain tumors. This Research Topic, *Neuro-detection: Advancements in Pattern Detection and Segmentation Techniques in Neuroscience*, presents a collection of nine high-quality studies that illustrate the growing impact of machine learning-driven methods in advancing medical diagnostics and understanding complex neurobiological conditions. The collection explores how AI techniques, particularly deep neural networks, explainable AI (XAI), unsupervised learning, and ensemble models, can be effectively applied to tasks such as interpreting complex medical imaging data, uncovering subtle biomarkers, and supporting clinical decision-making. Each contribution not only advances methodological concepts but also demonstrates the applicability and generalizability of these techniques in real-world diagnostic scenarios.

Brain tumor classification and segmentation remains a particularly challenging problem due to tumor heterogeneity, imaging variability, and anatomical differences across patients. Several contributions address these challenges with state-of-the-art approaches. One study proposes an ensemble model combining Vision Transformers (ViT) and EfficientNet-V2, optimized via a genetic algorithm-based weighted strategy. This hybrid system captures both local and global MRI features, outperforming individual models with 95% classification accuracy. In this contribution, Gasmi et al., demonstrate that such synergistic model integration can address multi-class medical classification problems with potential applicability beyond neuro-oncology.

Kiran et al., introduce a binary convolutional neural network (BCNN) for segmenting the ten most common brain tumor types, supported by a new dataset of 6,600 MRI images. Using adaptive thresholding and advanced morphological operations, the BCNN achieved 99.40% accuracy and a 99.28% F1-score, excelling at differentiating tumor grades and types. Another contribution, by Albalawi et al., presents a multi-task CNN architecture capable of simultaneously detecting tumors, classifying them by type and grade, and localizing them.

Trained on over 7,000 MRI images across four categories, the model achieved 99% overall accuracy, making it a promising candidate for integration into clinical workflows.

Beyond supervised segmentation, Arora et al., propose an unsupervised "unruly clustering" method based on integrated intuitionistic fuzzy logic with conditional spatial properties. By incorporating hesitation degrees and conditional spatial functions, the approach adapts to local image context, improving segmentation robustness in noisy or low-contrast conditions. Similarly, Iqbal et al., bridge handcrafted statistical radiomic features with deep spatial representations from a ResNet-inspired architecture. Their custom Fusion Net preserves critical information from both domains, reaching 97.53% accuracy and 97.77% precision on the Brains dataset, underscoring the benefits of combining engineered and learned features for complex classification tasks.

Moreover, AI applications in neurodegenerative disease diagnosis extend these innovations beyond oncology to conditions like Alzheimer's disease (AD) and Parkinson's disease (PD), where early detection can substantially influence patient outcomes. In AD, graph-based modeling has emerged as a powerful paradigm for capturing the complex interplay between brain regions. Alharbi et al., introduce a spectral graph Convolutional Neural Network (SGCNN) that represents MRI-derived brain connectivity as a graph, enabling the model to learn topological and spectral patterns linked to disease progression. Through targeted ablation experiments, they improved classification performance to 95%, outperforming conventional CNNs and confirming that graph-based architectures are well suited for mapping the connectivity disruptions and structural atrophy patterns characteristic of AD.

Building on this connectivity focus, Biswas and Sripada shift from *associative* to *causal* brain network analysis, introducing Causal Functional Connectivity (CFC) as a richer and potentially more prognostic biomarker. Using resting-state fMRI data from the Alzheimer's Disease Neuroimaging Initiative, they applied the Time-aware PC (TPC) algorithm—a directed graphical modeling method tailored for time series—to compute whole-brain causal connectomes for cognitively normal, mild cognitive impairment, and AD groups. Compared to Granger Causality and Sparse Partial Correlation approaches, TPC produced sparse, interpretable, and directionally meaningful maps that aligned with known connectivity patterns while revealing contemporaneous and directional influences often missed by standard methods. Edge-wise statistical analysis highlighted reduced causal influence from Heschl's gyrus, thalamus, and posterior cingulate cortex, alongside increased self-connections in the parahippocampal gyrus—alterations consistent with more than 30 prior neuroimaging studies. These findings capture both the loss of key network hubs and the emergence of compensatory hyperconnectivity during early disease stages, positioning TPC-based CFC analysis as a valuable complement to graph-based classifiers like SGCNN in early AD detection.

While AD studies highlight the role of connectivity modeling, PD research in this Research Topic demonstrates the potential of alternative, non-imaging biomarkers. Kim et al., present a voice-based diagnostic framework using self-supervised deep representation pattern learning (SS-DRPL), which learns micro-temporal and frequency features from unlabelled voice recordings. Combined with LSTM-RNN and DNN architectures, the model achieved an F1-score of 0.94, showing the promise of accessible, non-invasive diagnostic tools.

Another PD-focused study by Alharthi applies XAI techniques, particularly layer-wise relevance propagation (LRP), to neural network models trained on gait sensor data from Parkinson's and cognitively impaired individuals. The approach not only achieved a 98% F1-score on PD datasets and up to 90% ± 10% on healthy individuals under dual-task conditions, but also provided interpretable insights into how specific gait features influence predictions. Such transparency strengthens clinician trust and enables better understanding of how neurological deterioration manifests in motor patterns.

Together, these contributions present a coherent picture of progress at the intersection of deep learning, neuroimaging, and medical diagnostics. Across both tumor detection and neurodegenerative disease diagnosis, common themes emerge: integrating multi-modal and hybrid feature sets strengthens model performance; graph-based and causal network modeling open new frontiers for understanding disease mechanisms; and XAI techniques enhance transparency, paving the way for clinical adoption. The diversity of datasets, modalities, and analytic strategies showcased here demonstrates both the adaptability of AI methods to different neurological conditions and their potential to transform diagnostic practice in the years ahead.

Author contributions

NB: Formal analysis, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. SA: Formal analysis, Investigation, Methodology, Validation, Writing – review & editing. RE: Formal analysis, Investigation, Methodology, Validation, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

frontiers | Frontiers in Computational Neuroscience

TYPE Original Research
PUBLISHED 19 December 2023
DOI 10.3389/fncom.2023.1251301

OPEN ACCESS

EDITED BY
Junwei Duan,
Jinan University, China

REVIEWED BY
Junwei Jin,
Henan University of Technology, China
Lin Wang,
Chinese Academy of Sciences (CAS), China
Qiang Lin,
Sun Yat-sen University, China

*CORRESPONDENCE
Rahul Biswas
✉ rbiswas1@uw.edu

RECEIVED 01 July 2023
ACCEPTED 28 November 2023
PUBLISHED 19 December 2023

CITATION
Biswas R and Sripada S (2023) Causal functional connectivity in Alzheimer's disease computed from time series fMRI data.
Front. Comput. Neurosci. 17:1251301.
doi: 10.3389/fncom.2023.1251301

Causal functional connectivity in Alzheimer's disease computed from time series fMRI data

Rahul Biswas[1]* and SuryaNarayana Sripada[2]

[1]Department of Electrical and Computer Engineering, University of Washington, Seattle, WA, United States, [2]Center for Research on Science and Consciousness, Redmond, WA, United States

Functional connectivity between brain regions is known to be altered in Alzheimer's disease and promises to be a biomarker for early diagnosis. Several approaches for functional connectivity obtain an un-directed network representing stochastic associations (correlations) between brain regions. However, association does not necessarily imply causation. In contrast, Causal Functional Connectivity (CFC) is more informative, providing a directed network representing causal relationships between brain regions. In this paper, we obtained the causal functional connectome for the whole brain from resting-state functional magnetic resonance imaging (rs-fMRI) recordings of subjects from three clinical groups: cognitively normal, mild cognitive impairment, and Alzheimer's disease. We applied the recently developed Time-aware PC (TPC) algorithm to infer the causal functional connectome for the whole brain. TPC supports model-free estimation of whole brain CFC based on directed graphical modeling in a time series setting. We compared the CFC outcome of TPC with that of other related approaches in the literature. Then, we used the CFC outcomes of TPC and performed an exploratory analysis of the difference in strengths of CFC edges between Alzheimer's and cognitively normal groups, based on edge-wise p-values obtained by Welch's t-test. The brain regions thus identified are found to be in agreement with literature on brain regions impacted by Alzheimer's disease, published by researchers from clinical/medical institutions.

KEYWORDS

causal inference, functional connectivity, brain mapping, directed graphical modeling, Alzheimer's disease, functional magnetic resonance imaging

1 Introduction

Alzheimer's disease (AD) is the most common age-related progressive neurodegenerative disorder. It typically begins with a preclinical phase and advances through mild cognitive impairment (MCI) to clinically significant AD, a form of dementia (Querfurth and LaFerla, 2010). Despite substantial efforts to identify biomarkers for AD, it still relies on clinical diagnosis, and early and accurate disease prediction remains limited (Laske et al., 2015; Li et al., 2019). Abnormal resting-state functional connectivity (FC) between brain regions has been observed as early as two decades before brain atrophy and the emergence of AD symptoms (Ashraf et al., 2015; Nakamura et al., 2017). Therefore, resting-state FC can potentially determine the relative risk of developing AD (Sheline and Raichle, 2013; Brier et al., 2014).

Resting-state functional magnetic resonance imaging (rs-fMRI) records the blood-oxygen-level-dependent (BOLD) signals from different brain regions while individuals are awake and not engaged in any specific task. The BOLD signal is popularly used to infer FC between brain regions partly due to the advantage that BOLD signal provides high spatial resolution (Yamasaki et al., 2012; Sporns, 2013; Liu et al., 2015; Xue et al., 2019).

FC refers to the stochastic relationship between brain regions with respect to their activity over time. Popularly, FC involves measuring the statistical association between signals from different brain regions. The statistical association measures are either pairwise associations between pairs of brain regions, such as Pearson's correlation, or multivariate i.e., incorporating multi-regional interactions such as undirected graphical models (Biswas and Shlizerman, 2022a). Detailed technical explanations of FC in fMRI can be found in Chen et al. (2017), Keilholz et al. (2017), and Scarapicchia et al. (2018). The findings from studies using FC (Wang et al., 2007; Kim et al., 2016), and meta-analyses (Jacobs et al., 2013; Li et al., 2015; Badhwar et al., 2017) indicate a decrease in connectivity in several brain regions with AD, such as the posterior cingulate cortex and hippocampus. These regions play a role in attentional processing and memory. On the other hand, some studies have found an increase in connectivity within brain regions in the early stages of AD and MCI (Gour et al., 2014; Bozzali et al., 2015; Hillary and Grafman, 2017). Such an increase in connectivity is a well known phenomenon that occurs when the communication between other brain regions is impaired. Such hyperconnectivity has been interpreted as a compensatory mechanism where alternative paths within the brain's network are recruited (Hillary and Grafman, 2017; Oldham and Fornito, 2019; Marek and Dosenbach, 2022).

In contrast to Associative FC (AFC), Causal FC (CFC) represents functional connectivity between brain regions more informatively by a directed graph, with nodes as the brain regions, directed edges between nodes indicating causal relationships between the brain regions, and weights of the directed edges quantifying the strength of the corresponding causal relationship (Spirtes et al., 2000). However, functional connectomics studies in general, and those concerning fMRI from AD in particular, have predominantly used associative measures of FC (Reid et al., 2019). There are a few studies that deal with comparing broad hypotheses of alteration within the CFC in AD (Rytsar et al., 2011; Khatri et al., 2021). However, this area is largely unexplored, partly due to the lack of methods that can infer CFC in a desirable manner, as explained next.

Several properties are desirable in the context of causal modeling of FC (Smith et al., 2011; Biswas and Shlizerman, 2022a). Specifically, the CFC should represent causality while free of limiting assumptions such as linearity of interactions. In addition, since the activity of brain regions are related over time, such temporal relationships should be incorporated in defining causal relationships in neural activity. The estimation of CFC should be computationally feasible for the whole brain FC instead of limiting it to a smaller brain network. It is also desirable to capture beyond-pairwise multi-regional cause-and-effect interactions between brain regions. Furthermore, since the BOLD signal occurs and is sampled at a temporal resolution that is far slower than the neuronal activity, thereby causal effects often appear as contemporaneous (Granger, 1969; Smith et al., 2011). Therefore, the causal model in fMRI data should support contemporaneous interactions between brain regions.

Among the methods for finding CFC, *Dynamic Causal Model* (DCM) requires a mechanistic biological model and compares different model hypotheses based on evidence from data, and is unsuitable for estimating the CFC of the whole brain (Friston et al., 2003; Smith et al., 2011). On the other hand, Granger Causality (GC) typically assumes a vector auto-regressive linear model for the activity of brain regions over time, and it tells whether a regions's past is predictive of another's future (Granger, 2001). Furthermore, GC does not include contemporaneous interactions. This is a drawback since fMRI data often consists of contemporaneous interactions (Smith et al., 2011). In contrast, *Directed Graphical Modeling* (DGM) has the advantage that it does not require the specification of a parametric equation of the neural activity over time, it is predictive of the consequence of interventions, and supports estimation of whole brain CFC. Furthermore, the approach inherently goes beyond pairwise interactions to include multi-regional interactions between brain regions and estimating the cause and effect of such interactions. The *Time-aware PC* (TPC) algorithm is a recent method for computing the CFC based on DGM in a time series setting (Biswas and Shlizerman, 2022b). In addition, TPC also accommodates contemporaneous interactions among brain regions. A detailed comparative analysis of approaches to find CFC is provided in Biswas and Shlizerman (2022a,b). With the development of methodologies such as TPC, it would be possible to infer the whole brain CFC with the aforementioned desirable properties.

In this paper, we apply the TPC algorithm to infer the CFC between brain regions from resting-state fMRI data. The TPC algorithm estimates the subject-specific CFC for each subject from their fMRI data. We compare the CFC outcome of TPC with GC and Sparse Partial Correlation (SPC), which are approaches to find the CFC and AFC, respectively. We then use the CFC outcome of TPC to investigate the alteration of CFC in AD. In this regard, we conducted an exploratory analysis for the difference in strength of causal connections in AD compared to CN subjects (and MCI compared to CN subjects), based on their edge-wise p-values given by Welch's t-test. We reported the resulting CFC edges with lowest edge-wise p-values for altered connectivity in AD compared to CN subjects and their corresponding brain regions. The brain regions identified in those analyses are consistent with published literature on regions impacted by AD, with each such publication being a report from a team involving a clinical setting and at least one medical expert, thereby validating the approach.

2 Materials and methods

2.1 Participants

The resting fMRI and demographic data were downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI; http://adni.loni.usc.edu/). A total of 129 subjects were included in the study: 41 subjects who are CN, 54 subjects with MCI, and 34 subjects with AD.

Table 1 includes a summary of the participants' demographic and medical information. In the experiments, the subjects with AD presented significantly lower scores in the screening assessment cognitive test Mini-Mental State Examination (MMSE)

TABLE 1 **Summary of demographic information and Mini Mental State Examination (MMSE) for CN, MCI and AD subjects.**

Characteristic	CN	MCI	AD	p
Number of subjects	41	54	34	–
Sex (M/F)	19/22	29/26	16/18	0.16
Age (years)	74.9 ± 6.4	74.2 ± 7.1	74.4 ± 7.4	0.86
Education (years)	16.5 ± 2.3	15.7 ± 2.6	15.4 ± 2.5	0.22
MMSE	29.1 ± 1.4	27.8 ± 1.9	21.9 ± 4.2	$<10^{-14}$

The second to fourth columns present group characteristics, mean ± SD. The fifth column presents p-values for the statistical significance of the inter-group differences. Differences in Sex was assessed using a Chi-Squared test and differences in Age, Education and MMSE using non-parametric analysis of variance by Kruskal–Wallis test.

in comparison with the other groups. The subjects were age-matched (Kruskal–Wallis test: $p > 0.8$), gender-matched (Chi-Squared test: $p > 0.1$), and matching number of years of education (Kruskal–Wallis test: $p > 0.2$). As expected, MMSE scores had a significant difference between all pairs of groups (Kruskal–Wallis test: $p < 10^{-14}$).

2.2 Image acquisition

The acquisition of fMRI images was performed using Philips Medical Systems scanner. The fMRI images were obtained using an echo planar imaging sequence at a field strength of 3.0 Tesla, with a repetition time (TR) of 3 s, an echo time (TE) of 30 ms, and a flip angle of 80 degrees. The matrix size was 64 × 64 pixels, 140 volumes, 48 slices per volume, slice thickness of 3.3 mm, and voxel size of $3.3 \times 3.3 \times 3.3$ mm^3.

2.3 fMRI preprocessing

The fMRI pre-processing steps were carried out using the CONN toolbox version 21a, which utilizes the Statistical Parametric Mapping (SPM12), both of which are MATLAB-based cross-platform software (Friston et al., 1994; Nieto-Castanon and Whitfield-Gabrieli, 2021). We used the default pre-processing pipeline in CONN, consisting of the following steps in order: functional realignment and unwarp (subject motion estimation and correction), functional centering to (0, 0, 0) coordinates (translation), slice-time correction with interleaved slice order, outlier identification using Artifact Detection and Removal Tool, segmentation into gray matter, white matter and cerebrospinal fluid tissue, and direct normalization into standard Montreal Neurological Institute (MNI) brain space, and lastly, smoothing using spatial convolution with a Gaussian kernel of 8 mm full-width half maximum. This pipeline was followed by detrending and bandpass filtering (0.001–0.1 Hz) to remove low-frequency scanner drift and physiological noise in the fMRI images. The first four time points have been filtered out to remove any artifacts.

For the extraction of Regions-Of-Interest (ROIs), the automated anatomical labeling (AAL) atlas was utilized on the pre-processed rs-fMRI dataset (Tzourio-Mazoyer et al., 2002). The list of all regions in the AAL atlas is provided in Supplementary material along with their abbreviated, short, and full region names. This parcellation method has been demonstrated to be optimal for studying the FC between brain regions (Arslan et al., 2018). The voxels within each ROI were averaged, resulting in a time series for each ROI.

2.4 Inference of causal functional connectivity: Time-aware PC algorithm

The TPC Algorithm finds CFC between brain regions from time series based on DGM (Spirtes et al., 2000; Pearl, 2009; Biswas and Mukherjee, 2022; Biswas and Shlizerman, 2022a,b). While traditional DGM applies to static data, TPC extends the applicability of DGM to CFC inference in time series by first implementing the Directed Markov Property to model causal spatial and temporal interactions in the time series by an unrolled Directed Acyclic Graph (DAG) of the time series. The unrolled DAG consists of nodes (v, t), for region of interest v and time t, and edge $(v_1, t_1) \rightarrow (v_2, t_2)$ reflecting causal interaction from the BOLD signal in region v_1 at time t_1 to the BOLD signal in region v_2 at time t_2. The estimation of the unrolled DAG is carried out by first transforming the time series into sequential variables with a maximum time delay of interaction τ and then applying the Peter-Clark (PC) algorithm to infer the unrolled DAG based on the sequential variables (Kalisch and Bühlman, 2007). TPC then rolls the DAG back to obtain the CFC graph between the regions of interest (see Figure 1) (Biswas and Shlizerman, 2022b). We consider $\tau = 1$ for our analyses, which would include interactions of the BOLD signal between regions of interest with a maximum time delay of 3 s, the TR of the fMRI acquisition. The Python package *TimeAwarePC* is used for implementation (Biswas and Shlizerman, 2022b).

The CFC outcome of this methodology is interpretable in the following manner: An edge from region $i \rightarrow j$ in the CFC estimate represents significant causal interaction from brain region i at preceding times to region j at following times. The model and the approach are non-parametric, meaning that it does not require the specification of a parametric dynamical equation for neural activity. The method captures beyond-pairwise multivariate interactions between brain regions. It also supports the estimation of the CFC for the whole brain in a computationally feasible manner. It also allows for time delays in interactions between the brain units and the presence of feedback loops. Furthermore, it has been shown that if the neural activity obeys an arbitrary dynamical process, the model outcome of TPC is consistent with respect to the causal relationships implied by the dynamical process and is predictive of counterfactual queries such as ablation or modulation (Biswas and Shlizerman, 2022b).

It is noteworthy that implementing the Directed Markov Property on the unrolled DAG to model causal relationships over time enables contemporaneous interactions e.g., from region u to region v at time t (Biswas and Shlizerman, 2022b). Such contemporaneous interactions are represented by the edge $(u, t) \rightarrow (v, t)$ in the unrolled DAG, and the presence of such an edge in the unrolled DAG would be reflected as an edge $u \rightarrow v$ in the Rolled

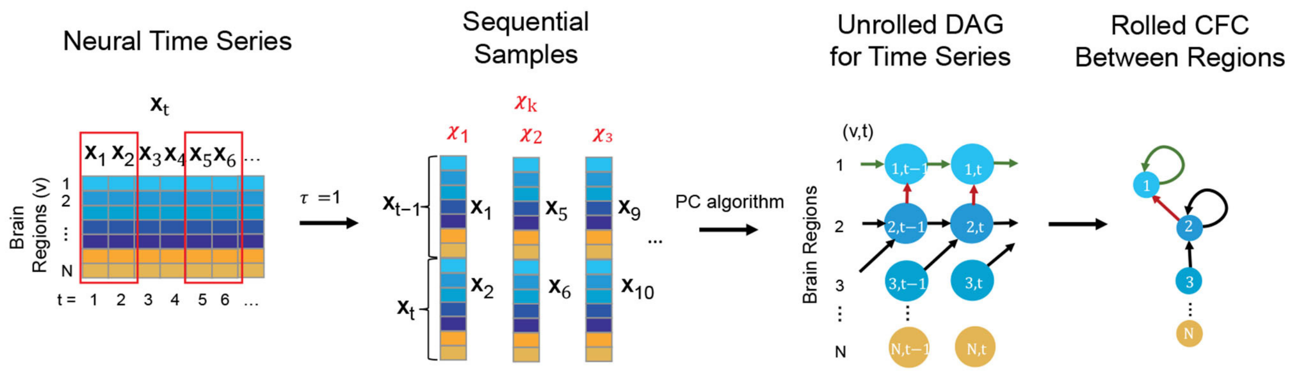


FIGURE 1
Steps conveying the concept of the TPC algorithm to infer CFC from observed neural time series data: First the neural time series is transformed to form sequential samples with a maximum time delay of interaction, τ (here $\tau = 1$). Then, Peter-Clark (PC) algorithm is applied on the sequential samples to obtain the unrolled DAG satisfying the Directed Markov Property. Finally the unrolled DAG is transformed to obtain the Rolled CFC between regions.

CFC outcome. Such contemporaneous interactions are especially relevant in fMRI due to the relatively slow temporal resolution of the BOLD signal compared to the underlying neural activity (Smith et al., 2011).

2.5 Comparison with functional connectivity using other approaches

In Biswas and Shlizerman (2022b), the authors have demonstrated that TPC performs better in computing CFC compared to other methods such as GC on simulated and public benchmarking datasets as well as on a real neurobiological dataset of single neuron signals obtained using Neuropixels. Additionally, the authors have drawn contrast [in Biswas and Shlizerman (2022b)] with SPC, which is a popular method for inferring AFC. In this paper, we computed AFC using SPC and CFC using GC from fMRI data (Deshpande et al., 2009; Schouten et al., 2016). We compared these two outputs with the CFC obtained by TPC from fMRI data. The GC graph is computed using the Nitime Python library, which fits a Multi-variate Auto-Regressive (MVAR) model followed by the use of GrangerAnalyzer to compute the GC (Rokem et al., 2009). We consider MVAR model of order 1, and GC likelihood ratio statistic of greater than 95 percentile as indicating edges (Schmidt et al., 2016). The SPC was estimated by Graphical Lasso penalized Maximum Likelihood Estimation, whose optimal penalization was obtained by a five-fold cross-validation (Friedman et al., 2008).

2.6 Alterations of CFC edges in Alzheimer's disease

We perform an exploratory analysis of statistical trends for edge-wise inter-group differences. Using the subject-specific CFC computed by TPC algorithm, for each detected CFC edge, we reported the p-value in the Welch's t-test for greater average edge weight in one clinical group compared to another clinical group (Yuen, 1974). Specifically, we listed the CFC edges with 10 lowest p-values for greater average weight in CN compared to AD group (and for greater average weight in AD compared to CN). For a CFC edge from region u to region v, we refer to u as the *source* brain region and v as the *destination* brain region. The source brain regions of the CFC edges with lowest p-values are found to be in agreement with literature for regions impacted by AD.

3 Results

3.1 Subject-specific causal functional connectivity

Figure 2 shows the CFC estimated using the TPC algorithm for an example subject (ID: 129_S_4396) in the CN group. In Figure 2A, the CFC is represented in the form of a matrix, whose entry (i, j) indicates the presence of connectivity from region index $i \rightarrow j$, and the value at entry (i, j) represents the weight of that causal connection. A positive value (blue) of the weight indicates excitatory influence, whereas a negative value (red) indicates inhibitory influence. The diagonal of the matrix representing self-connections for regions has been filtered out. In Figure 2B, the CFC is represented by a directed graph overlayed on schematics of the brain. The schematics of the brain comprise 2-dimensional brain projections in the Frontal, Axial, and Lateral planes. The nodes of the CFC graph correspond to the centers of brain regions in the AAL atlas. The nodes are colored light to dark gray according to their depth in the brain, with light gray representing superficial and dark gray representing deeper brain regions. The CFC graph provides a highly informative map of causal interactions between brain regions.

It is noteworthy that the CFC computed by TPC is sparse since the edges are filtered by conditional dependence tests. We quantified the sparsity of a CFC graph by its edge density. Edge density of a directed graph is the proportion of the number of edges in the directed graph over the total number of edges in the corresponding fully connected graph. Therefore, the edge density of an empty graph is 0 and that of a fully connected graph is 1. For

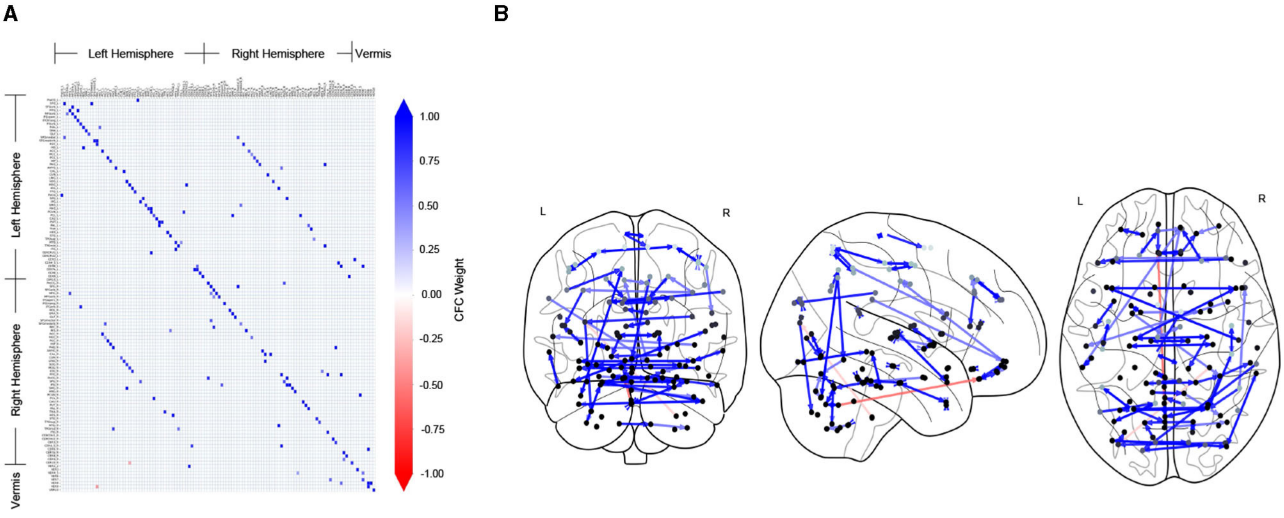


FIGURE 2
CFC for an example subject who is CN, estimated by TPC algorithm. **(A)** The estimated CFC is represented by its adjacency matrix, whose non-zero entry (i, j) represents the connection of region $i \rightarrow j$. **(B)** The CFC is visualized with directed graph edges on the Frontal, Axial and Lateral brain maps (left to right). The nodes correspond to brain region centers, ranging from superficial (light gray) to deeper (darker gray) regions, in the AAL brain atlas.

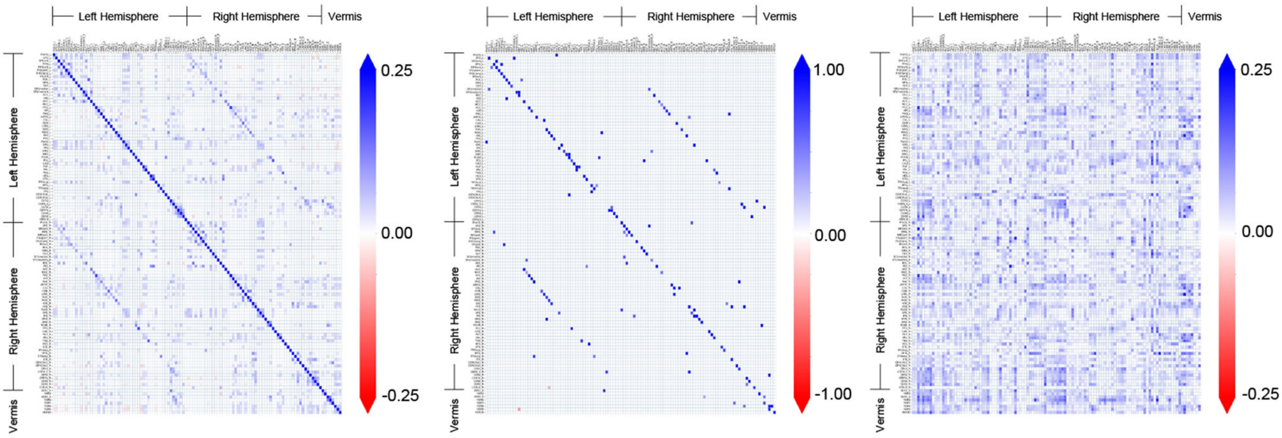


FIGURE 3
Comparison and demonstration of FC inferred by three methods: Associative FC using SPC, and Causal FC using TPC and GC. The estimated FC is represented by its adjacency matrix with edge weights, which is symmetric for Associative FC and asymmetric for Causal FC. In the adjacency matrices, a non-zero entry in (i, j) represents the connection of region $i \rightarrow j$.

the CFC graphs computed by TPC, the edge density for subjects in the CN group is (mean ± standard deviation) 0.0117 ± 0.0008, MCI group is 0.0118 ± 0.0009, and AD group is 0.0118 ± 0.0008, indicating a sparse CFC outcome of TPC for subjects in each of the groups.

3.2 Comparison with functional connectivity using other approaches

Figure 3 shows the adjacency matrices for the FC obtained by different methods for an example subject (ID: 129_S_4396) in the CN group. The AFC constitutes a distinct pattern of associative connectivity among the regions. It is expected that the CFC will be a directed subgraph of the AFC and be consistent with the overall patterns present in the AFC (Dadgostar et al., 2016; Wang et al., 2016). However, the patterns present in the CFC obtained by GC do not match with the AFC upon visual inspection. In comparison, the overall patterns present in the CFC obtained by TPC indeed match with the AFC obtained by SPC. On a detailed level, there are differences between TPC-CFC and AFC: TPC results in a directed graph thereby its adjacency matrix is asymmetric while AFC is an undirected graph with symmetric adjacency matrix. Furthermore, the CFC obtained by TPC includes self-loops represented by the diagonals of the adjacency matrix in contrast to GC, and results in a sparse matrix devoid of noise since the connections are filtered by conditional dependence tests.

3.3 Alterations of CFC edges in Alzheimer's disease

Figure 4 shows the edge-wise p-values for greater average edge weight in one clinical group compared to another, based on Welch's t-test. This provides insights into statistical trends for CFC edges that have an increase (or decrease) in strength in CN compared to MCI, CN compared to AD, and MCI compared to AD subjects.

In Table 2, we report 10 CFC edges that show the lowest p-values for greater average strength in CN subjects compared to AD, and their source brain regions. Similarly, we report another list of 10 CFC edges corresponding to greater average strength in subjects with AD compared to CN, and their source brain regions. The reported brain regions are in agreement with published medical literature cited in Table 2.

4 Discussion

In this study, we have obtained the CFC of the whole brain from its resting state fMRI time series. We used the recently developed TPC algorithm based on directed graphical modeling in time series, to compute the CFC. In the dataset, the subjects belonged to three clinical categories: CN, MCI, and AD. We computed the subject-specific CFC using TPC and compared it with those obtained by other approaches, such as GC. We then used the CFC outcomes of TPC for further investigation into the alteration of CFC in AD. In this regard, we explored statistical trends for edges that have a difference in strength between clinical categories, based on their edge-wise p-values obtained by Welch's t-test. We reported the causal connections with lowest p-values for greater strength in CN compared to AD (and greater strength in AD compared to CN) and their corresponding brain regions. The brain regions identified in the above analyses were found to be in agreement with medical literature for regions impacted by AD.

In Figure 4 and Table 2, the presence of CFC edges with weight in AD greater than that in CN (in addition to edges with weight in AD less than that in CN) is consistent with published studies in the literature. While several studies have concluded decreased connectivity in MCI and AD compared to CN (Jacobs et al., 2013; Li et al., 2015; Badhwar et al., 2017), others have highlighted that MCI and early stages of AD can involve an increase in FC between brain regions (Fredericks et al., 2018; Penalba-Sánchez et al., 2023). This increase occurs when the communication between specific brain regions is impaired and has been interpreted as a compensatory mechanism where alternative paths within the brain's network are recruited (Hillary and Grafman, 2017; Oldham and Fornito, 2019; Marek and Dosenbach, 2022). In the short term, the augmentation of FC along alternative pathways exhibits efficiency and adaptability of the brain. However, it is imperative to acknowledge the susceptibility of these densely interconnected hubs to beta-amyloid deposition, which can elicit secondary damage through metabolic stress, ultimately culminating in system breakdown (Hillary and Grafman, 2017). Consequently, the initial state of hyperconnectivity observed in neurodegenerative disorders may gradually transition into hypoconnectivity among the engaged pathways, thereby contributing to cognitive decline as the disease advances (Marek and Dosenbach, 2022).

In Table 2A, the Heschl's gyrus (Heschl's gyrus Left → Rolandic operculum Left with edge-wise p-value 0.0008) is prominent for lower CFC weight in AD compared to CN subjects. The Heschl's gyrus is not only important for language comprehension, but it also has a crucial role in speech production, phonologic retrieval, and semantic processing (Warrier et al., 2009; Fernández et al., 2020), and has been reported in the literature to be impacted by AD (Hänggi et al., 2011; Dhanjal et al., 2013). The Thalamus is also present among the list of regions in Table 2A (Thalamus Right → Thalamus Left with edge-wise p-value 0.002). The Thalamus functions as a relay station between different sub-cortical areas and the cerebral cortex and also plays a role in sleep, wakefulness, consciousness, and memory (Steriade and Llinás, 1988; Gazzaniga et al., 2002; Aggleton et al., 2010; Bruno et al., 2013), and is also known to be impacted by AD (Braak and Braak, 1991; de Jong et al., 2008). Also present in the table is the Posterior cingulate gyrus (Posterior cingulate Left → Angular gyrus Left with edge-wise p-value of 0.006), which plays an essential role in memory integration and attentional processing, and is widely considered to be impacted by AD (Villain et al., 2008; Jacobs et al., 2013; Li et al., 2015; Badhwar et al., 2017). The Hippocampus, which is involved in long-term memory formation and memory retrieval, is not in the list of regions, yet exhibits a trend of reduction in CFC weight in AD compared to CN (Hippocampus Right → Parahippocampal gyrus Right, edge-wise p-value 0.033) (Boutet et al., 2014; Rao et al., 2022). Self-connections in Hippocampus have been reported to be often involved in compensatory mechanisms leading to increased strength in AD (Pasquini et al., 2015). In Table 2B, the self-connection Parahippocampal gyrus Right → Parahippocampal gyrus Right (edge-wise p-value 0.0008) is prominent for greater weight in AD compared to CN. It is known that the Parahippocampal gyrus is highly impacted by AD and is the focus of damage during disease onset, in a manner such that its connectivity to other regions of the brain decreases with AD, while its activity and intrinsic connectivity within the region increases with AD (Van Hoesen et al., 2000; Chen et al., 2014; Pasquini et al., 2015, 2016; Tahmasian et al., 2015).

TPC identified 1,475 edges in the CFCs across subjects with CN and AD. To obtain the subset of edges which have significant inter-group difference at Bonferonni family-wise error rate of 0.05 requires a total of 6,352 unique subjects across three groups (2,117 per group) to ensure a family-wise power of 0.95 in detecting mean differences of a quarter of the standard deviation (Cohen's $D = 0.25$), computed by power_t_test function in MESS package in R (Ekstrøm, 2023). None of the databases that are available publicly have so many subjects. For example, ADNI has under 2,000 subjects (Weiner et al., 2017, https://adni.loni.usc.edu/adni-3/), and the Australian database has 2,359 subjects (Fowler et al., 2021, https://aibl.org.au/about/). Therefore we took a subset of the ADNI dataset

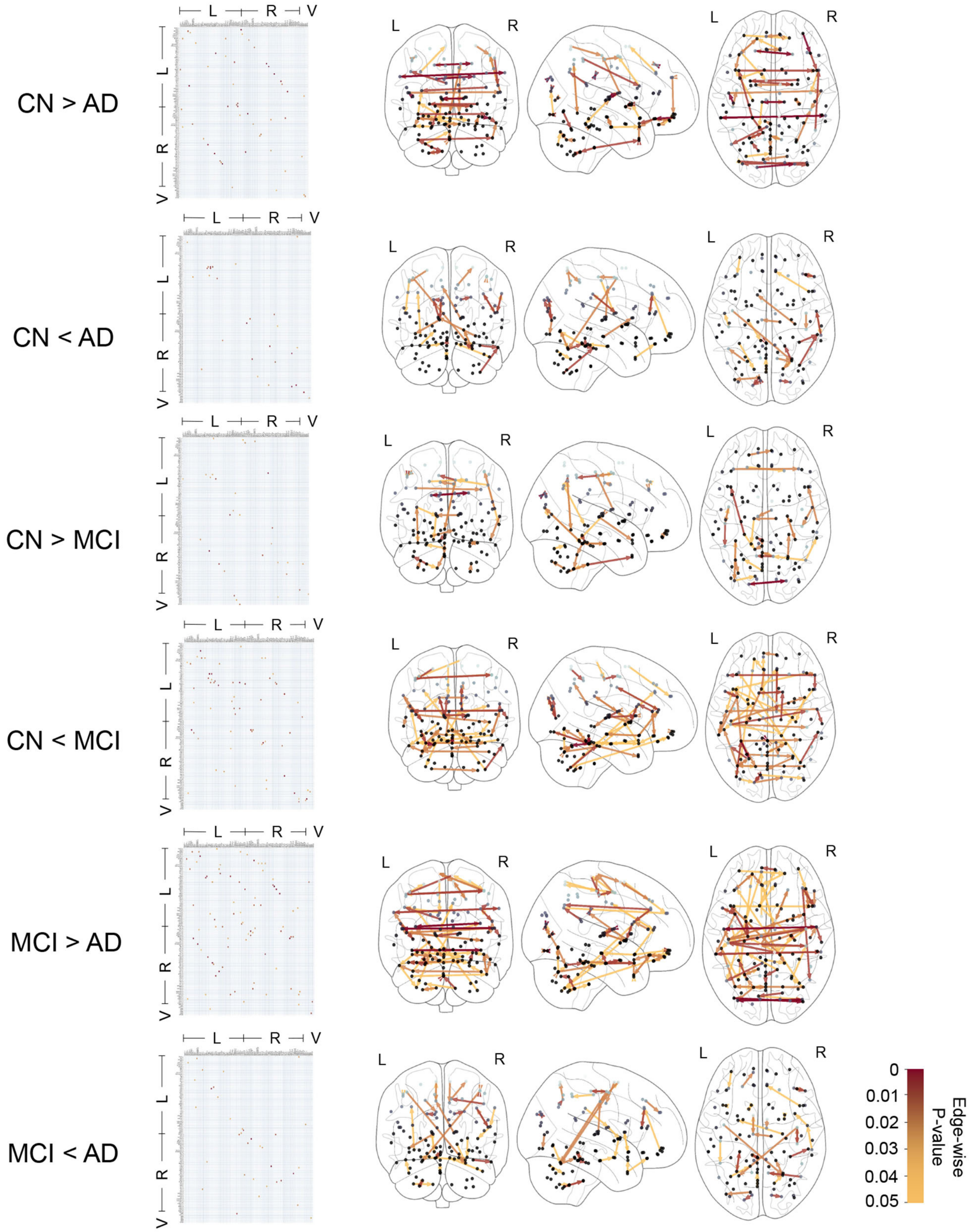


FIGURE 4
Causal functional connections with edge-weights differing between clinical groups with edge-wise p-values ranging in $0 - 0.05$ based on t-test. The edge-wise p-values are represented by a matrix whose entry in (i, j) corresponds to the edge $i \rightarrow j$ and also represented by graph edges on brain schematics. The brain regions are annotated by Left (L) and Right (R) hemispheres of the brain and Vermis (V).

TABLE 2 CFC edges with lowest edge-wise p-values for (a) greater weight in CN compared to AD group and (b) greater weight in AD compared to CN group.

Edge	p-value	Region name	Reported by
(a) CN > AD			
HES_L → ROL_L	0.0008	Heschl's gyrus	Hänggi et al., 2011; Dhanjal et al., 2013
ITG_R → ITG_R	0.001	Inferior temporal gyrus	Palmer and Burns, 1994; Scheff et al., 2011
SOG_L → SOG_R	0.001	Superior occipital gyrus	Beyer et al., 2009; Mao et al., 2021
SFG_L → SFG_R	0.001	Superior frontal gyrus	Brachova et al., 1993; Lue et al., 1996
MFGorb_R → IFGorb_R	0.002	Middle frontal gyrus	Neufang et al., 2011; Zhou et al., 2013
THA_R → THA_L	0.002	Thalamus	Braak and Braak, 1991; de Jong et al., 2008
SMG_L → SMG_R	0.002	SupraMarginal gyrus	Grignon et al., 1998; Desikan et al., 2009
IOG_L → IOG_R	0.005	Inferior occipital gyrus	Johnen et al., 2015; Wu et al., 2023
SMG_R → SMG_L	0.005	SupraMarginal gyrus	Grignon et al., 1998; Desikan et al., 2009
PCC_L → ANG_L	0.006	Posterior cingulate gyrus	Villain et al., 2008; Mascali et al., 2015; Caminiti et al., 2020
(b) CN < AD			
PHG_R → PHG_R	0.0008	Parahippocampal gyrus	Van Hoesen et al., 2000; Thangavel et al., 2008
REC_R → REC_R	0.004	Gyrus rectus	Mölsä et al., 1987; Nochlin et al., 1993; Sheline et al., 2010
CER6_R → CER4_5_R	0.008	Cerebellum	Joachim et al., 1989; Jacobs et al., 2018
IFGtriang_R → MFG_R	0.008	Inferior frontal gyrus	Eliasova et al., 2014; Cajanus et al., 2019
CER7b_L → ITG_R	0.008	Cerebellum	Joachim et al., 1989; Jacobs et al., 2018
FFG_R → ITG_R	0.008	Fusiform gyrus	Whitwell, 2010; Ma et al., 2020
CUN_L → SOG_L	0.014	Cuneus	He et al., 2007; Niskanen et al., 2011
VER1_2 → VER3	0.014	Vermis	Sjöbeck and Englund, 2001; A Mavroudis et al., 2013
CAL_L → SOG_L	0.014	Calcarine fissure	Ren et al., 2020; Yang et al., 2020
PCL_L → PCL_L	0.017	Paracentral lobule	Garcia Martin et al., 2013; Yang et al., 2019

The corresponding source brain regions are in agreement with regions reported in literature (right column) as impacted by AD.

that is captured using 3T fMRI scanner while matching education and age levels for exploratory analysis.

Based on the whole-brain CFC outcome alone, this study obtained brain regions that have been reported across more than 30 different studies of altered connectivity in AD, using different feature extraction methods and advanced imaging technologies (see Table 2). This demonstrates the promise of CFC computed by the TPC algorithm based on directed graphical models in a time series setting. Given the nature of AD, progressively more and more regions of the brain get impacted. Therefore, we make the case for the collection of larger datasets to enable the identification, at desirable levels of significance, of various subnetworks that alter with AD. This would promote the maturation and the use of the TPC-CFC (and other approaches) for prognostic and diagnostic purposes for AD.

It is noteworthy that machine-learning-based classifiers can help predict the clinical category of subjects and diagnose AD (Zhang et al., 2011, 2014; Gray et al., 2013; Salehi et al., 2020; Wen et al., 2020). Recently, researchers have proposed robust multi-class classification methods in the presence of incorrect labeling of classes using the broad learning system (Jin et al., 2021, 2023). Such classifiers would be able to classify a subject as belonging to one of the clinical categories, given a subject's fMRI time-series data as input. However, such classifiers do not compute the CFC between brain regions. Computing the CFC can nicely complement a classifier by providing insights into specific causal functional connections and subnetworks that are altered by AD (Chen et al., 2011; Du et al., 2018). Abnormal resting-state FC between brain regions is known to predate brain atrophy and the emergence of AD symptoms by upto two decades or more (Sheline and Raichle, 2013; Brier et al., 2014; Ashraf et al., 2015; Nakamura et al., 2017). Therefore, a subject's computed CFC can shed light on such abnormalities and promises to be a biomarker for early diagnosis and prognosis of the disease.

In this paper, we have demonstrated the following: (a) Application of the TPC algorithm to compute whole-brain CFC for each subject, (b) Comparison of CFCs computed using other approaches, (c) Interpretation of CFC in the context of AD using domain (neuropathological) knowledge, and (d) Exploratory analysis for edge-wise differences and corresponding brain regions with altered connectivity in subjects with AD compared to CN.

The findings are consistent with published medical literature. In summary, our results show the promise of computing the whole-brain CFC from fMRI data using the TPC algorithm to gain prognostic and diagnostic insights.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: http://adni.loni.usc.edu.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the patients/ participants or patients/participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

RB and SS initiated the study, developed the methods, verified the results, and wrote and edited the manuscript. RB implemented the methods. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors would like to thank the reviewers for their insightful comments and suggestions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncom.2023.1251301/full#supplementary-material

References

Aggleton, J. P., O'Mara, S. M., Vann, S. D., Wright, N. F., Tsanov, M., Erichsen, J. T., et al. (2010). Hippocampal-anterior thalamic pathways for memory: uncovering a network of direct and indirect actions. *Eur. J. Neurosci.* 31, 2292–2307. doi: 10.1111/j.1460-9568.2010.07251.x

Arslan, S., Ktena, S. I., Makropoulos, A., Robinson, E. C., Rueckert, D., Parisot, S., et al. (2018). Human brain mapping: a systematic comparison of parcellation methods for the human cerebral cortex. *Neuroimage* 170, 5–30. doi: 10.1016/j.neuroimage.2017.04.014

Ashraf, A., Fan, Z., Brooks, D., and Edison, P. (2015). Cortical hypermetabolism in mci subjects: a compensatory mechanism? *Eur. J. Nucl. Med. Mol. Imaging* 42, 447–458. doi: 10.1007/s00259-014-2919-z

Badhwar, A., Tam, A., Dansereau, C., Orban, P., Hoffstaedter, F., Bellec, P., et al. (2017). Resting-state network dysfunction in Alzheimer's disease: a systematic review and meta-analysis. *Alzheimers Dement.* 8, 73–85. doi: 10.1016/j.dadm.2017.03.007

Beyer, N., Coulson, D. T., Heggarty, S., Ravid, R., Irvine, G. B., Hellemans, J., et al. (2009). Znt3 mRNA levels are reduced in Alzheimer's disease post-mortem brain. *Mol. Neurodegener.* 4, 1–10. doi: 10.1186/1750-1326-4-53

Biswas, R., and Mukherjee, S. (2022). Consistent causal inference from time series with pc algorithm and its time-aware extension. *arXiv* [preprint]. doi: 10.48550/arXiv.2210.09038

Biswas, R., and Shlizerman, E. (2022a). Statistical perspective on functional and causal neural connectomics: a comparative study. *Front. Syst. Neurosci.* 16, 817962. doi: 10.3389/fnsys.2022.817962

Biswas, R., and Shlizerman, E. (2022b). Statistical perspective on functional and causal neural connectomics: the time-aware pc algorithm. *PLoS Comput. Biol.* 18, 1–27. doi: 10.1371/journal.pcbi.1010653

Boutet, C., Chupin, M., Lehéricy, S., Marrakchi-Kacem, L., Epelbaum, S., Poupon, C., et al. (2014). Detection of volume loss in hippocampal layers in Alzheimer's disease using 7 t MRI: a feasibility study. *Neuroimage Clin.* 5, 341–348. doi: 10.1016/j.nicl.2014.07.011

Bozzali, M., Dowling, C., Serra, L., Spano, B., Torso, M., Marra, C., et al. (2015). The impact of cognitive reserve on brain functional connectivity in Alzheimer's disease. *J. Alzheimers Dis.* 44, 243–250. doi: 10.3233/JAD-141824

Braak, H., and Braak, E. (1991). Alzheimer's disease affects limbic nuclei of the thalamus. *Acta Neuropathol.* 81, 261–268. doi: 10.1007/BF00305867

Brachova, L., Lue, L.-F., Schultz, J., El Rashidy, T., and Rogers, J. (1993). Association cortex, cerebellum, and serum concentrations of C1q and factor B in Alzheimer's disease. *Mol. Brain Res.* 18, 329–334. doi: 10.1016/0169-328X(93)90097-9

Brier, M. R., Thomas, J. B., Fagan, A. M., Hassenstab, J., Holtzman, D. M., Benzinger, T. L., et al. (2014). Functional connectivity and graph theory in preclinical Alzheimer's disease. *Neurobiol. Aging* 35, 757–768. doi: 10.1016/j.neurobiolaging.2013.10.081

Bruno, M.-A., Laureys, S., and Demertzi, A. (2013). Coma and disorders of consciousness. *Handb. Clin. Neurol.* 118, 205–213. doi: 10.1016/B978-0-444-53501-6.00017-2

Cajanus, A., Solje, E., Koikkalainen, J., Lötjönen, J., Suhonen, N.-M., Hallikainen, I., et al. (2019). The association between distinct frontal brain volumes and behavioral symptoms in mild cognitive impairment, Alzheimer's disease, and frontotemporal dementia. *Front. Neurol.* 10, 1059. doi: 10.3389/fneur.2019.01059

Caminiti, S., Sala, A., Pilotto, A., Presotto, L., Garibotto, V., D'Amelio, M., et al. (2020). Imaging dopamine system transporter activity and connectivity in Alzheimer's dementia: neuroimaging/new imaging methods. *Alzheimers Dement.* 16, e043304. doi: 10.1002/alz.043304

Chen, G., Ward, B. D., Chen, G., and Li, S.-J. (2014). Decreased effective connectivity from cortices to the right parahippocampal gyrus in Alzheimer's disease subjects. *Brain Connect.* 4, 702–708. doi: 10.1089/brain.2014.0295

Chen, G., Ward, B. D., Xie, C., Li, W., Wu, Z., Jones, J. L., et al. (2011). Classification of alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional mr imaging. *Radiology* 259, 213–221. doi: 10.1148/radiol.10100734

Chen, J. E., Rubinov, M., and Chang, C. (2017). Methods and considerations for dynamic analysis of functional mr imaging data. *Neuroimaging Clin.* 27, 547–560. doi: 10.1016/j.nic.2017.06.009

Dadgostar, M., Setarehdan, S. K., Shahzadi, S., and Akin, A. (2016). Functional connectivity of the pfc via partial correlation. *Optik* 127, 4748–4754. doi: 10.1016/j.ijleo.2016.01.139

de Jong, L. W., van der Hiele, K., Veer, I. M., Houwing, J., Westendorp, R., Bollen, E., et al. (2008). Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain* 131, 3277–3285. doi: 10.1093/brain/awn278

Deshpande, G., LaConte, S., James, G. A., Peltier, S., and Hu, X. (2009). Multivariate granger causality analysis of fMRI data. *Hum. Brain Mapp*. 30, 1361–1373. doi: 10.1002/hbm.20606

Desikan, R. S., Cabral, H. J., Hess, C. P., Dillon, W. P., Glastonbury, C. M., Weiner, M. W., et al. (2009). Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. *Brain* 132, 2048–2057. doi: 10.1093/brain/awp123

Dhanjal, N. S., Warren, J. E., Patel, M. C., and Wise, R. J. (2013). Auditory cortical function during verbal episodic memory encoding in Alzheimer's disease. *Ann. Neurol*. 73, 294–302. doi: 10.1002/ana.23789

Du, Y., Fu, Z., and Calhoun, V. D. (2018). Classification and prediction of brain disorders using functional connectivity: promising but challenging. *Front. Neurosci*. 12, 525. doi: 10.3389/fnins.2018.00525

Ekstrøm, C. T. (2023). *MESS: Miscellaneous Esoteric Statistical Scripts*. R package version 0.5.12. Available online at: https://cran.r-project.org/web/packages/MESS/MESS.pdf (accessed November 12, 2023).

Eliasova, I., Anderkova, L., Marecek, R., and Rektorova, I. (2014). Non-invasive brain stimulation of the right inferior frontal gyrus may improve attention in early Alzheimer's disease: a pilot study. *J. Neurol. Sci*. 346, 318–322. doi: 10.1016/j.jns.2014.08.036

Fernández, L., Velásquez, C., Porrero, J. A. G., de Lucas, E. M., and Martino, J. (2020). Heschl's gyrus fiber intersection area: a new insight on the connectivity of the auditory-language hub. *Neurosurg. Focus* 48, E7. doi: 10.3171/2019.11.FOCUS19778

Fowler, C., Rainey-Smith, S. R., Bird, S., Bomke, J., Bourgeat, P., Brown, B. M., et al. (2021). Fifteen years of the australian imaging, biomarkers and lifestyle (AIBL) study: progress and observations from 2,359 older adults spanning the spectrum from cognitive normality to Alzheimer's disease. *J. Alzheimers Dis. Rep*. 5, 443–468. doi: 10.3233/ADR-210005

Fredericks, C. A., Sturm, V. E., Brown, J. A., Hua, A. Y., Bilgel, M., Wong, D. F., et al. (2018). Early affective changes and increased connectivity in preclinical Alzheimer's disease. *Alzheimers Dement*. 10, 471–479. doi: 10.1016/j.dadm.2018.06.002

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9, 432–441. doi: 10.1093/biostatistics/kxm045

Friston, K. J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302. doi: 10.1016/S1053-8119(03)00202-7

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., Frackowiak, R. S., et al. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp*. 2, 189–210. doi: 10.1002/hbm.460020402

Garcia Martin, E., De Hoz, R., Rojas, B., Gil, P., Yubero, R., and Ramirez, J. (2013). Macular nerve-fiber-layer measurement in early stage Alzheimer's disease using optical coherence tomography. *Acta Ophthalmol*. 91. doi: 10.1111/j.1755-3768.2013.3771.x

Gazzaniga, M. S., Ivry, R. B., and Mangun, G. R. (2002). *Cognitive Neuroscience?: The Biology of the Mind (2nd ed.)*. Norton.

Gour, N., Felician, O., Didic, M., Koric, L., Gueriot, C., Chanoine, V., et al. (2014). Functional connectivity changes differ in early and late-onset Alzheimer's disease. *Hum. Brain Mapp*. 35, 2978–2994. doi: 10.1002/hbm.22379

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438. doi: 10.2307/1912791

Granger, C. W. (2001). *Essays in Econometrics: Collected Papers of Clive WJ Granger, Volume 32*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511753978

Gray, K. R., Aljabar, P., Heckemann, R. A., Hammers, A., Rueckert, D., Initiative, A. D. N., et al. (2013). Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *Neuroimage* 65, 167–175. doi: 10.1016/j.neuroimage.2012.09.065

Grignon, Y., Duyckaerts, C., Bennecib, M., and Hauw, J.-J. (1998). Cytoarchitectonic alterations in the supramarginal gyrus of late onset Alzheimer's disease. *Acta Neuropathol*. 95, 395–406. doi: 10.1007/s004010050816

Hänggi, J., Streffer, J., Jäncke, L., and Hock, C. (2011). Volumes of lateral temporal and parietal structures distinguish between healthy aging, mild cognitive impairment, and Alzheimer's disease. *J. Alzheimers Dis*. 26, 719–734. doi: 10.3233/JAD-2011-101260

He, Y., Wang, L., Zang, Y., Tian, L., Zhang, X., Li, K., et al. (2007). Regional coherence changes in the early stages of Alzheimer's disease: a combined structural and resting-state functional MRI study. *Neuroimage* 35, 488–500. doi: 10.1016/j.neuroimage.2006.11.042

Hillary, F. G., and Grafman, J. H. (2017). Injured brains and adaptive networks: the benefits and costs of hyperconnectivity. *Trends Cogn. Sci*. 21, 385–401. doi: 10.1016/j.tics.2017.03.003

Jacobs, H. I., Hopkins, D. A., Mayrhofer, H. C., Bruner, E., van Leeuwen, F. W., Raaijmakers, W., et al. (2018). The cerebellum in Alzheimer's disease: evaluating its role in cognitive decline. *Brain* 141, 37–47. doi: 10.1093/brain/awx194

Jacobs, H. I., Radua, J. Lückmann, H. C., and Sack, A. T. (2013). Meta-analysis of functional network alterations in Alzheimer's disease: toward a network biomarker. *Neurosci. Biobehav. Rev*. 37, 753–765. doi: 10.1016/j.neubiorev.2013.03.009

Jin, J., Li, Y., and Chen, C. P. (2021). Pattern classification with corrupted labeling via robust broad learning system. *IEEE Trans. Knowl. Data Eng*. 34, 4959–4971. doi: 10.1109/TKDE.2021.3049540

Jin, J., Geng, B., Li, Y., Liang, J., Xiao, Y., Chen, C. P., et al. (2023). Flexible label-induced manifold broad learning system for multiclass recognition. *IEEE Trans. Neural Netw. Learn. Syst*. 1–15. doi: 10.1109/TNNLS.2023.3291793

Joachim, C. L., Morris, J. H., and Selkoe, D. J. (1989). Diffuse senile plaques occur commonly in the cerebellum in Alzheimer's disease. *Am. J. Pathol*. 135, 309.

Johnen, A., Brandstetter, L., Lohmann, H., and Duning, T. (2015). P12 neural correlates of apraxia in mild dementia of Alzheimer's disease-a voxel-based morphometry study. *Clin. Neurophysiol*. 126, e93. doi: 10.1016/j.clinph.2015.04.133

Kalisch, M., and Bühlman, P. (2007). Estimating high-dimensional directed acyclic graphs with the pc-algorithm. *J. Mach. Learn. Res*. 8,613–636.

Keilholz, S., Caballero-Gaudes, C., Bandettini, P., Deco, G., and Calhoun, V. (2017). Time-resolved resting-state functional magnetic resonance imaging analysis: current status, challenges, and new directions. *Brain Connect*. 7, 465–481. doi: 10.1089/brain.2017.0543

Khatri, U., Lama, R. K., and Kwon, G.-R. (2021). "Diagnosis of Alzheimer's disease using effective connectivity of rs-fMRI," in *2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)* (Jeju: IEEE), 1–4. doi: 10.1109/ITC-CSCC52171.2021.9501447

Kim, H. J., Cha, J., Lee, J.-M., Shin, J. S., Jung, N.-Y., Kim, Y. J., et al. (2016). Distinctive resting state network disruptions among Alzheimer's disease, subcortical vascular dementia, and mixed dementia patients. *J. Alzheimers Dis*. 50, 709–718. doi: 10.3233/JAD-150637

Laske, C., Sohrabi, H. R., Frost, S. M., López-de Ipiña, K., Garrard, P., Buscema, M., et al. (2015). Innovative diagnostic tools for early detection of Alzheimer's disease. *Alzheimers Dement*. 11, 561–578. doi: 10.1016/j.jalz.2014.06.004

Li, H., Habes, M., Wolk, D. A., Fan, Y., Initiative, A. D. N., et al. (2019). A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data. *Alzheimers Dement*. 15, 1059–1070. doi: 10.1016/j.jalz.2019.02.007

Li, H.-J., Hou, X.-H., Liu, H.-H., Yue, C.-L., He, Y., Zuo, X.-N., et al. (2015). Toward systems neuroscience in mild cognitive impairment and Alzheimer's disease: a meta-analysis of 75 fMRI studies. *Hum. Brain Mapp*. 36, 1217–1232. doi: 10.1002/hbm.22689

Liu, S., Cai, W., Liu, S., Zhang, F., Fulham, M., Feng, D., et al. (2015). Multimodal neuroimaging computing: the workflows, methods, and platforms. *Brain Inform*. 2, 181–195. doi: 10.1007/s40708-015-0020-4

Lue, L.-F., Brachova, L., Civin, W. H., and Rogers, J. (1996). Inflammation, Aβ deposition, and neurofibrillary tangle formation as correlates of Alzheimer's disease neurodegeneration. *J. Neuropathol. Exp. Neurol*. 55, 1083–1088. doi: 10.1097/00005072-199655100-00008

Ma, D., Fetahu, I. S., Wang, M., Fang, R., Li, J., Liu, H., et al. (2020). The fusiform gyrus exhibits an epigenetic signature for Alzheimer's disease. *Clin. Epigenetics* 12, 1–16. doi: 10.1186/s13148-020-00916-3

Mao, Y., Liao, Z., Liu, X., Li, T., Hu, J., Le, D., et al. (2021). Disrupted balance of long and short-range functional connectivity density in Alzheimer's disease (AD) and mild cognitive impairment (MCI) patients: a resting-state fMRI study. *Ann. Transl. Med*. 9, 65. doi: 10.21037/atm-20-7019

Marek, S., and Dosenbach, N. U. (2022). The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. *Dialogues Clin. Neurosci*. 20, 133–140. doi: 10.31887/DCNS.2018.20.2/smarek

Mascali, D., DiNuzzo, M., Gili, T., Moraschi, M., Fratini, M., Maraviglia, B., et al. (2015). Intrinsic patterns of coupling between correlation and amplitude of low-frequency fMRI fluctuations are disrupted in degenerative dementia mainly due to functional disconnection. *PLoS ONE* 10, 1–18. doi: 10.1371/journal.pone.0120988

Mavroudis, A. I., Manani, M. G., Petrides, F., Petsoglou, K., Njau, S. D., Costa, V. G., et al. (2013). Dendritic and spinal pathology of the purkinje cells from the human cerebellar vermis in Alzheimer's disease. *Psychiatr. Danub*. 25, 221–226.

Mölsä, P., Säkö, E., Paljärvi, L., Rinne, J., and Rinne, U. (1987). Alzheimer's disease: neuropathological correlates of cognitive and motor disorders. *Acta Neurol. Scand*. 75, 376–384. doi: 10.1111/j.1600-0404.1987.tb05465.x

Nakamura, A., Cuesta, P., Kato, T., Arahata, Y., Iwata, K., Yamagishi, M., et al. (2017). Early functional network alterations in asymptomatic elders at risk for Alzheimer's disease. *Sci. Rep*. 7, 1–11. doi: 10.1038/s41598-017-06876-8

Neufang, S., Akhrif, A., Riedl, V., Förstl, H., Kurz, A., Zimmer, C., et al. (2011). Disconnection of frontal and parietal areas contributes to impaired attention in very early Alzheimer's disease. *J. Alzheimers Dis*. 25, 309–321. doi: 10.3233/JAD-2011-102154

Nieto-Castanon, A., and Whitfield-Gabrieli, S. (2021). *CONN functional connectivity toolbox: RRID SCR_009550, release 21.* Boston, MA. doi: 10.56441/hilbertpress.2161.7292

Niskanen, E., Könönen, M., Määttä, S., Hallikainen, M., Kivipelto, M., Casarotto, S., et al. (2011). New insights into Alzheimer's disease progression: a combined tms and structural MRI study. *PLoS ONE* 6, e26113. doi: 10.1371/journal.pone.0026113

Nochlin, D., Van Belle, G., Bird, T., and Sumi, S. (1993). Comparison of the severity of neuropathologic changes in familial and sporadic Alzheimer's disease. *Alzheimer Dis. Assoc. Disord.* 7, 212–222.

Oldham, S., and Fornito, A. (2019). The development of brain network hubs. *Dev. Cogn. Neurosci.* 36, 100607. doi: 10.1016/j.dcn.2018.12.005

Palmer, A. M., and Burns, M. A. (1994). Selective increase in lipid peroxidation in the inferior temporal cortex in Alzheimer's disease. *Brain Res.* 645, 338–342. doi: 10.1016/0006-8993(94)91670-5

Pasquini, L., Scherr, M., Tahmasian, M., Meng, C., Myers, N. E., Ortner, M., et al. (2015). Link between hippocampus' raised local and eased global intrinsic connectivity in AD. *Alzheimers Dement.* 11, 475–484. doi: 10.1016/j.jalz.2014.02.007

Pasquini, L., Scherr, M., Tahmasian, M., Myers, N. E., Ortner, M., Kurz, A., et al. (2016). Increased intrinsic activity of medial-temporal lobe subregions is associated with decreased cortical thickness of medial-parietal areas in patients with Alzheimer's disease dementia. *J. Alzheimers Dis.* 51, 313–326. doi: 10.3233/JAD-150823

Pearl, J. (2009). *Causality.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511803161

Penalba-Sánchez, L., Oliveira-Silva, P., Sumich, A. L., and Cifre, I. (2023). Increased functional connectivity patterns in mild Alzheimer's disease: a rsfMRI study. *Front. Aging Neurosci.* 14, 1037347. doi: 10.3389/fnagi.2022.1037347

Querfurth, H. W., and LaFerla, F. M. (2010). Alzheimer's disease. *N. Engl. J. Med.* 362, 329–344. doi: 10.1056/NEJMra0909142

Rao, Y. L., Ganaraja, B., Murlimanju, B., Joy, T., Krishnamurthy, A., Agrawal, A., et al. (2022). Hippocampus and its involvement in Alzheimer's disease: a review. *3 Biotech* 12, 55. doi: 10.1007/s13205-022-03123-4

Reid, A. T., Headley, D. B., Mill, R. D., Sanchez-Romero, R., Uddin, L. Q., Marinazzo, D., et al. (2019). Advancing functional connectivity research from association to causation. *Nat. Neurosci.* 22, 1751–1760. doi: 10.1038/s41593-019-0510-4

Ren, S., Huang, Q., Jiang, D., Huang, L., Wang, Y., Guan, Y., et al. (2020). Brain amyloid accumulation and glucose hypometabolism in Chinese Alzheimer's disease population: neuroimaging/multi-modal comparisons. *Alzheimers Dement.* 16, e043567. doi: 10.1002/alz.043567

Rokem, A., Trumpis, M., and Perez, F. (2009). "Nitime: time-series analysis for neuroimaging data," in *Proceedings of the 8th Python in Science Conference* (Pasadena, CA), 68–75.

Rytsar, R., Fornari, E., Frackowiak, R. S., Ghika, J. A., and Knyazeva, M. G. (2011). Inhibition in early Alzheimer's disease: an fMRI-based study of effective connectivity. *Neuroimage* 57, 1131–1139. doi: 10.1016/j.neuroimage.2011.05.029

Salehi, A. W., Baglat, P., Sharma, B. B., Gupta, G., and Upadhya, A. (2020). "A CNN model: earlier diagnosis and classification of alzheimer disease using MRI," in *2020 International Conference on Smart Electronics and Communication (ICOSEC)* (Trichy: IEEE), 156–161. doi: 10.1109/ICOSEC49089.2020.9215402

Scarapicchia, V., Mazerolle, E. L., Fisk, J. D., Ritchie, L. J., and Gawryluk, J. R. (2018). Resting state bold variability in Alzheimer's disease: a marker of cognitive decline or cerebrovascular status? *Front. Aging Neurosci.* 10, 39. doi: 10.3389/fnagi.2018.00039

Scheff, S. W., Price, D. A., Schmitt, F. A., Scheff, M. A., and Mufson, E. J. (2011). Synaptic loss in the inferior temporal gyrus in mild cognitive impairment and Alzheimer's disease. *J. Alzheimers Dis.* 24, 547–557. doi: 10.3233/JAD-2011-101782

Schmidt, C., Pester, B., Schmid-Hertel, N., Witte, H., Wismüller, A., and Leistritz, L. (2016). A multivariate granger causality concept towards full brain functional connectivity. *PLoS ONE* 11, e0153105. doi: 10.1371/journal.pone.0153105

Schouten, T. M., Koini, M., De Vos, F., Seiler, S., Van Der Grond, J., Lechner, A., et al. (2016). Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease. *Neuroimage Clin.* 11, 46–51. doi: 10.1016/j.nicl.2016.01.002

Sheline, Y. I., and Raichle, M. E. (2013). Resting state functional connectivity in preclinical Alzheimer's disease. *Biol. Psychiatry* 74, 340–347. doi: 10.1016/j.biopsych.2012.11.028

Sheline, Y. I., Raichle, M. E., Snyder, A. Z., Morris, J. C., Head, D., Wang, S., et al. (2010). Amyloid plaques disrupt resting state default mode network connectivity in cognitively normal elderly. *Biol. Psychiatry* 67, 584–587. doi: 10.1016/j.biopsych.2009.08.024

Sjöbeck, M., and Englund, E. (2001). Alzheimer's disease and the cerebellum: a morphologic study on neuronal and glial changes. *Dement. Geriatr. Cogn. Disord.* 12, 211–218. doi: 10.1159/000051260

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for fMRI. *Neuroimage* 54, 875–891. doi: 10.1016/j.neuroimage.2010.08.063

Spirtes, P., Glymour, C. N., Scheines, R., and Heckerman, D. (2000). *Causation, Prediction, and Search.* Cambridge, MA: MIT press. doi: 10.7551/mitpress/1754.001.0001

Sporns, O. (2013). The human connectome: origins and challenges. *Neuroimage* 80, 53–61. doi: 10.1016/j.neuroimage.2013.03.023

Steriade, M., and Llinás, R. R. (1988). The functional states of the thalamus and the associated neuronal interplay. *Physiol. Rev.* 68, 649–742. doi: 10.1152/physrev.1988.68.3.649

Tahmasian, M., Pasquini, L., Scherr, M., Meng, C., Förster, S., Bratec, S. M., et al. (2015). The lower hippocampus global connectivity, the higher its local metabolism in alzheimer disease. *Neurology* 84, 1956–1963. doi: 10.1212/WNL.0000000000001575

Thangavel, R., Van Hoesen, G. W., and Zaheer, A. (2008). Posterior parahippocampal gyrus pathology in Alzheimer's disease. *Neuroscience* 154, 667–676. doi: 10.1016/j.neuroscience.2008.03.077

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978

Van Hoesen, G. W., Augustinack, J. C., Dierking, J., Redman, S. J., and Thangavel, R. (2000). The parahippocampal gyrus in Alzheimer's disease: clinical and preclinical neuroanatomical correlates. *Ann. N. Y. Acad. Sci.* 911, 254–274. doi: 10.1111/j.1749-6632.2000.tb06731.x

Villain, N., Desgranges, B., Viader, F., De La Sayette, V., Mézenge, F., Landeau, B., et al. (2008). Relationships between hippocampal atrophy, white matter disruption, and gray matter hypometabolism in Alzheimer's disease. *J. Neurosci.* 28, 6174–6181. doi: 10.1523/JNEUROSCI.1392-08.2008

Wang, K., Liang, M., Wang, L., Tian, L., Zhang, X., Li, K., et al. (2007). Altered functional connectivity in early Alzheimer's disease: a resting-state fMRI study. *Hum. Brain Mapp.* 28, 967–978. doi: 10.1002/hbm.20324

Wang, Y., Kang, J., Kemmer, P. B., and Guo, Y. (2016). An efficient and reliable statistical method for estimating functional connectivity in large scale brain networks using partial correlation. *Front. Neurosci.* 10, 123. doi: 10.3389/fnins.2016.00123

Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., et al. (2009). Relating structure to function: Heschl's gyrus and acoustic processing. *J. Neurosci.* 29, 61–69. doi: 10.1523/JNEUROSCI.3489-08.2009

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., et al. (2017). The Alzheimer's disease neuroimaging initiative 3: continued innovation for clinical trial improvement. *Alzheimers Dement.* 13, 561–571. doi: 10.1016/j.jalz.2016.10.006

Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., et al. (2020). Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. *Med. Image Anal.* 63, 101694. doi: 10.1016/j.media.2020.101694

Whitwell, J. L. (2010). Progression of atrophy in Alzheimer's disease and related disorders. *Neurotox. Res.* 18, 339–346. doi: 10.1007/s12640-010-9175-1

Wu, H., Song, Y., Yang, X., Chen, S., Ge, H., Yan, Z., et al. (2023). Functional and structural alterations of dorsal attention network in preclinical and early-stage Alzheimer's disease. *CNS Neurosci. Therap.* 29, 1512–1524. doi: 10.1111/cns.14092

Xue, C., Yuan, B., Yue, Y., Xu, J., Wang, S., Wu, M., et al. (2019). Distinct disruptive patterns of default mode subnetwork connectivity across the spectrum of preclinical Alzheimer's disease. *Front. Aging Neurosci.* 11, 307. doi: 10.3389/fnagi.2019.00307

Yamasaki, T., Muranaka, H., Kaseda, Y., Mimori, Y., and Tobimatsu, S. (2012). Understanding the pathophysiology of Alzheimer's disease and mild cognitive impairment: a mini review on fMRI and erp studies. *Neurol. Res. Int.* 2012, 719056. doi: 10.1155/2012/719056

Yang, H., Xu, H., Li, Q., Jin, Y., Jiang, W., Wang, J., et al. (2019). Study of brain morphology change in Alzheimer's disease and amnestic mild cognitive impairment compared with normal controls. *Gen. Psychiatry* 32, e100005. doi: 10.1136/gpsych-2018-100005

Yang, L., Yan, Y., Li, Y., Hu, X., Lu, J., Chan, P., et al. (2020). Frequency-dependent changes in fractional amplitude of low-frequency oscillations in Alzheimer's disease: a resting-state fMRI study. *Brain Imaging Behav.* 14, 2187–2201. doi: 10.1007/s11682-019-00169-6

Yuen, K. K. (1974). The two-sample trimmed t for unequal population variances. *Biometrika* 61, 165–170. doi: 10.1093/biomet/61.1.165

Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., Initiative, A. D. N., et al. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867. doi: 10.1016/j.neuroimage.2011.01.008

Zhang, Y.-D., Wang, S., and Dong, Z. (2014). Classification of alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. *Prog. Electromagn. Res.* 144, 171–184. doi: 10.2528/PIER13121310

Zhou, B., Liu, Y., Zhang, Z., An, N., Yao, H., Wang, P., et al. (2013). Impaired functional connectivity of the thalamus in Alzheimer's disease and mild cognitive impairment: a resting-state fMRI study. *Curr. Alzheimer Res.*, 10, 754–766. doi: 10.2174/15672050113109990146

frontiers | Frontiers in Computational Neuroscience

TYPE Original Research
PUBLISHED 12 June 2024
DOI 10.3389/fncom.2024.1418546

OPEN ACCESS

EDITED BY
Ridha Ejbali,
University of Gabes, Tunisia

REVIEWED BY
Poonam Chaudhary,
The Northcap University, India
Anchal Garg,
University of Bolton, United Kingdom
Bishwajeet Kumar Pandey,
Astana University, Kazakhstan
M. Murugappan,
Kuwait College of Science and Technology,
Kuwait

*CORRESPONDENCE
Eid Albalawi
✉ ealbalawi@kfu.edu.sa
Muhammad Shahid Anwar
✉ shahidanwar786@gachon.ac.kr

RECEIVED 16 April 2024
ACCEPTED 23 May 2024
PUBLISHED 12 June 2024

CITATION
Albalawi E, Thakur A, Dorai DR, Bhatia Khan S, Mahesh TR, Almusharraf A, Aurangzeb K and Anwar MS (2024) Enhancing brain tumor classification in MRI scans with a multi-layer customized convolutional neural network approach.
Front. Comput. Neurosci. 18:1418546.
doi: 10.3389/fncom.2024.1418546

Enhancing brain tumor classification in MRI scans with a multi-layer customized convolutional neural network approach

Eid Albalawi[1]*, Arastu Thakur[2], D. Ramya Dorai[3], Surbhi Bhatia Khan[4,5], T. R. Mahesh[2], Ahlam Almusharraf[6], Khursheed Aurangzeb[7] and Muhammad Shahid Anwar[8]*

[1]Department of Computer Science, King Faisal University, Al-Ahsa, Saudi Arabia, [2]Department of Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bangalore, India, [3]Department of Information Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bangalore, India, [4]School of Science, Engineering and Environment, University of Salford, Manchester, United Kingdom, [5]Department of Electrical and Computer Engineering, Lebanese American University, Byblos, Lebanon, [6]Department of Management, College of Business Administration, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia, [7]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia, [8]Department of AI and Software, Gachon University, Seongnam-si, Republic of Korea

Background: The necessity of prompt and accurate brain tumor diagnosis is unquestionable for optimizing treatment strategies and patient prognoses. Traditional reliance on Magnetic Resonance Imaging (MRI) analysis, contingent upon expert interpretation, grapples with challenges such as time-intensive processes and susceptibility to human error.

Objective: This research presents a novel Convolutional Neural Network (CNN) architecture designed to enhance the accuracy and efficiency of brain tumor detection in MRI scans.

Methods: The dataset used in the study comprises 7,023 brain MRI images from figshare, SARTAJ, and Br35H, categorized into glioma, meningioma, no tumor, and pituitary classes, with a CNN-based multi-task classification model employed for tumor detection, classification, and location identification. Our methodology focused on multi-task classification using a single CNN model for various brain MRI classification tasks, including tumor detection, classification based on grade and type, and tumor location identification.

Results: The proposed CNN model incorporates advanced feature extraction capabilities and deep learning optimization techniques, culminating in a groundbreaking paradigm shift in automated brain MRI analysis. With an exceptional tumor classification accuracy of 99%, our method surpasses current methodologies, demonstrating the remarkable potential of deep learning in medical applications.

Conclusion: This study represents a significant advancement in the early detection and treatment planning of brain tumors, offering a more efficient and accurate alternative to traditional MRI analysis methods.

KEYWORDS

diagnosis of brain tumors, convolutional neural networks, deep learning, classification of medical images, MRI imaging

1 Introduction

The diagnosis of brain tumors represents a critical intersection of neurology and oncology, necessitating precise and efficient methodologies for accurate identification and characterization. Magnetic Resonance Imaging (MRI) stands as a cornerstone in this endeavor, offering detailed visualization of brain anatomy crucial for detecting abnormal growths or lesions indicative of tumors. However, the manual interpretation of MRI scans relies heavily on radiologists' expertise, presenting challenges such as time consumption and susceptibility to human error, ultimately affecting diagnosis accuracy and treatment planning. In Figure 1, different sample images of different tumor types are shown to make it clear why manual interpretation is difficult.

Globally, the incidence rates of brain tumors have been on the rise, underscoring the urgency for more effective diagnostic approaches. Brain tumors exhibit considerable diversity in type, size, location, and malignancy level, further complicating their diagnosis (Zhang and Sejdić, 2019). The study of brain tumor segmentation and classification through neuroimaging methodologies has gained significant importance in recent years due to the potential fatality of undetected tumors (Kumar and Kumar, 2023). Proper classification aids clinicians in providing appropriate treatment, and deep learning, particularly convolutional neural networks (CNN), has achieved notable success in these tasks (Kumar and Kumar, 2023). This study utilized a 25-layer CNN model to classify brain tumors from public MRI datasets, showing superior performance over previous methods, achieving classification accuracies of 86.23 and 81.6% using different optimizers, yet the technological gap remains in enhancing real-time processing and integration with clinical workflows (Sarkar et al., 2023). Another research employed AlexNet CNN with various classifiers, achieving up to 100% accuracy, highlighting the model's effectiveness; however, the gap lies in the need for more extensive datasets and robustness against diverse MRI quality and protocols (Bairagi et al., 2023). The necessity for automatic and reliable detection systems is underscored due to the complex and time-consuming nature of manual tumor detection (Tong and Wang, 2023). The proposed CNN-based system achieved 98.67% accuracy using AlexNet on a specific dataset, yet a gap exists in validating across larger and more varied datasets to ensure generalizability (Tong and Wang, 2023). Furthermore, a dual tri-path CNN system demonstrated high reproducibility and quality in segmentation tasks, crucial for practical application, but the challenge remains in reducing computational complexity without sacrificing accuracy (Tong and Wang, 2023). Lastly, a study on federated learning (FL) combined with CNN ensemble architectures showed promising results in privacy-protected brain tumor classification, achieving 91.05% accuracy, slightly lower than the traditional approach but maintaining data privacy; the technological gap here involves improving the FL model's performance to match centralized models while ensuring scalability and efficiency (Islam et al., 2023). To address these shortcomings, our research introduces a multi-layer customized CNN architecture designed specifically for the nuanced task of brain tumor classification from MRI scans. Our model leverages advanced feature extraction techniques and optimization algorithms to improve diagnostic accuracy and efficiency significantly. Unlike existing models, our approach emphasizes robustness and adaptability across different imaging settings, enhancing its practical utility in diverse clinical environments.

The dataset used in the study comprises 7,023 human brain MRI images sourced from figshare, SARTAJ, and Br35H, categorized into four classes: glioma, meningioma, no tumor, and pituitary. The "no tumor" images are from the Br35H dataset, and due to classification issues in the SARTAJ dataset's glioma class, these images were replaced with those from figshare to ensure accuracy.

1.1 Motivation

The motivation behind this research is to utilize the capabilities of CNNs to improve the accuracy and efficiency of diagnosing brain tumors from MRI scans. Through the development of a specialized CNN architecture, this study aims to tackle the unique challenges of analyzing brain tumor MRIs. The goal is to provide a tool that can help radiologists make quicker and more precise diagnoses, ultimately enhancing patient care. The objectives of this research paper are to:

- Develop a novel convolutional neural network (CNN) architecture that significantly improves the accuracy of brain tumor classification from MRI scans.
- Design the CNN model to effectively generalize across different MRI protocols and imaging conditions, ensuring reliable performance in diverse clinical settings.
- Optimize the model to reduce computational demands, enabling faster processing times suitable for real-time diagnostic applications.

1.2 Contribution of the paper

This research introduces a customized CNN architecture tailored for the classification of brain tumors from MRI images. The goal of this research paper focus on enhancing the accuracy and efficiency of

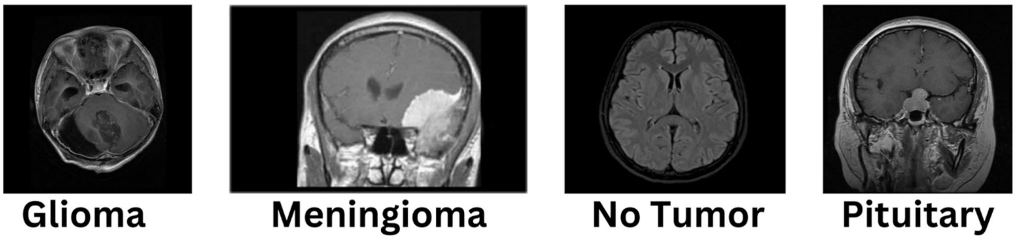


FIGURE 1
Annotated images of different tumors.

diagnosing brain tumors from MRI scans using a tailored Convolutional Neural Network (CNN) architecture. The numerical achievements underscore the significance of these objectives: the proposed model achieved a remarkable tumor classification accuracy of 99%. This level of accuracy is a considerable improvement compared to traditional methods, which often suffer from lower accuracy due to human error and the time-intensive nature of manual interpretations. Such high performance not only validates the efficacy of the specialized CNN in medical imaging tasks but also emphasizes its potential to significantly improve diagnostic processes, thereby enhancing patient care by allowing for quicker and more accurate diagnosis and treatment planning. This achievement highlights the practical relevance and impact of the research, affirming the objectives centered on technical advancement in medical diagnostics.

1.3 Organization of the paper

Following this introduction, the paper is organized into several sections: The next section reviews related work, establishing the context and justifying the need for optimized CNN. The methodology section details the design of custom CNN, the dataset, and training procedures. The results section presents a comparative analysis of proposed method performance against other models. Finally, the discussion and conclusion sections reflect on the findings, their implications for clinical practice, and directions for future research.

2 Related work

The use of artificial intelligence (AI) in medical imaging, specifically employing convolutional neural networks (CNNs) for diagnosing brain tumors from MRI scans, has been a highly researched area with notable advancements. This section delves into various methodologies developed in recent years, highlighting their contributions and limitations, and setting the stage for the introduction of proposed method.

Initially, traditional machine learning techniques such as Support Vector Machines (SVMs) and Random Forests were used for classification, relying on extracted features from MRI scans (Alzubaidi et al., 2021). However, these methods lacked dynamic feature-learning abilities and relied heavily on expert-driven feature selection, potentially overlooking critical details. Early CNN models were shallow due to computational constraints, limiting their ability to capture complex features (Alzubaidi et al., 2021). Deeper architectures like AlexNet and VGG improved feature extraction but faced challenges such as overfitting and the need for extensive labeled datasets (Zhao, 2023). Transfer learning addressed data scarcity issues by fine-tuning models pretrained on large datasets like ImageNet. Integrating multimodal MRI data improved analysis accuracy, although synchronizing features from different modalities posed challenges (Ahmmed et al., 2023). Attention mechanisms enhanced interpretability by focusing on relevant regions, while 3D CNNs preserved spatial relationships for volumetric analysis but introduced computational complexities (Aboussaleh et al., 2023). Ensemble learning improved accuracy but increased computational demands, and domain adaptation aimed to generalize models across different MRI scanners and protocols (Zhao, 2023). Federated learning addressed privacy concerns by training models collaboratively across institutions but faced challenges such as data heterogeneity and communication overhead (Dufumier et al., 2021). A summary of some studies is presented in Table 1.

Recent studies have focused on various advanced methods for brain tumor detection and classification. One study aimed to create a metaheuristic-based system using an enhanced seagull optimization algorithm for feature selection and classification with a deep belief network (Hu and Razmjooy, 2021). Another research developed an automated diagnosis system employing evolutionary algorithms, reinforcement learning, and transfer learning for multi-classification of brain tumors (Sadad et al., 2021). A hybrid deep learning model, DeepTumorNet, used a modified GoogLeNet architecture to classify glioma, meningioma, and pituitary tumors (Nickparvar, 2021). An automated method utilizing morphological-based segmentation was proposed for precise tumor detection in MRI images (Albalawi et al., 2024). Deep learning techniques, specifically a 2D CNN, were employed for early detection of various brain tumors (Mahesh et al., 2024), while an Improved Residual Network (ResNet) aimed to enhance segmentation accuracy (Aggarwal et al., 2023). An FPGA-based accelerator was introduced to improve segmentation speed and accuracy (Xiong et al., 2021), and a YOLO2-based transfer learning approach achieved high classification accuracy (Kumar Sahoo et al., 2023). A deep semi-supervised learning framework integrated CNN features and GAN-generated synthetic MRIs for glioma classification (Ge et al., 2020). Transfer-learning-based models and a CNN called BRAIN-TUMOR-net were developed for classifying MRI images, achieving high accuracy across different datasets (Taher et al., 2022).

However, despite these advancements, the field continues to confront challenges, particularly in the context of brain MRI analysis. The unique complexities of brain anatomy and the diverse manifestations of tumors demand a tailored approach in AI model development. Our study is situated within this specialized domain, introducing a custom-designed CNN architecture optimized for the intricate task of detecting brain tumors in MRI scans. Our proposed model builds on foundational research, integrating state-of-the-art feature extraction and deep learning optimization strategies to tackle the specific challenges of brain MRI data. By enhancing and advancing CNN capabilities in this specialized context, our research contributes to the continual evolution of AI in medical imaging, with the aim of establishing a new standard in accuracy and efficiency for brain tumor diagnosis.

3 Methodology

The proposed method signifies a substantial advancement in utilizing convolutional neural networks (CNNs) for analyzing brain tumor MRI scans. This innovative network architecture is tailored to tackle the complex challenges of brain tumor classification and segmentation, harnessing deep learning to improve diagnostic accuracy and efficiency. Figure 2 illustrates the basic workflow of the model, providing a clearer understanding of the operational mechanism of the proposed architecture.

The novelty of proposed methodology lies in its specialized architecture, which is meticulously crafted to capture the complex patterns and features inherent in brain tumor MRI scans. Unlike generic CNN models, custom CNN incorporates advanced layers and

TABLE 1 Summary of studies.

Study	Objective	Remarks
Hu and Razmjooy (2021)	The aim is to create a metaheuristic-based system for the early detection of brain tumors, utilizing automated procedures. The focus is on tumor segmentation, feature extraction, and classification, employing a deep belief network.	The proposed method incorporates an enhanced version of the seagull optimization algorithm for both feature selection and image classification
Sadad et al. (2021)	To develop an automated computer-assisted diagnosis system for early detection of tumors in brain, with a focus on segmentation, classification, and performance enhancement through preprocessing and data augmentation.	Evolutionary algorithms and reinforcement learning, along with transfer learning, are employed for multi-classification of brain tumors, showcasing a comprehensive approach to diagnosis.
Raza et al. (2022)	The aim was to introduce DeepTumorNet, a hybrid deep learning model designed for precise classification of three types of brain tumors (glioma, meningioma, and pituitary tumor). This model utilizes a modified GoogLeNet architecture and employs the leaky ReLU activation function.	DeepTumorNet utilizes a modified GoogLeNet architecture with 15 additional layers, enhancing the expressiveness of the model for feature extraction.
Gurunathan and Krishnan (2021)	The objective was to create an automated computer-aided method for detecting and locating brain tumors in MRI images. This method was utilizing deep learning algorithms and consist of three sub-modules: preprocessing, classification, and segmentation.	Morphological-based segmentation methodology is utilized for precise identification of tumor regions.
Methil (2021)	To develop a novel method using image preprocessing and a convolutional neural network (CNN) to detect brain tumors from diverse brain images.	The proposed method, combining histogram equalization and CNN, achieved impressive recall rates of 98.55% on the training set and 99.73% on the validation set, demonstrating its effectiveness in accurately detecting brain tumors across various shapes, sizes, textures, and locations.
Aggarwal et al. (2023)	The proposal aims to introduce an efficient method for brain tumor segmentation utilizing an Improved Residual Network (ResNet). This method addresses the gradient diffusion issue in Deep Neural Networks (DNN) and aims to enhance segmentation accuracy in MRI images.	The study highlights the potential of Improved ResNet in advancing brain tumor segmentation, with promising implications for medical diagnosis and treatment planning.
Xiong et al. (2021)	The goal was to create an FPGA-based accelerator for brain tumor segmentation. This aims to enhance segmentation speed, reduce computational complexity, and maintain high accuracy.	The FPGA-based accelerator presents a promising approach for automatic segmentation and remote diagnosis of brain tumors. This contributes to enhancing efficiency and accuracy in medical imaging analysis.
Kumar Sahoo et al. (2023)	The aim was to develop an intelligent system for automatically extracting and identifying brain tumors from 2D contrast-enhanced MRI images. This system was addressing challenges related to accurate diagnosis and the time-consuming nature of manual examination.	The YOLO2 based transfer learning approach achieves a high classification accuracy, further enhancing the diagnostic capability of the system.
Ge et al. (2020)	The objective was to tackle the challenge of glioma classification from MRI scans through a proposed deep semi-supervised learning framework. This framework integrates deep CNN features and a novel 3D-2D consistent constraint. Additionally, it leverages synthetic MRIs generated by Generative Adversarial Networks (GANs) to augment the training data.	The proposed scheme achieves promising results on two glioma datasets, demonstrating good performance in IDH-mutation prediction and glioma grading, with accuracies of 86.53 and 90.70% on TCGA and MICCAI datasets, respectively.
Taher et al. (2022)	The aim was to develop transfer-learning-based models and a Convolutional Neural Network (CNN) called BRAIN-TUMOR-net for classifying brain MRI images into tumor or normal cases. The performance of these models will be compared with pre-trained models (InceptionResNetv2, Inceptionv3, and ResNet50) and tested on publicly available datasets.	Transfer-learning-based models and BRAIN-TUMOR-net are introduced for classification, with BRAIN-TUMOR-net achieving the highest accuracy levels across different MRI datasets.
Khan et al. (2022)	The aim is to develop a hierarchical deep learning method using a convolutional neural network (CNN) to detect and classify brain tumors into glioma, meningioma, pituitary, and no-tumor categories.	The proposed HDL2BT system demonstrated high precision (92.13%) and a low miss rate (7.87%), outperforming previous methods in detecting and segmenting brain tumors, thus providing valuable clinical assistance to physicians.

(Continued)

TABLE 1 (Continued)

Study	Objective	Remarks
Mahmud et al. (2023)	To develop a convolutional neural network (CNN) architecture for efficient identification and classification of brain tumors using MRI images.	The proposed CNN model achieved an accuracy of 93.3%, an AUC of 98.43%, a recall of 91.19%, and a loss of 0.25, outperforming ResNet-50, VGG16, and Inception V3, indicating its reliability for early detection of brain tumors.
Chattopadhyay and Maitra (2022)	The proposed model aims to introduce a highly accurate automatic method using a convolutional neural network (CNN) to segment brain tumors from 2D MRI images.	The proposed CNN-based model achieved an impressive accuracy of 99.74%, surpassing existing methods and significantly aiding doctors in the accurate and timely detection of brain tumors from MRI images.
Bitto et al. (2023)	To identify brain tumors in MRI images using convolutional neural network designs and data preprocessing techniques to achieve competitive performance.	The study combines MRI-based image datasets, employs various data preprocessing techniques and image augmentation methods, and utilizes five pre-trained models to achieve high accuracy and precision in brain tumor identification, with ResNet-50 performing the best at 96.76% accuracy.
Gayathri et al. (2023)	Assess the effectiveness of the VGG-16 architecture, a Convolutional Neural Network (CNN) model, for accurate brain tumor detection through deep learning.	The fine-tuned VGG-16 model achieved a high accuracy of 94% after hyperparameter optimization, demonstrating strong sensitivity, specificity, precision, recall, and F1 scores compared to other techniques for brain tumor detection.

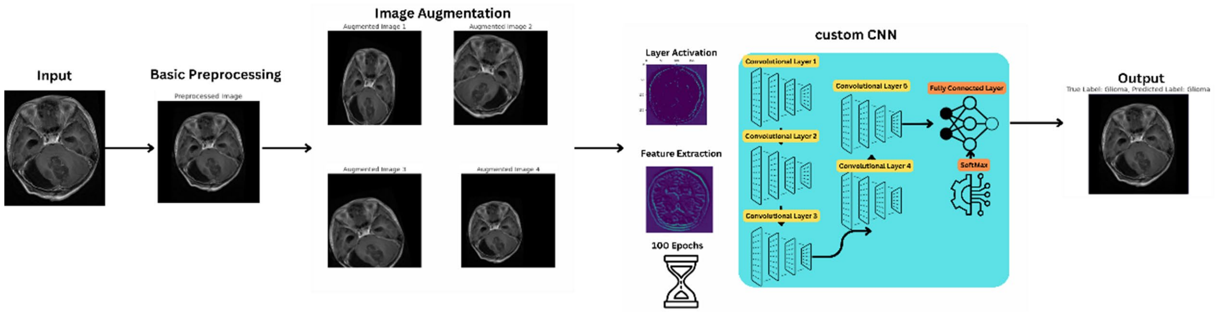


FIGURE 2
Workflow of the proposed model.

structures optimized for medical imaging, ensuring a deeper and more context-aware analysis. Its design considers the specific variations and characteristics of brain tumors, enabling the network to achieve high accuracy and reliability in tumor identification and categorization.

3.1 Dataset description

The dataset used for training and evaluating the proposed method consists of a comprehensive collection of brain MRI scans, carefully selected to encompass a diverse range of brain tumor types. This dataset includes images of glioma, meningioma, pituitary tumors, and non-tumorous brain tissue, ensuring that the model is exposed to a wide spectrum of tumor characteristics and variations. The dataset contained images of 512*512.

Sourced from a freely available medical imaging database, the dataset consists of several thousand MRI scans, each labeled with the corresponding tumor type or the absence of a tumor. The dataset's size and diversity are instrumental in training proposed method to recognize and differentiate between various brain tumor manifestations (Nickparvar, 2021).

The dataset utilized in this study comprises 1,621 images of gliomas, 1,645 images of meningiomas, 2000 images of pituitary tumors, and 1757 images representing non-tumorous tissues, ensuring a comprehensive representation of common brain tumor types. To address potential class imbalances, we employed stratified sampling to maintain a uniform distribution across training and validation sets.

In Table 2 a summary of the dataset has been given.

Alongside normalization, data augmentation techniques are employed on the dataset to bolster the robustness and generalizability of the proposed method. These techniques include rotations, translations, scaling, and flipping of the MRI images, creating variations that simulate different imaging conditions and perspectives. This augmentation process is crucial for preventing overfitting and ensuring that custom CNN maintains high performance across diverse and unseen MRI data. In Figure 3 images after resizing and applying the basic techniques are being shown.

By meticulously preparing and augmenting the dataset, the proposed method is equipped with a rich and varied foundation of MRI scans, enabling it to learn and generalize effectively, thereby demonstrating superior performance in brain tumor analysis.

3.2 Proposed architecture

The proposed method embodies a sophisticated convolutional neural network architecture meticulously crafted to tackle the intricate task of analyzing brain tumor MRI scans. Central to this architecture are a series of convolutional layers that progressively delve deeper into the MRI images, extracting a wide range of features from basic textures and edges to intricate patterns associated with various types of brain tumors. These layers play a crucial role in enabling the proposed method to discern and characterize the nuanced manifestations of brain tumors within the MRI scans. Each convolutional layer in the proposed method is followed by a non-linear activation function, such as the Rectified Linear Unit (ReLU). This function introduces the necessary non-linearity into the model, enabling it to capture and model the complex, non-linear relationships inherent in the MRI data (Albalawi et al., 2024). This capability is crucial for the network's capacity to learn and adapt to the varied presentations of brain tumors. To sharpen the model's focus on salient features and alleviate the computational load, pooling layers are incorporated into the architecture. These layers reduce the spatial dimensions of the feature maps while preserving essential information. Mathematically, the convolutional layer can be defined as follow in equation 1.

TABLE 2 Dataset description.

Type	Training	Testing
Glioma	1,321	300
Meningioma	1,339	306
No Tumor	1,595	405
Pituitary	1,457	300

$$\mathit{Convolution\ Operation}: a_{ij}^{l} = \sum_{m}\sum_{n} I_{m,n} \cdot K_{i-m,\,j-n}^{l} \tag{1}$$

where $\left(a_{ij}^{l}\right)$ is the activationat layer (l),(I) is the input image,and (K) is the kernel.

The ReLU activation function can be mathematically defined as equation 2 followed by maxpooling in equation 3, batch normalization in equation 4, dropout at equation 5, softmax at equation 6 and categorical cross entropy loss in equation 7.

$$f(x) = \max(0,x) \tag{2}$$

Where,

- x: Input value to the ReLU activation function.
- $f(x)$: Output value of the ReLU activation function, which is xx if xx is positive, and 0 otherwise

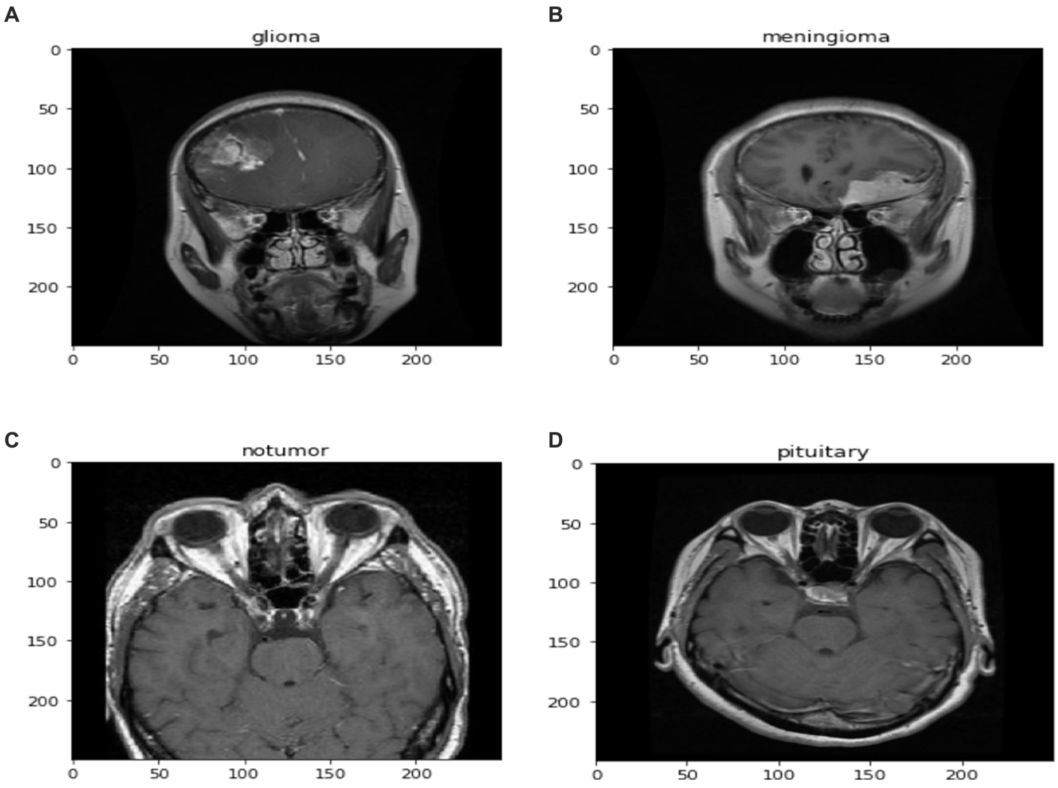


FIGURE 3
Images after resizing and applying basic pre-processing techniques.

$$a_{ij}^{l} = \max\left(\text{region from input}\right) \tag{3}$$

Where,

- a_{ij}^{l}: The output value of the max pooling operation at position (i,j) in the l-th layer.
- region from input: A specific region from the input feature map over which the max operation is performed. Typically, this region is defined by a pooling window.

$$y_i = \gamma\left(\frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}\right) + \beta \tag{4}$$

Where,

- x_i: Input value to the batch normalization layer.
- μ_B: Mean of the batch.
- $\sigma_B{}^2$: Variance of the batch.
- ϵ: Small constant added for numerical stability.
- γ: Scale parameter learned during training.
- β: Shift parameter learned during training.
- y_i: Output value of the batch normalization.

$$y_i = x_i \cdot d_i \tag{5}$$

$$Where\left(d_i \sim \text{Bernoulli}\left(p\right)\right),$$

- x_i: Input value to the dropout layer.
- d_i: Dropout mask value for the ii-th input, drawn from a Bernoulli distribution with probability pp.
- p: Probability of retaining a unit (i.e., not dropping it out).
- y_i: Output value after applying the dropout mask.

$$\sigma\left(z_i\right) = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{6}$$

Where,

- z_i: Input value to the softmax function for the ii-th class.
- $\sigma(z_i)$): Output probability of the ii-th class after applying the softmax function.
- $\sum_j e^{zj}$: Sum of exponentials of all input values for normalization.

$$L = -\sum_i y_i \log\left(p_i\right) \tag{7}$$

Where,

- y_i: Ground truth binary indicator (0 or 1) if class label ii is the correct classification for the observation.
- p_i: Predicted probability of the observation belonging to class ii (output from the softmax function).
- L: Categorical cross-entropy loss.

The network also integrates batch normalization, a technique that normalizes the inputs of each layer to enhance training stability and efficiency. This is particularly advantageous in expediting the training process and ensuring consistent performance across various training batches. To mitigate the risk of overfitting—a prevalent challenge in deep learning models, especially when handling complex medical imaging data—the custom CNN includes dropout layers. These layers randomly exclude a subset of features during training, forcing the network to learn more robust and generalized representations of the data.

As the network progresses, the extracted features are funneled into fully connected layers, which synthesize the high-level information gleaned from the MRI scans to facilitate the final classification task. The culmination of this architecture is a SoftMax output layer, providing a probabilistic interpretation of each tumor type, offering a clear and interpretable decision basis for clinicians. The proposed method is described further in Algorithm 1.

ALGORITHM 1: MRI brain tumor classification using CNN.

```
Input: MRI brain images dataset with four categories: glioma, meningioma, no tumor, and pituitary tumor.
Output: Classification of MRI images into one of the four categories.

1. Preprocessing:
   • Load the MRI brain images from the dataset.
   • Resize the images to 200×200 pixels for standardization.
2. Model Architecture:
   • Initialize a Sequential model.
   • Add six convolutional layers with ReLU activation:
      • First layer: 64 filters of size 7×7, padding = 'same', input shape (200, 200, 1).
      • Second layer: 128 filters of size 7×7, padding = 'same'.
      • Third layer: 128 filters of size 7×7, padding = 'same'.
      • Fourth layer: 256 filters of size 7×7, padding = 'same'.
      • Fifth layer: 256 filters of size 7×7, padding = 'same'.
      • Sixth layer: 512 filters of size 7×7, padding = 'same'.
   • After each convolutional layer, add a batch normalization layer and a max-pooling layer with pool size (2,2).
   • Flatten the output to feed into the fully connected layers.
   • Add two fully connected layers with ReLU activation, 1,024 and 512 neurons, respectively, each followed by a dropout layer with a dropout rate of 0.25.
   • Add an output layer with four neurons (corresponding to the four categories) with softmax activation.
3. Compilation:
   • Compile the model using the SGD optimizer with a learning rate of 0.001, loss function as 'categorical_crossentropy', and metric as 'categorical_accuracy'.
4. Data Augmentation:
   • Use ImageDataGenerator for real-time data augmentation, including rescaling and horizontal flipping.
5. Training:
   • Train the model on the training dataset using the flow_from_directory method with a batch size of 32 and 100 epochs, employing callbacks for early stopping and learning rate reduction on plateau.
6. Evaluation:
   • Evaluate the model on a separate test dataset.
   • Compute and plot the training and validation accuracy and loss over the epochs.
   • Generate a confusion matrix to evaluate the model's classification performance.
```

Custom CNN's training is meticulously orchestrated using advanced optimization techniques like Adam and SGD (Stochastic

Gradient Descent), which fine-tune the network's weights to minimize a carefully chosen loss function, typically categorical cross-entropy in multi-class classification scenarios. This loss function plays a crucial role in guiding the network's learning process, ensuring that the model's predictions closely align with the actual tumor classifications. The SGD update rule, adam update rule, learning rate decay, early stopping criterion, flattening, feature map size after convolution, feature map size after pooling and gradient computation can be mathematically represented by equations 8–15, respectively.

Equation (8): This represents the standard gradient descent update rule, where θ (the model parameters) are updated by subtracting the gradient of the loss function J(θ) with respect to θ, scaled by a learning rate η.

Equation (9): This is a component of the Adam optimization algorithm, where v_t and s_t are exponentially decaying moving averages of the gradient and its square, respectively. β_1 is a parameter controlling the exponential decay rates.

Equation (10): Another component of Adam, updating the squared gradients moving average.

Equation (11): The learning rate decay mechanism in Adam, which reduces the learning rate η over time.

Equation (12): A notation indicating flattening of a tensor, commonly used when transitioning from convolutional layers to fully connected layers in neural networks.

Equations (13) and (14): These formulas calculate the output size (W_{out}) of a convolutional layer given the input size (W_{in}), filter size (F), padding (P), and stride (S). They differ depending on whether padding is applied.

Equation (15): Simply represents the gradient of the loss function J with respect to the model parameters θ.

$$\theta = \theta - \eta \cdot \nabla_\theta J(\theta) \tag{8}$$

$$v_{t+1} = \beta_1 v_t + (1-\beta_1)\nabla_\theta J(\theta) \tag{9}$$

$$s_{t+1} = \beta_2 s_t + (1-\beta_2)\left(\nabla_\theta J(\theta)\right)^2 \tag{10}$$

$$\eta_{t+1} = \eta_t \cdot \text{decay\_rate} \tag{11}$$

$$\mathit{If}\left(\text{val\_loss}_{t+1} > \text{val\_loss}_t\right) \mathit{for}\,(n)\, \mathit{epochs}, \mathit{stop\ training}.$$

$$\left(a^{l+1} = \text{flatten}\left(a^l\right)\right) \tag{12}$$

$$W_{out} = \frac{W_{in} - F + 2P}{S} + 1(13) \tag{13}$$

$$W_{out} = \frac{W_{in} - F}{S} + 1 \tag{14}$$

$$\nabla_\theta J(\theta) = \frac{\partial J}{\partial \theta}(15) \tag{15}$$

In essence, the proposed method is a meticulously crafted network that merges deep learning innovations with domain-specific adaptations to excel in the realm of brain tumor MRI analysis. In Figure 4 a detailed visual about how different layers of the model extract features from the images is given.

Its architecture is not just a series of layers but a well-orchestrated symphony of components each playing a critical role in ensuring the network's effectiveness in diagnosing and classifying brain tumors with high precision and reliability. Through its advanced feature extraction capabilities, adaptability to various MRI modalities, and a design conducive to clinical interpretability, the proposed method stands as a pioneering tool poised to transform the landscape of medical imaging analysis.

3.3 Preprocessing and data augmentation

Prior to being fed into the custom CNN, the MRI images undergo a series of preprocessing steps to ensure they are in an optimal format for analysis. These steps are crucial for standardizing the input data, which helps in reducing model complexity and improving its learning efficiency. Initially, the MRI images are resized to a consistent dimension, balancing the need for detail retention and computational efficiency. This standardization is essential for the network to process images uniformly, regardless of their original resolution.

In the preprocessing phase, each MRI scan was resized to a uniform dimension of 200×200 pixels to standardize input size for the CNN. Pixel intensity values were then normalized to a range of 0 to 1 to mitigate variations in image brightness and contrast, which are prevalent across different MRI machines and scanning parameters. Additional steps included applying Gaussian smoothing filters to reduce image noise and enhance feature extraction by the CNN layers.

Normalization is another crucial preprocessing step in which the pixel intensity values of the MRI images are scaled to a standard range, typically between 0 and 1.This scaling is vital for stabilizing the network's training process, as it ensures that the model is not biased by variations in image brightness or contrast, which are common in medical images due to differences in scan protocols and equipment. By normalizing the images, proposed architecture can focus on learning the relevant features that indicate the presence and type of brain tumors, rather than being influenced by extraneous imaging artifacts. Table 3 presents the augmentation technique with the values.

Data augmentation is crucial for improving the resilience and adaptability of a given architecture. Considering the diverse nature of tumor characteristics and the limited availability of labeled MRI data, augmentation methods are utilized to effectively broaden the scope of the training dataset. These methods entail generating altered renditions of the training images by means of operations like rotation, scaling, and flipping. For instance, the images might be rotated by various angles or flipped horizontally or vertically to simulate different perspectives of tumor presentations. Scaling adjustments are also made to mimic variations in tumor size across different patients.

These augmented images help the network learn to recognize tumors from a broader range of angles and appearances, increasing its ability to generalize from the training data to new, unseen images. The augmentation process introduces a level of diversity to the training set that mimics the variability in proposed method which it will encounter in real-world clinical settings, thereby preparing it to perform accurately and reliably across a wide range of scenarios.

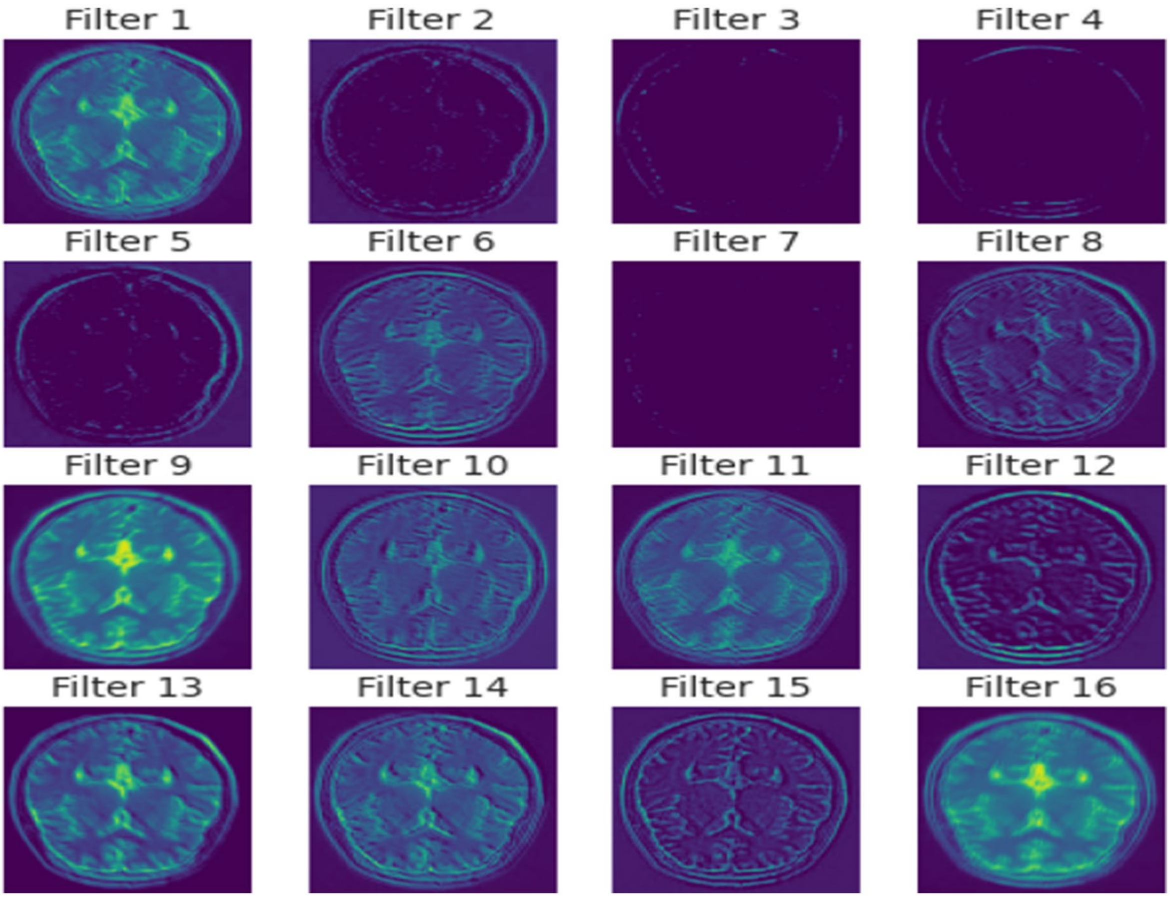


FIGURE 4
Filter-wise activation.

TABLE 3 Augmentation technique.

Augmentation technique	Value
rescale	1./255
featurewise_center	False
samplewise_center	False
featurewise_std_normalization	False
samplewise_std_normalization	False
zca_whitening	False
rotation_range	0
zoom_range	0
width_shift_range	0
height_shift_range	0
horizontal_flip	True
vertical_flip	False

Through this meticulous preprocessing and data augmentation, custom CNN is trained on a dataset that not only represents the complexity and variability of brain tumors but also reflects the diverse conditions under which clinical MRI scans are performed. This preparation is crucial for enabling the proposed method to effectively analyze MRI images of brain tumors, rendering it a strong and adaptable tool for assisting in the diagnosis and categorization of such tumors.

3.4 Training process

The training process of a custom CNN is a crucial phase where the network learns to accurately interpret and classify brain tumor MRI images. This process begins with a careful division of the available dataset into three distinct sets: training, validation, and testing. The training set, being the largest portion, is used to train the model and adjust the weights of the network. The validation set is utilized to fine-tune the model's hyperparameters and prevent overfitting by providing an independent evaluation of the model's performance during training. Finally, the testing set is used to assess the model's generalization capabilities on unseen data, ensuring that the performance metrics reflect the model's effectiveness in a real-world clinical setting. Table 4 presents the hyperparameter. The optimal value of each hyperparameter is chosen based on the continuous assessment of the code under different conditions.

During training, a specific loss function is employed to quantify the discrepancy between the predicted outputs and the actual labels. For a multi-class classification task such as brain tumor categorization,

TABLE 4 Hyperparameters.

Hyperparameter	Value
Monitor	'loss'
min_delta	1e-11
Patience	12
Verbose	1
Monitor	'val_loss'
Factor	0.2
Patience	6
Verbose	1
Monitor	'val_categorical_accuracy'
save_best_only	True
Verbose	1
steps_per_epoch	178
Epochs	100
validation_steps	40

categorical cross-entropy is typically chosen as the loss function due to its effectiveness in handling multiple classes. This function provides a measure of the model's prediction accuracy, guiding the network's adjustments to minimize errors during the training iterations.

Optimization of the network is achieved through sophisticated algorithms like Stochastic Gradient Descent (SGD) or Adam, which are instrumental in updating the model's weights and minimizing the loss function. These optimizers are selected based on their proven efficiency in navigating the complex landscape of high-dimensional weight space to find optimal values that minimize the loss. Figure 5 represents the model architecture with parameters of the model.

To bolster the model's generalization and prevent overfitting, several strategies are employed during the training process. Regularization techniques, such as L2 regularization, are incorporated to penalize large weights, encouraging the model to develop simpler, more general patterns that are robust to variations in the input data. Dropout is another crucial technique used, randomly deactivating a subset of neurons during training to compel the network to learn more distributed representations of the data, thus enhancing its generalization capabilities.

Furthermore, the training process involves periodic evaluations on the validation set to monitor the model's performance and make adjustments to the hyperparameters as necessary. This iterative evaluation helps in identifying the best model configuration that balances accuracy and generalizability, ensuring that proposed model performs optimally not just on the training data but also on new, unseen MRI images. Through this comprehensive and iterative training process, the model is finely tuned to excel in the complex task of classifying brain tumors from MRI scans, demonstrating its potential as a valuable tool in medical imaging analysis.

3.5 Model evaluation and validation

The evaluation and validation of model are pivotal stages in the development process, ensuring the model's efficacy and reliability in classifying brain tumors from MRI scans. These phases are designed to rigorously assess the model's performance using a range of metrics and benchmarks, providing insights into its accuracy, robustness, and clinical applicability (Mahesh et al., 2024).

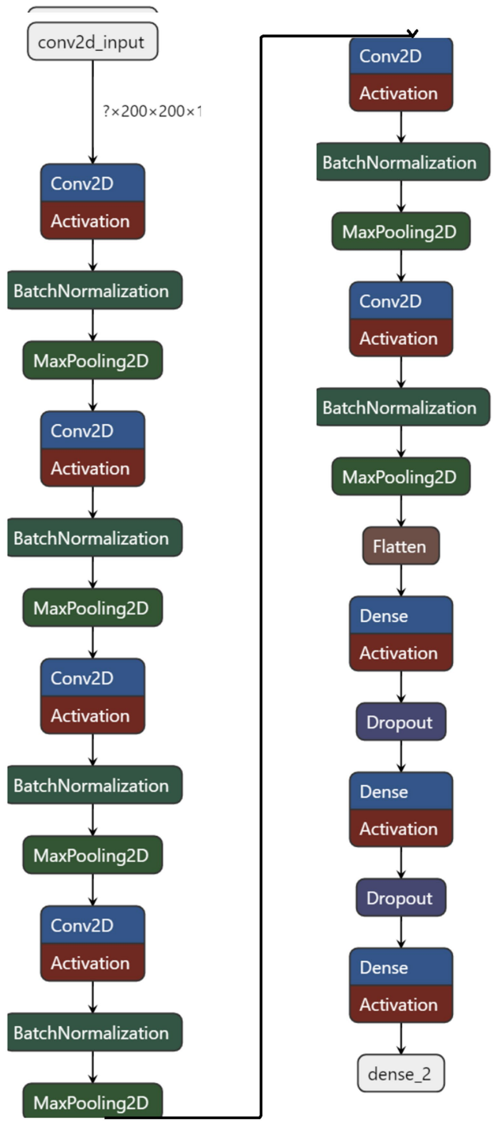


FIGURE 5
Model architecture.

3.5.1 Accuracy

These primary metric measures the proportion of correct predictions out of all predictions made, offering a straightforward assessment of the model's overall performance. It can be achieved by equation 16.

$$A = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \tag{16}$$

3.5.2 Precision and recall

Precision (the proportion of true positive results in all positive predictions) and recall (the proportion of true positive results in all actual positives) are crucial for understanding the model's performance in the context of each tumor type, especially in imbalanced datasets where some tumor types may be underrepresented. Precision and recall can be calculated using the following equations 17, 18.

$$P = \frac{TP}{TP + FP} \tag{17}$$

$$R = \frac{TP}{TP + FN} \tag{18}$$

where TP, FP, TN, and FN stand for True Positive, False Positive, True Negative, and False Negative, respectively.

3.5.3 F1 score

The F1 score combines precision and recall into a single metric by calculating their harmonic mean, providing a balanced view of the model's performance, particularly in scenarios where the cost of false positives and false negatives is significant. The F1 Score is calculated using the following equation 19.

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{19}$$

3.5.4 Area under the receiver operating characteristic curve (AUC-ROC)

This metric evaluates the model's ability to distinguish between classes at various threshold settings, which is particularly important for medical diagnosis where decision thresholds may vary based on clinical contexts. Additionally, the error metrics and advanced metrics like Mean Squared Error, Mean Absolute Error and F2 Score were calculated and they can be interpreted by equations 20–22, respectively.

Equation (20): This represents the Mean Squared Error (MSE), a commonly used metric for assessing the performance of regression models. It calculates the average squared difference between the actual values (Y_i) and the predicted values ($\hat{Y}_i$) over a dataset of size *n*.

Equation (21): This is the Mean Absolute Error (MAE), another metric for evaluating regression models. It computes the average absolute difference between the actual values (Y_i) and the predicted values ($\hat{Y}_i$).

Equation (22): This formula calculates the F-beta score, denoted as F2 in this case. It combines precision (P) and recall (R) into a single metric, with emphasis on recall. The value of beta determines the weight of recall in the calculation, where higher beta values place more importance on recall. In this case, beta is set to 2, giving more weight to recall.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 \tag{20}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| Y_i - \hat{Y}_i \right| \tag{21}$$

$$F2 = \left(1 + 2^2\right) \cdot \frac{P \cdot R}{\left(2^2 \cdot P\right) + R} \tag{22}$$

Model's performance is benchmarked against established models and industry standards to ascertain its effectiveness and advancement in brain tumor MRI analysis. These comparisons help in contextualizing Model's performance within the broader landscape of medical imaging AI.

Benchmarking involves comparing model's performance metrics with those from previous studies or conventional methodologies in brain tumor diagnosis. Such comparative analysis not only highlights the improvements but also identifies areas where model may require further enhancement.

By employing these rigorous evaluation and validation methods, the effectiveness of model in classifying brain tumors is thoroughly assessed, ensuring that the model is not only statistically sound but also practically significant in a clinical setting. This comprehensive evaluation framework underpins the model's potential to serve as a reliable and robust tool in enhancing the accuracy and efficiency of brain tumor diagnostics.

4 Experimentation and results

Experimentation and results sections delves into the different metrics result on which model is evaluated along with it the comparison with the existing is provided.

The experimental setup for assessing the proposed model involved an extensive training and validation regimen using a dataset comprising 7,023 MRI images categorized into four groups: glioma, meningioma, no tumor, and pituitary tumors. These images underwent preprocessing to standardize their dimensions to 200×200 pixels and conversion to grayscale, which simplified the input while preserving crucial structural details essential for accurate classification.

The model underwent training utilizing a stochastic gradient descent optimizer with a learning rate set at 0.001, with the objective of minimizing the categorical cross-entropy loss function—a suitable choice for tasks involving multi-class classification. Figure 6 illustrates the epoch-wise accuracy and loss of the proposed CNN model, while Figure 7 depicts the epoch-wise rate of accuracy improvement. Additionally, Figure 8 presents the learning rate schedule employed in the training process.

During training, early stopping mechanisms, learning rate reduction on plateau, and model checkpointing were employed to enhance training efficiency and prevent overfitting. The training process spanned multiple epochs, during which the dataset was partitioned into distinct training, validation, and testing sets. This division ensured thorough evaluation and validation of the model's performance, as well as its ability to generalize to unseen data.

4.1 Results presentation

The model exhibited exceptional performance metrics when evaluated on the testing set, reflecting its robustness and effectiveness in distinguishing brain tumors from MRI scans. Its achieved accuracy was notably high, reaching a rate of 99%, demonstrating its capability to accurately identify and categorize the vast majority of cases.

The precision for detecting glioma was perfect at 1.00, with a recall of 0.96, indicating a high true positive rate and few false negatives. The model displayed strong predictive power and sensitivity specifically

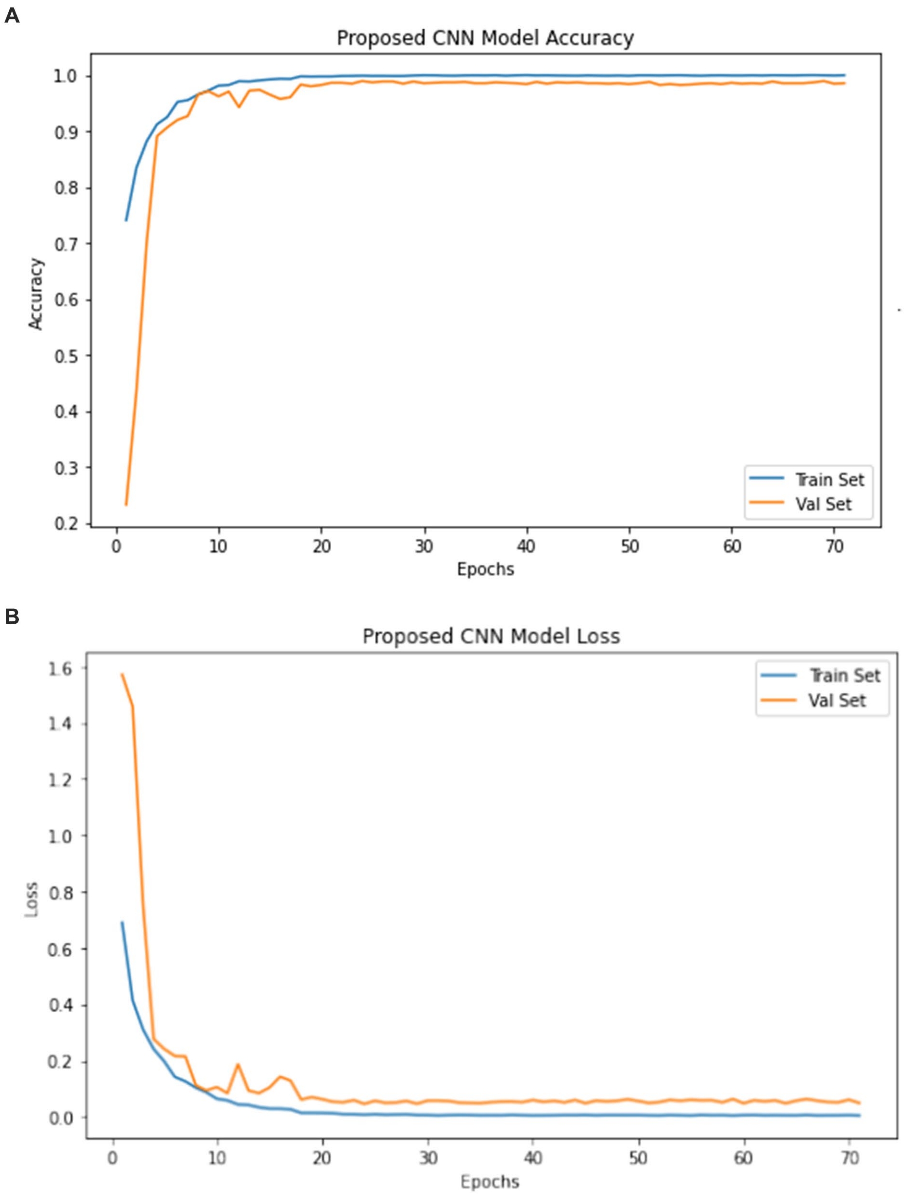


FIGURE 6
Epoch wise accuracy and lose.

for meningioma, with a precision of 0.96 and a recall of 0.98. These metrics highlight the model's ability to accurately identify and classify cases of meningioma, emphasizing its effectiveness in this particular category. The precision and recall for notumor and pituitary cases were equally impressive, showcasing the model's comprehensive learning and classification capabilities across various tumor types. The class 0, 1, 2, and 3 represents Glioma, Meningioma, No Tumor and Pituitary, respectively.

To highlight the advantages of our CNN model, we compared its performance against several established methods in brain tumor classification. For instance, traditional machine learning techniques such as SVM and Random Forests, though useful, lack the dynamic feature-learning capability that deep learning offers. Recent models like AlexNet and VGG, while deeper, still suffer from overfitting and require extensive labeled datasets. Our model's use of advanced regularization and data augmentation strategies positions it favorably against these methods, demonstrating superior accuracy and generalization in our tests.

Table 5 provides a comprehensive summary of the classification report, detailing various performance metrics such as precision, recall, and F1-score for each class.

Figure 9 gives a visual representation of normalized confusion matrix followed by precision-recall curve and roc-auc curve in Figures 10, 11 respectively.

In terms of error metrics, the model demonstrated low mean squared error (MSE) and mean absolute error (MAE), along with a high F2 score, underscoring its precision and reliability in prediction. The MSE of 0.026 indicates a small average squared difference between estimated values and actual values. Additionally, the MAE of 0.0168 represents the model's average absolute error across all predictions.

The F2 score, which strikes a balance between precision and recall, was exceptionally high at 0.986 This high F2 score underscores the model's effectiveness in classifying brain tumors, with a particular emphasis on minimizing false negatives—a critical consideration in medical diagnosis contexts. Figure 12 represents the error metrics.

4.2 Comparison with baseline models

When compared to traditional methods or earlier CNN-based models, custom CNN's performance stands out significantly. Traditional machine learning models or shallow CNNs typically achieve lower accuracy and precision metrics due to their limited feature extraction and learning capabilities. In contrast, proposed model's advanced architecture and training regimen have propelled its performance metrics well beyond these baseline models, demonstrating the effectiveness of its deep learning approach in medical image analysis. In Table 6 a comparative analysis between the previous methodology and the proposed methodology has been given.

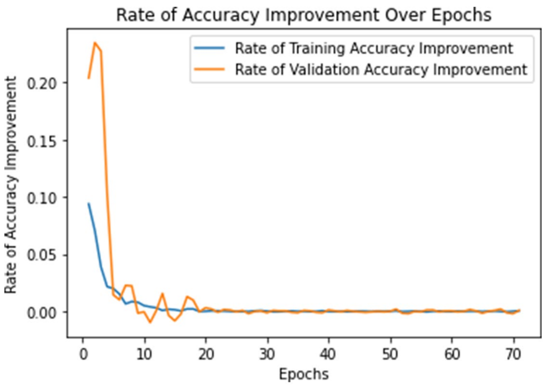


FIGURE 7
Rate of accuracy improvement over epochs.

The custom architecture and training strategy employed in the proposed model, combined with its remarkable performance metrics, highlight its potential to establish a new standard in the domain of brain tumor classification from MRI scans. The model's capacity to achieve high accuracy, alongside detailed metrics for different tumor types, offers robust quantitative evidence supporting its adoption and further investigation in clinical environments.

4.3 Ablation study

In the ablation study conducted to assess the robustness and significance of each layer within our brain tumor classification model, we systematically eliminated layers and documented the resulting effects on model performance. Initial results indicated a moderate degree of robustness, with the overall accuracy slightly declining from 0.89 after removing one layer to 0.92 after removing up to four layers. The precision, recall, and F1-score for each tumor type demonstrated only minor fluctuations, suggesting that the model preserves its discriminatory power up to a certain depth. However, a stark degradation was observed when all layers were removed, plummeting the overall accuracy to 0.69. This highlights the layers' collective importance in achieving high diagnostic accuracy. Conversely, the proposed model, which integrates all layers, displayed exceptional performance, achieving near-perfect precision and recall across all categories and culminating in an exemplary overall accuracy of 0.99. The comparison between the layer-ablated versions and the complete model underscores the intricate balance between model depth and performance, as summarized in the following Table 7.

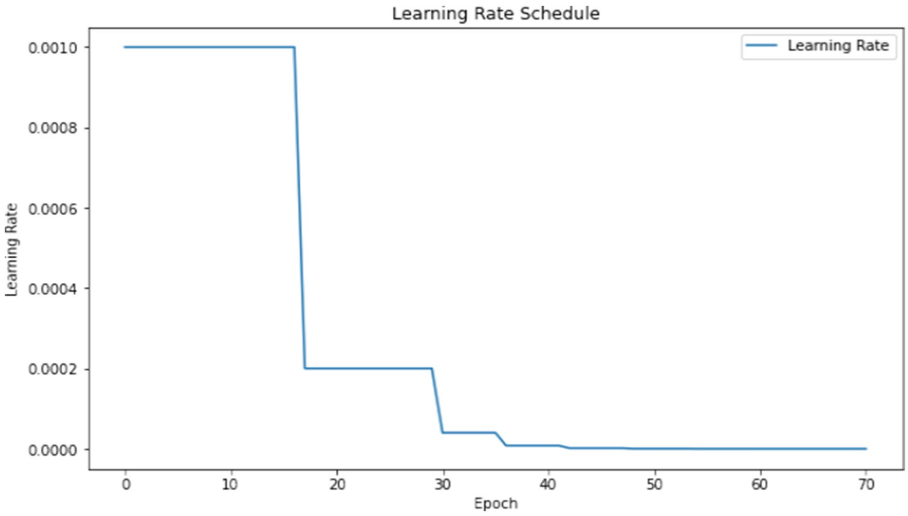


FIGURE 8
Learning rate schedule.

TABLE 5 Summary of classification report.

Type	Precision	Recall	F1-Score
Glioma	1	0.96	0.98
Meningioma	0.96	0.98	0.97
No tumor	1	1	1
Pituitary	0.99	1	0.99

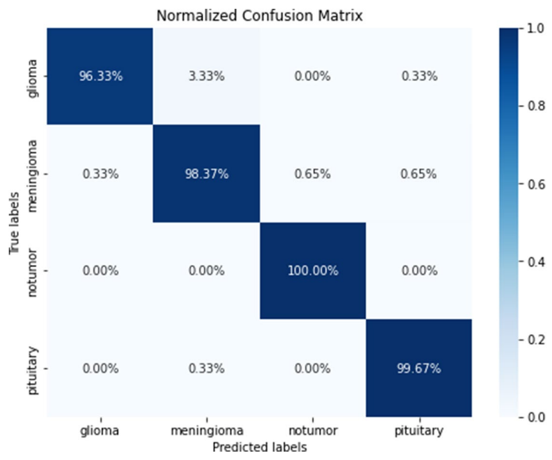


FIGURE 9
Normalized confusion matrix.

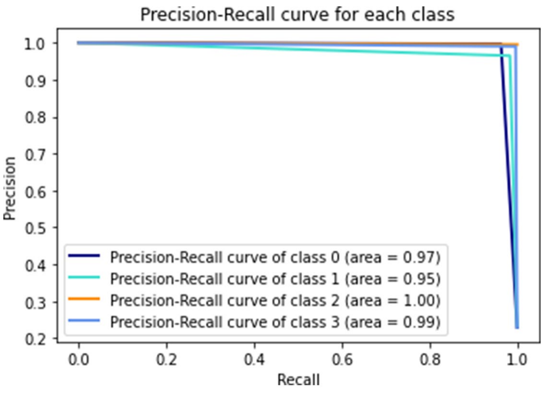


FIGURE 10
Precision-recall curve.

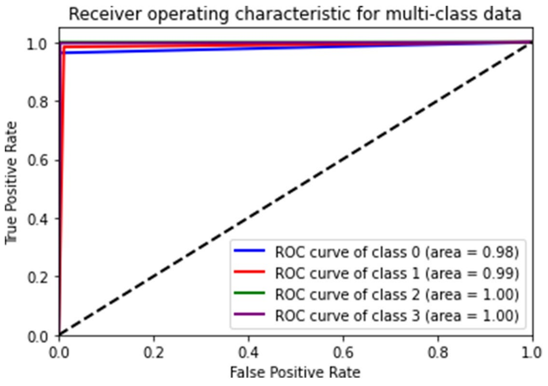


FIGURE 11
ROC-AUC score.

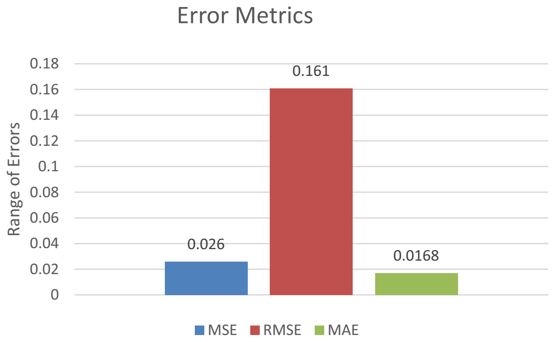


FIGURE 12
Error metrics.

The ablation study conducted to assess the robustness and significance of each layer within our brain tumor classification model revealed insightful findings. We systematically removed layers and observed the resulting effects on model performance. Interestingly, the model displayed a moderate degree of robustness, with only minor fluctuations in precision, recall, and F1-score when one to four layers were eliminated. However, a significant drop in accuracy was observed when all layers were removed, highlighting the collective importance of the layers in achieving high diagnostic accuracy. Conversely, the proposed model, which integrated all layers, exhibited exceptional performance, with near-perfect precision and recall across all categories and an exemplary overall accuracy of 0.99. This comparison underscores the delicate balance between model depth and performance, emphasizing the critical role of each layer in optimizing classification outcomes.

The comprehensive model clearly demonstrates the necessity of each layer, offering a robust framework for accurate brain tumor classification.

5 Discussion

The outcomes yielded by the proposed model are highly encouraging, signifying a notable advancement in leveraging convolutional neural networks for analyzing brain tumor MRI scans. With an accuracy rate of 99%, the model demonstrates exceptional proficiency in distinguishing between various types of brain tumors, as well as accurately identifying non-tumor regions within the brain. Such elevated accuracy holds immense importance in medical diagnostics, where the repercussions of false positives or negatives can be significant.

Moreover, the precision and recall metrics across different tumor types offer a nuanced insight into the model's performance. The high precision observed for glioma and meningioma indicates that when the model predicts these tumor types, it does so with high reliability. Similarly, the high recall rates indicate the model's effectiveness in

TABLE 6 Comparative analysis with the proposed model.

Study	Technique	Accuracy
Pedada et al. (2023)	U-Net Model for Brats 2017 and 2018 dataset segmentation	93.40 and 92.20%
Saeedi et al. (2023)	2D CNN with ensemble machine learning techniques	96.47%
Mahmud et al. (2023)	Redefined CNN Model with modified classification	93.3%
Wang et al. (2022)	Deep CNN on OCT Images	94.90%
Vidyarthi et al. (2022)	CNN with NN Classifier	95.86%
Lamrani et al. (2022)	CNN with Enhanced Classifiers	96%
Yildirim et al. (2023)	Convolutional Neural Network (CNN)-based hybrid model	95.4%
Bacak et al. (2023)	Convolutional Neural Network (CNN) using TensorFlow	90%
Khan et al. (2022)	Deep Learning Models (Convolutional Neural Networks - CNN)	Up to 97.8%
Gómez-Guzmán et al. (2023)	Evaluation of deep convolutional neural network (CNN) models for brain tumor classification	97.12%
Sharma and Shukla (2022)	CNN for brain tumor classification	93.38%
Nayak et al. (2022)	Efficient net on T1 Weighted Data	98.78%
Guan et al. (2021)	Agglomerative Clustering Based Approach	98.04%
Rajput et al. (2024)	VGG19, Inception-v3, and ResNet50	90%
Suryawanshi and Patil (2024)	Convolutional Neural Network (CNN) & VGG19	98.01%
Prasad et al. (2024)	CNN	98.93%
Schiavon et al. (2023)	convolutional neural networks (CNNs)	96%
Rasool et al. (2022)	CNN & SqueezeNet	96.5%
Sarada et al. (2024)	ResNet0V2	96.34%
Rahman and Islam (2023)	Parallel deep convolutional neural networks	98%
Özkaraca et al. (2023)	VGG 16 and basic CNN Architecture	97%
Prabha et al. (2023)	Efficeint Net model Using transfer learning	98.27%
Proposed model	Custom CNN with advance layer arrangement.	99%

identifying the majority of actual cases for each tumor type, reducing the risk of missed diagnoses.

The F2 score, which emphasizes the importance of recall (minimizing false negatives), is particularly relevant in a medical context (Zhou et al., 2023). A high F2 score, as achieved by proposed model, underscores the model's capability in correctly identifying positive cases, a critical aspect when early detection can significantly influence treatment outcomes.

Proposed research introduces several innovative elements to the domain of medical imaging, particularly in how deep learning can be tailored to enhance diagnostic precision. The network architecture's design, which integrates deep convolutional layers with advanced regularization and normalization techniques, is specifically optimized for the complex task of brain tumor identification and classification. This bespoke approach, which diverges from the application of generic CNN models, is a significant contributor to the model's success.

The impact of these findings extends beyond the technical domain, potentially revolutionizing how brain tumors are diagnosed and classified in clinical settings. By providing a tool that can rapidly and accurately analyze MRI scans, proposed model could assist radiologists in making more informed decisions, facilitating early and accurate diagnoses, and ultimately improving patient care and outcomes.

The model's performance, while tested on a robust dataset, might still be limited by the diversity and volume of the data available. Real-world applicability will require continual testing and validation on a broader array of MRI scans from diverse patient demographics and equipment.

Although the model achieves high accuracy, it's important to acknowledge the potential limitation posed by the "black box" nature of deep neural networks (Salahuddin et al., 2022) Integrating attention mechanisms or employing explainable AI frameworks could significantly enhance the interpretability of the proposed model, thereby increasing its clinical utility. These techniques offer insights into the model's decision-making process, providing clinicians with a deeper understanding of how and why specific diagnoses are made (Jiang et al., 2023). By elucidating the rationale behind the model's predictions, these methods can improve trust and confidence in its outputs, ultimately facilitating more informed clinical decision-making. The current version of proposed model is optimized for a specific MRI dataset. Its ability to generalize across different MRI machines and imaging modalities remains to be thoroughly tested. Future work could focus on expanding the model's adaptability to various imaging conditions, enhancing its robustness and applicability (Chaudhary et al., 2024). The model primarily focuses on cross-sectional MRI data. Incorporating longitudinal and multi-modal imaging data, such as merging MRI with CT or PET scans, has the potential to offer a more holistic understanding of tumor features and development, thereby enriching diagnostic capabilities (Sharma and Chaudhary, 2023).

TABLE 7 Ablation study.

		Precision	Recall	F1-Score
After removing 1 layer	Glioma	0.98	0.73	0.84
	Meningioma	0.77	0.8	0.79
	No tumor	0.91	0.99	0.95
	Pituitary	0.91	1	0.95
	Overall accuracy = 0.89			
After removing 2 layers	Glioma	0.97	0.76	0.85
	Meningioma	0.79	0.87	0.83
	No tumor	0.93	0.99	0.96
	Pituitary	0.96	0.99	0.98
	Overall accuracy = 0.91			
After removing 3 layers	Glioma	0.96	0.81	0.88
	Meningioma	0.82	0.87	0.84
	No tumor	0.93	0.99	0.96
	Pituitary	0.97	0.99	0.98
	Overall accuracy = 0.92			
After removing 4 layers	Glioma	0.96	0.82	0.89
	Meningioma	0.82	0.85	0.83
	No tumor	0.92	0.99	0.96
	Pituitary	0.97	0.99	0.97
	Overall accuracy = 0.92			
After removing all layers	Glioma	0.64	0.53	0.58
	Meningioma	0.52	0.37	0.43
	No tumor	0.79	0.9	0.84
	Pituitary	0.72	0.91	0.8
	Overall accuracy = 0.69			
Proposed model	Glioma	1	0.96	0.98
	Meningioma	0.96	0.98	0.97
	No Tumor	1	1	1
	Pituitary	0.99	1	0.99
	Overall accuracy = 0.99			

In Figure 13 one instance which was misclassified has been given followed by correct predictions in Figure 14.

The proposed model demonstrates significant advancements in brain tumor MRI analysis, a conscious effort to address these limitations through continuous research and iterative refinement will be essential. Enhancing data diversity, interpretability, and cross-modality generalization will be crucial steps in evolving proposed model from a promising model to a reliable tool in clinical practice.

However, there are major limitations to consider. The model's performance, while tested on a robust dataset, might be limited by data diversity and volume. Real-world applicability will require validation on a broader array of MRI scans from diverse demographics and equipment. Additionally, the "black box" nature of deep neural networks poses interpretability challenges. Integrating explainable AI techniques could enhance the model's transparency and clinical utility. Future work should focus on enhancing data diversity, interpretability, and cross-modality generalization, along with extensive clinical validation for integration into clinical workflows. Furthermore, exploring the integration of multimodal imaging data and adapting the model to different populations or tumor types represents promising directions for future research.

One of the significant challenges in enhancing the generalization of datasets for brain tumor classification using MRI scans is the diversity and variability inherent in medical imaging data. MRI scans can vary widely in terms of imaging protocols, machine calibration, and patient demographics, all of which can influence the appearance of the images and, consequently, the performance of classification models. Additionally, the limited availability of labeled medical images due to privacy concerns and the cost of expert annotation poses a challenge for training robust models.

To make the dataset more generalized and comprehensive, it is crucial to include a broader array of MRI scans from diverse populations and multiple healthcare settings. Incorporating images from different MRI machines and including variations in scan settings can help the model learn to recognize tumors across different imaging conditions. Extending the dataset to include multi-modal imaging data, such as combining MRI scans with CT or PET scans, can enrich the dataset and provide more comprehensive features for the model to learn. This approach can improve diagnostic accuracy and help the model generalize better to new, unseen cases. Furthermore, synthetic data generation techniques like Generative Adversarial Networks (GANs) can be employed to augment the dataset, providing a wider array of training examples without compromising patient privacy. These strategies would enhance the model's robustness and its applicability in diverse clinical environments.

The integration of our CNN model into clinical workflows could significantly enhance the diagnostic process by providing rapid, preliminary analysis of MRI scans. This tool could serve as a second opinion to assist radiologists in identifying subtle or ambiguous tumor signs, potentially speeding up the diagnosis and reducing the likelihood of human error (Chaudhary and Agrawal, 2021). Challenges for integration include the need for extensive clinical validation to ensure accuracy and reliability, as well as adjustments to existing medical software systems to accommodate the new AI capabilities.

Future research could explore the integration of multimodal imaging data, combining MRI with CT or PET scans to enrich the dataset and potentially improve diagnostic accuracy. Additionally, further studies could focus on adapting the model to different populations or other types of tumors, enhancing its applicability. Another promising direction is the incorporation of explainable AI techniques to provide insights into the model's decision-making processes, increasing its transparency and trustworthiness for clinical use.

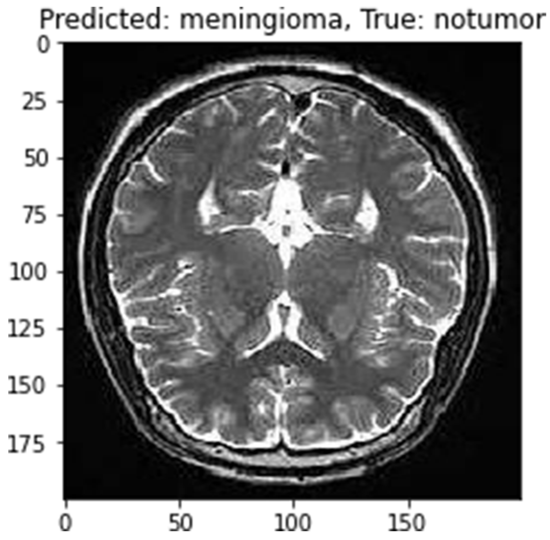


FIGURE 13
Misclassified instances.

6 Conclusion

This study advances the application of CNNs in the classification of brain tumors from MRI scans, demonstrating a significant improvement over existing methods. The customized CNN architecture introduced novel aspects tailored specifically for medical imaging, setting a new benchmark for accuracy and efficiency in this field. Tailored specifically for the nuanced task of brain tumor classification, proposed method demonstrated an impressive 99% accuracy rate in proposed study, alongside high precision and recall across various tumor categories, underscoring its potential as a robust diagnostic aid in clinical settings. The implications of these discoveries are significant for the realm of medical imaging and diagnostics. The capability of the model to precisely classify brain tumors from MRI scans has the potential to transform diagnostic procedures, leading to heightened accuracy, shortened analysis durations, and potentially better patient outcomes by enabling earlier and more accurate diagnoses. This research emphasizes the value of developing specialized, task-specific AI models for medical imaging, which can address the unique challenges of the field more effectively than general-purpose models.

Looking ahead, there are several promising directions for future research. Expanding the diversity of the training and validation datasets can enhance model's generalizability and robustness. Improving the model's interpretability would make it more valuable in clinical contexts, where understanding the basis for its predictions is crucial. Extending its capabilities to multi-modal and longitudinal analyses could offer deeper insights into tumor progression and response to treatment. Finally, rigorous clinical validation and integration into clinical workflows will be essential steps toward realizing proposed model's potential to improve diagnostic practices and patient care in the realm of brain tumor treatment. By pursuing these avenues, we can build on the solid foundation laid by this study to further advance the application of AI in medical diagnostics, ultimately contributing to better health outcomes and enhanced clinical decision-making.

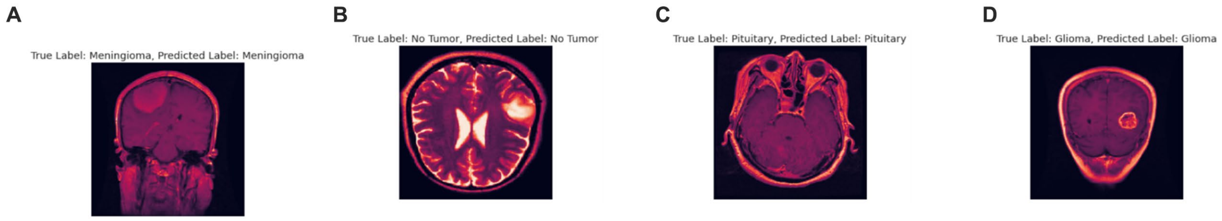


FIGURE 14
Correct predictions.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

EA: Conceptualization, Investigation, Writing – original draft. AT: Data curation, Methodology, Software, Writing – original draft. DD: Formal analysis, Supervision, Writing – review & editing. SB: Conceptualization, Writing – original draft. TM: Resources, Visualization, Writing – review & editing. AA: Funding acquisition, Validation, Writing – review & editing. KA: Formal analysis, Project administration, Writing – review & editing. MA: Formal analysis, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R432), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Acknowledgments

This research is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R432), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This work was supported by the Deanship of Scientific Research, Vice President for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [GrantA376].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aboussaleh, I., Riffi, J., Fazazy, K. E., Mahraz, M. A., and Tairi, H. (2023). Efficient U-net architecture with multiple encoders and attention mechanism decoders for brain tumor segmentation. *Diagnostics* 13:872. doi: 10.3390/diagnostics13050872

Aggarwal, M., Tiwari, A. K., Sarathi, M. P., and Bijalwan, A. (2023). An early detection and segmentation of brain tumor using deep neural network. *BMC Med. Inform. Decis. Mak.* 23:78. doi: 10.1186/s12911-023-02174-8

Ahmmed, S., Podder, P., Mondal, M., Rahman, S., Kannan, S., Hasan, M., et al. (2023). Enhancing brain tumor classification with transfer learning across multiple classes: an in-depth analysis. *BioMedInformatics* 3, 1124–1144. doi: 10.3390/biomedinformatics 3040068

Albalawi, , Thakur, A., Ramakrishna, M. T., Bhatia Khan, S., SankaraNarayanan, S., Almarri, B., et al. (2024). Oral squamous cell carcinoma detection using EfficientNet on histopathological images. *Front. Med.* 10:1349336:1349336. doi: 10.3389/fmed.2023.1349336

Alzubaidi, L., Zhang, J., Humaidi, A. J., al-Dujaili, A., Duan, Y., al-Shamma, O., et al. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* 8:53. doi: 10.1186/s40537-021-00444-8

Bacak, A., Şenel, M., and GünaY, O. (2023). Convolutional neural network (CNN) prediction on meningioma, glioma with Tensorflow. *Int. J. Comput. Experiment. Sci. Engin.* 9, 197–204. doi: 10.22399/ijcesen.1306025

Bairagi, V. K., Gumaste, P. P., Rajput, S. H., and Chethan, K. S. (2023). Automatic brain tumor detection using CNN transfer learning approach. *Med. Biol. Eng. Comput.* 61, 1821–1836. doi: 10.1007/s11517-023-02820-3

Bitto, A. K., Bijoy, M. H. I., Yesmin, S., Mahmud, I., Mia, M. J., Biplob, B., et al. (2023). Tumor-net: convolutional neural network modeling for classifying brain tumors from MRI images. *Int. J. Advanc. Intelligent Inform.* 9:148. doi: 10.26555/ijain.v9i2.872

Chattopadhyay, A., and Maitra, M. (2022). MRI-based brain tumour image detection using CNN based deep learning method. *Neuroscience Inform.* 2:100060. doi: 10.1016/j.neuri.2022.100060

Chaudhary, P., and Agrawal, R. (2021). Sensory motor imagery EEG classification based on non-dyadic wavelets using dynamic weighted majority ensemble classification. *Intelligent Decision Technol.* 15, 33–43. doi: 10.3233/IDT-200005

Chaudhary, P., Varshney, Y. V., Srivastava, G., and Bhatia, S. (2024). Motor imagery classification using sparse nonnegative matrix factorization and convolutional neural networks. *Neural Comput. & Applic.* 36, 213–223. doi: 10.1007/s00521-022-07861-7

Dufumier, B., Gori, P., Battaglia, I., Victor, J., Grigis, A., Duchesnay, E. (2021). Benchmarking CNN on 3D anatomical brain MRI: architectures, data augmentation and deep ensemble learning. *arXiv preprint arXiv:2106.01132.*

Gayathri, P., Dhavileswarapu, A., Ibrahim, S., Paul, R., and Gupta, R. (2023). Exploring the potential of vgg-16 architecture for accurate brain tumor detection using deep learning. *J. Comput. Mechanic. Manag.* 2, –23056. doi: 10.57159/gadl.jcmm.2.2.23056

Ge, I. Y.-H. G., Jakola, A. S., and Yang, J. (2020). Deep semi-supervised learning for brain tumor classification. *BMC Med. Imaging* 20:87. doi: 10.1186/s12880-020-00485-0

Gómez-Guzmán, M. A., Jiménez-Beristaín, L., García-Guerrero, E. E., López-Bonilla, O. R., Tamayo-Perez, U. J., Esqueda-Elizondo, J. J., et al. (2023). Classifying brain tumors on magnetic resonance imaging by using convolutional neural networks. *Electronics* 12:955. doi: 10.3390/electronics12040955

Guan, Y, Aamir, M, Rahman, Z, Ali, A, Abro, WA, Dayo, ZA, et al. (2021). A framework for efficient brain tumor classification using MRI images. *Math Biosci Eng.* 18, 5790–5815. doi: 10.3934/mbe.2021292

Gurunathan, A., and Krishnan, B. (2021). Detection and diagnosis of brain tumors using deep learning convolutional neural networks. *Int. J. Imaging Syst. Technol.* 31, 1174–1184. doi: 10.1002/ima.22532

Hu, A., and Razmjooy, N. (2021). Brain tumor diagnosis based on metaheuristics and deep learning. *Int. J. Imaging Syst. Technol.* 31, 657–669. doi: 10.1002/ima.22495

Islam, M., Reza, M. T., Kaosar, M., and Parvez, M. Z. (2023). Effectiveness of federated learning and CNN ensemble architectures for identifying brain tumors using MRI images. *Neural. Process. Lett.* 55, 3779–3809. doi: 10.1007/s11063-022-11014-1

Jiang, X., Hu, Z., Wang, S., and Zhang, Y. (2023). Deep learning for medical image-based Cancer diagnosis. *Cancers* 15:3608. doi: 10.3390/cancers15143608

Khan, A. H., Abbas, S., Khan, M. A., Farooq, U., Khan, W. A., Siddiqui, S. Y., et al. (2022). Intelligent model for brain tumor identification using deep learning. *Appl. Comput. Intelligence Soft Comput.* 2022, 1–10. doi: 10.1155/2022/8104054

Khan, M. S. I., Rahman, A., Debnath, T., Karim, M. R., Nasir, M. K., Band, S. S., et al. (2022). Accurate brain tumor detection using deep convolutional neural network. *Comput. Struct. Biotechnol. J.* 20, 4733–4745. doi: 10.1016/j.csbj.2022.08.039

Kumar, S., and Kumar, D. (2023). Human brain tumor classification and segmentation using CNN. *Multimed. Tools Appl.* 82, 7599–7620. doi: 10.1007/s11042-022-13713-2

Kumar Sahoo, P., Parida, K. M., and Dash, S. (2023). Efficient simultaneous segmentation and classification of brain tumors from MRI scans using deep learning. *Biocybernet. Biomed. Engin.* 43, 616–633. doi: 10.1016/j.bbe.2023.08.003

Lamrani, B., Cherradi, O. E., Gannour, M. A., Bouqentar, M. A., and Bahatti, L. (2022). Brain tumor detection using MRI images and convolutional neural network. *Int. J. Adv. Comput. Sci. Appl.* 13. doi: 10.14569/ijacsa.2022.0130755

Mahmud, M. I., Mamun, M., and Abdelgawad, A. (2023). A deep analysis of brain tumor detection from mr images using deep learning networks. *Algorithms* 16:176. doi: 10.3390/a16040176

Methil, A. S. (2021). "Brain tumor detection using deep learning and image processing" in *2021 international conference on artificial intelligence and smart systems (ICAIS)* (IEEE), 100–108.

Nayak, D. R., Padhy, N., Mallick, P. K., Zymbler, M., and Kumar, S. (2022). Brain tumor classification using dense efficient-net. *Axioms* 11:34. doi: 10.3390/axioms11010034

Nickparvar, M., "Brain tumor MRI dataset," Kaggle, (2021). *[pre print]* Available at: https://www.kaggle.com/dsv/2645886

Özkaraca, O., Bağrıaçık, O. İ., Gürüler, H., Khan, F., Hussain, J., Khan, J., et al. (2023). Multiple brain tumor classification with dense CNN architecture using brain MRI images. *Life* 13:349. doi: 10.3390/life13020349

Pedada, K. R., Bhujanga Rao, A., Patro, K. K., Allam, J. P., Jamjoom, M. M., and Samee, N. A. (2023). A novel approach for brain tumour detection using deep learning based technique. *Biomed. Signal Process. Control* 82:104549. doi: 10.1016/j.bspc.2022.104549

Prabha, P. L., Kulshreshtha, A., and Patel, S. (2023). "Automated brain tumor detection from MRI images using transfer learning techniques" in *AIP conference proceedings*, vol. *2603* (AIP Publishing).

Prasad, C. R., Srividya, K., Jahnavi, K., Srivarsha, T., Kollem, S., and Yelabaka, S. (2024). "Comprehensive CNN model for brain tumour identification and classification using MRI images" in *In 2024 IEEE international conference for women in innovation, Technology & Entrepreneurship (ICWITE)* (IEEE), 524–528.

Rahman, T., and Islam, M. S. (2023). MRI brain tumor detection and classification using parallel deep convolutional neural networks. *Measurement [pre print]* 26:100694.

Rajput, I. S., Gupta, A., Jain, V., and Tyagi, S. (2024). A transfer learning-based brain tumor classification using magnetic resonance images. *Multimed. Tools Appl.* 83, 20487–20506. doi: 10.1007/s11042-023-16143-w

Rasool, M., Ismail, N. A., Al-Dhaqm, A., Yafooz, W. M., and Alsaeedi, A. (2022). A novel approach for classifying brain tumours combining a squeezenet model with svm and fine-tuning. *Electronics* 12:149. doi: 10.3390/electronics12010149

Raza, , Ayub, H., Khan, J. A., Ahmad, I., Salama, A. S., Daradkeh, Y. I., et al. (2022). A hybrid deep learning-based approach for brain tumor classification. *Electronics* 11:1146. doi: 10.3390/electronics11071146

Sadad, T., Rehman, A., Munir, A., Saba, T., Tariq, U., Ayesha, N., et al. (2021). Brain tumor detection and multi-classification using advanced deep learning techniques. *Microsc. Res. Tech.* 84, 1296–1308. doi: 10.1002/jemt.23688

Saeedi, S., Rezayi, S., Keshavarz, H., and Niakan Kalhori, S. R. (2023). MRI-based brain tumor detection using convolutional deep learning methods and chosen machine learning techniques. *BMC Med. Inform. Decis. Mak.* 23:16. doi: 10.1186/s12911-023-02114-6

Salahuddin, Z., Woodruff, H. C., Chatterjee, A., and Lambin, P. (2022). Transparency of deep neural networks for medical image analysis: A review of interpretability methods. *Comput. Biol. Med.* 140:105111. doi: 10.1016/j.compbiomed.2021.105111

Sarada, B., Narasimha Reddy, K., Babu, R., and Ramesh Babu, B. S. S. V. (2024). Brain tumor classification using modified ResNet50V2 deep learning model. *Int. J. Comput. Digital Syst. [pre print]* 16, 1–10.

Sarkar, A., Maniruzzaman, M., Alahe, M. A., and Ahmad, M. (2023). An effective and novel approach for brain tumor classification using AlexNet CNN feature extractor and multiple eminent machine learning classifiers in MRIs. *J Sens* 2023, 1–19. doi: 10.1155/2023/1224619

Schiavon, D. E. B., Becker, C. D. L., Botelho, V. R., and Pianoski, T. A. (2023). "Interpreting convolutional neural networks for brain tumor classification: an explainable artificial intelligence approach" in *Brazilian conference on intelligent systems* (Springer Nature Switzerland: Cham), 77–91.

Sharma, S., and Chaudhary, P. (2023). Machine learning and deep learning. *Quantum Comput. Artif. Intell. Train. Mach. Deep Learn. Algorithms Quantum Comput*, 71–84. doi: 10.1515/9783110791402-004

Sharma, P., and Shukla, A. P. (2022). Brain tumor classification using convolution neural network. *Lecture Notes Networks Syst.*, 579–588. doi: 10.1007/978-981-16-7118-0_50

Suryawanshi, S., and Patil, S. B. (2024). Efficient brain tumor classification with a hybrid CNN-SVM approach in MRI. *J. Advances Inform. Technol.* 15, 340–354. doi: 10.12720/jait.15.3.340-354

Taher, F., Shoaib, M. R., Emara, H. M., Abdelwahab, K. M., Abd El-Samie, F. E., and Haweel, M. T. (2022). Efficient framework for brain tumor detection using different deep learning techniques. *Front. Public Health* 10. doi: 10.3389/fpubh.2022.959667

Mahesh, T., Thakur, A., Gupta, M., Sinha, D. K., Mishra, K. K., Venkatesan, V. K., et al. (2024). Transformative breast Cancer diagnosis using CNNs with optimized ReduceLROnPlateau and early stopping enhancements. *Int. J. Comput. Intell. Syst.* 17. doi: 10.1007/s44196-023-00397-1

Tong, J., and Wang, C. (2023). A dual tri-path CNN system for brain tumor segmentation. *Biomed. Signal Process. Control* 81:104411. doi: 10.1016/j.bspc.2022.104411

Vidyarthi, R., Agarwal, D., Gupta, R., Sharma, D. D., and Tiwari, P. (2022). Machine learning assisted methodology for multiclass classification of malignant brain tumors. *IEEE Access* 10, 50624–50640. doi: 10.1109/access.2022.3172303

Wang, N., Lee, C. Y., Park, H. C., Nauen, D. W., Chaichana, K. L., Quinones-Hinojosa, A., et al. (2022). Deep learning-based optical coherence tomography image analysis of human brain cancer. *Biomed. Opt. Express* 14, 81–88. doi: 10.1364/boe.477311

Xiong, S., Wu, G., Fan, X., Feng, X., Huang, Z., Cao, W., et al. (2021). MRI-based brain tumor segmentation using FPGA-accelerated neural network. *BMC Bioinformatics* 22:421. doi: 10.1186/s12859-021-04347-6

Yildirim, M., Cengil, E., Eroglu, Y., and Cinar, A. (2023). Detection and classification of glioma, meningioma, pituitary tumor, and normal in brain magnetic resonance imaging using deep learning-based hybrid model. *Iran J. Comput. Sci.* 6, 455–464. doi: 10.1007/s42044-023-00139-8

Zhang, Z., and Sejdić, E. (2019). Radiological images and machine learning: trends, perspectives, and prospects. *Comput. Biol. Med.* 108, 354–370. doi: 10.1016/j.compbiomed.2019.02.017

Zhao, R. (2023). Brain tumor identification based on AlexNet and VGG. *Highlights Sci. Engin. Technol.* 57, 57–61. doi: 10.54097/hset.v57i.9897

Zhou, Z., Qiu, Q., Liu, H., Ge, X., Li, T., Xing, L., et al. (2023). Automatic detection of brain metastases in T1-weighted construct-enhanced MRI using deep learning model. *Cancers* 15:4443. doi: 10.3390/cancers15184443

Frontiers in Computational Neuroscience

TYPE Original Research
PUBLISHED 12 June 2024
DOI 10.3389/fncom.2024.1414462

OPEN ACCESS

EDITED BY
Sadique Ahmad,
Prince Sultan University, Saudi Arabia

REVIEWED BY
Farhan Ullah,
Northwestern Polytechnical University, China
Misbah Abbas,
Nencki Institute of Experimental Biology (PAS), Poland
Muhammad Imad,
Ulster University, United Kingdom

*CORRESPONDENCE
Tae Hoon Kim
✉ 323020@zust.edu.cn
Moez Krichen
✉ moez.krichen@redcad.org

RECEIVED 08 April 2024
ACCEPTED 13 May 2024
PUBLISHED 12 June 2024

CITATION
Kim TH, Krichen M, Ojo S, Sampedro GA and Alamro MA (2024) SS-DRPL: self-supervised deep representation pattern learning for voice-based Parkinson's disease detection. *Front. Comput. Neurosci.* 18:1414462. doi: 10.3389/fncom.2024.1414462

SS-DRPL: self-supervised deep representation pattern learning for voice-based Parkinson's disease detection

Tae Hoon Kim[1]*, Moez Krichen[2]*, Stephen Ojo[3], Gabriel Avelino Sampedro[4,5] and Meznah A. Alamro[6]

[1]School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou, Zhejiang, China, [2]FCSIT, Al-Baha University, Al-Baha, Saudi Arabia, [3]Department of Electrical and Computer Engineering, College of Engineering, Anderson University, Anderson, SC, United States, [4]Faculty of Information and Communication Studies, University of the Philippines Open University, Los Baños, Philippines, [5]Gokongwei College of Engineering, De La Salle University, Manila, Philippines, [6]Department of Information Technology, College of Computer and Information Science, Princess Nourah Bint Abdul Rahman University, Riyadh, Saudi Arabia

Parkinson's disease (PD) is a globally significant health challenge, necessitating accurate and timely diagnostic methods to facilitate effective treatment and intervention. In recent years, self-supervised deep representation pattern learning (SS-DRPL) has emerged as a promising approach for extracting valuable representations from data, offering the potential to enhance the efficiency of voice-based PD detection. This research study focuses on investigating the utilization of SS-DRPL in conjunction with deep learning algorithms for voice-based PD classification. This study encompasses a comprehensive evaluation aimed at assessing the accuracy of various predictive models, particularly deep learning methods when combined with SS-DRPL. Two deep learning architectures, namely hybrid Long Short-Term Memory and Recurrent Neural Networks (LSTM-RNN) and Deep Neural Networks (DNN), are employed and compared in terms of their ability to detect voice-based PD cases accurately. Additionally, several traditional machine learning models are also included to establish a baseline for comparison. The findings of the study reveal that the incorporation of SS-DRPL leads to improved model performance across all experimental setups. Notably, the LSTM-RNN architecture augmented with SS-DRPL achieves the highest F1-score of 0.94, indicating its superior ability to detect PD cases using voice-based data effectively. This outcome underscores the efficacy of SS-DRPL in enabling deep learning models to learn intricate patterns and correlations within the data, thereby facilitating more accurate PD classification.

KEYWORDS

Parkinson's disease, artificial intelligence, self-supervised deep representation pattern learning, machine learning, FT-HV

1 Introduction

The neurological disorder known as Parkinson's disease (PD) has a substantial impact on a considerable proportion of the worldwide population (Muangpaisan et al., 2011; Rocca, 2018). The dopaminergic neurons in the substantia nigra region of the brain gradually deteriorate, which distinguishes them. As a consequence, individuals have motor symptoms, including tremors, bradykinesia, stiffness, and postural instability, which have a substantial influence on their daily functioning and overall quality of life. Due to the chronic nature of the condition and the way these symptoms are advancing over time, comprehensive therapy and care are necessary. Timely detection and prompt intervention play a critical role in minimizing the consequences of PD and maximizing the effectiveness of treatment. The ability to identify the subtle indications and manifestations of PD at its early stages enables the implementation of suitable therapeutic measures with the goal of mitigating symptoms, maintaining motor abilities, and enhancing the overall quality of life (Kanagaraj et al., 2024; Xia et al., 2024).

Individuals with PD may benefit from early intervention measures such as lifestyle adjustments, rehabilitative therapy, and pharmaceutical interventions that are specifically designed to address their unique requirements and limitations (Olanrewaju et al., 2014; Bhat et al., 2018; Lang and Espay, 2018). In order to effectively address a broader spectrum of non-motor symptoms, including cognitive impairment, psychiatric symptoms, and autonomic dysfunction, neurologists, movement disorder specialists, rehabilitation therapists, nurses, social workers, and other healthcare professionals must collaborate in a multidisciplinary approach to care (Seppi et al., 2019). The progress made in diagnostic technology, neuroimaging techniques, and biomarker identification has greatly contributed to the timely and precise identification of PD, allowing healthcare practitioners to intervene during the initial phases of the condition. Nevertheless, there are still obstacles that remain, such as the requirement for more accurate diagnostic biomarkers, the advancement of neuroprotective medicines, and the enhancement of customized therapy approaches for each patient. The increasing societal and economic impact of PD highlights the imperative for ongoing research, lobbying, and public awareness campaigns aimed at improving healthcare accessibility, enhancing treatment efficacy, and finally achieving a cure for this incapacitating disorder (Liu et al., 2024; Sardar and Pahari, 2024).

Recent developments in the field of deep learning have demonstrated potential for enhancing the precision and effectiveness of PD identification (Ali et al., 2023). The use of Self Supervised Representation learning (SSRL) has become increasingly prominent as an effective methodology for acquiring representations from data that lacks labels (Ericsson et al., 2021). The SSRL technique has become a reliable approach for extracting significant representations from data that lacks labels. This method shows potential for enhancing the effectiveness of voice-based PD detection systems. SSRL fundamentally entails acquiring representations from data without the need for explicit supervision (Mohamed et al., 2022). SSRL algorithms differ from standard supervised learning approaches in that they can predict a specific part of the input data based on another part without requiring explicitly defined target outputs. The utilization of this self-supervised methodology allows models to exploit the intrinsic structure and patterns that exist within the data, resulting in enhanced learning efficiency and effectiveness. The success of SSRL algorithms in several areas is indicative of their adaptability since they have exhibited the capacity to improve the efficiency of learning tasks by enhancing sample efficiency (Zhang et al., 2017).

SSRL shows potential in the field of voice-based PD identification. Through the utilization of unlabeled data, these algorithms can reveal latent patterns and characteristics that are suggestive of PD, consequently enhancing the precision of predictive models (Nekoui and Cheng, 2021). It is noteworthy that the numerous benefits of improved PD identification are made possible by self-supervised deep representational learning (Jiang et al., 2021). The timely identification and Early diagnosis of PD is very important so that treatments can be started right away, which could slow the disease's progression and improve patient outcomes (Mei et al., 2021; Öksüz et al., 2022). SSRL algorithms are very important for making detection methods more accurate and reliable, which speeds up the process of early diagnosis and allows healthcare practitioners to intervene during the initial phases of the disease. Moreover, the economic and systemic consequences of improved Parkinson's disease identification are of utmost importance. The utilization of more precise diagnostic instruments not only leads to enhanced patient care but also holds the capacity to substantially mitigate healthcare expenses linked to misdiagnosis or delayed diagnosis (Painuli et al., 2022). SSRL-driven breakthroughs in PD detection enhance healthcare delivery by simplifying the diagnostic procedure and increasing the efficiency of healthcare systems. Furthermore, the utilization of SSRL approaches not only enhances the precision of PD identification but also exhibits the potential for furthering our comprehension of the fundamental mechanisms associated with this condition. These algorithms offer valuable insights into the subtle connections and biomarkers linked to voice-based PD by revealing intricate patterns and representations within unlabeled data. This enhanced comprehension can contribute to the advancement of therapeutic approaches and therapies designed to alleviate the consequences of the condition (Güvenç Paltun et al., 2021).

Motivation: PD is a significant healthcare challenge, demanding precise and timely diagnostic methods for effective intervention. Traditional approaches may lack accuracy, prompting the exploration of advanced techniques like self-supervised deep representation pattern learning (SS-DRPL) to extract meaningful features from voice data. By integrating SS-DRPL with deep learning algorithms, this study aims to enhance the accuracy of voice-based PD detection. Evaluation of various machine learning models seeks to identify optimal approaches for early PD identification, ultimately improving patient outcomes through timely intervention and personalized treatment strategies.

Contributions: This paper makes the following contributions.

- **Self-supervised deep representational pattern learning technique:** Propose Self-supervised representational pattern learning techniques for the extraction of latent features and structures from data, resulting in enhanced discriminative capabilities of the models. Self-supervised learning has enabled the more efficient utilization of available data in prediction for Parkinson's disease detection by predicting one part of the input data based on another part without relying on manually provided target outputs.
- **Voice-based PD classification with machine learning classifier:** Proposed SS-DRPL based LSTM-RNN, DNN, trained on voice dataset for the detection of PD. The DNN

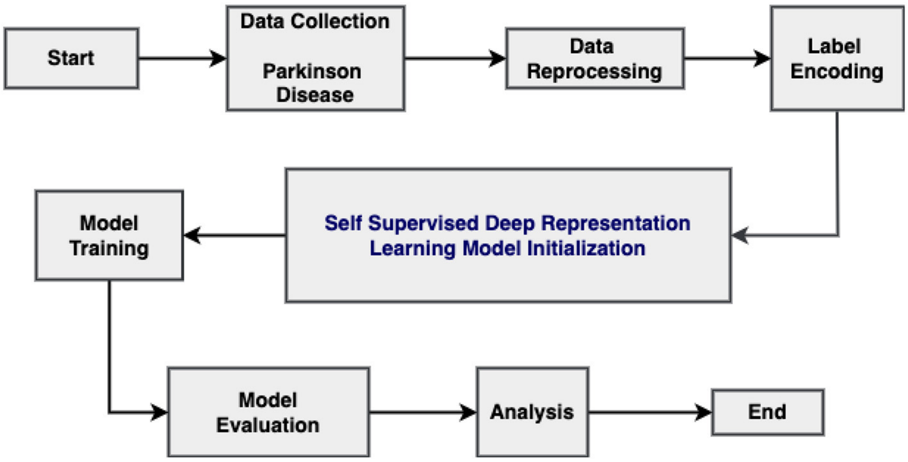


FIGURE 1
Employed methodology.

```
Data gathering and preprocessing:
    • Collect data for Parkinson's disease
    • Data preprocessing
Input: Unlabeled dataset $\mathcal{D}_s = \{x_i^{(s)}\}_{i=1}^{M}$
for $i$ from 1 to $M$ do
    Sample $x_a \sim T(x_i^{(s)})$
    Sample $x_+ \sim T(x_i^{(s)})$
    for $k$ from 1 to $K$ do
        Sample j ∼ U(1, M) {Choose a different raw input}
        Sample $x_{-k} \sim T(x_j^{(s)})$ {Have a random transform}
    end for
    $x_i \leftarrow \{(x_a, x_+), (x_a, x_{-1}), \ldots, (x_a, x_{-K})\}$
    $z_i \leftarrow \{1, 0, \ldots, 0\}$
end for
Output: $\{x_i, z_i\}_{i=1}^{M}$
Model training:
    • DNN, and LSTM-RNN
Results analysis:
    • Analyze model performance metrics.
    • Discuss future research suggestions and implications.
    • Calculate and print results.
```

Algorithm 1. SS-DRPL for voice-based Parkinson's disease detection.

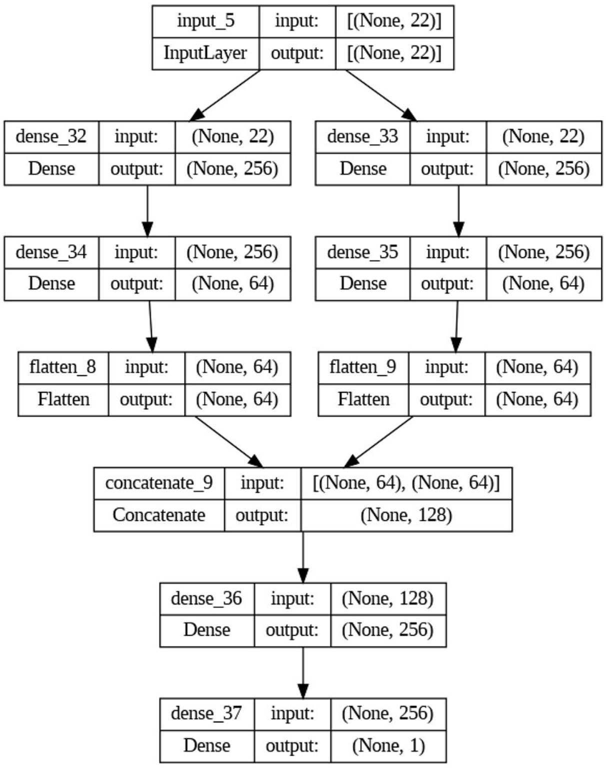


FIGURE 2
DNN architecture.

demonstrated a performance of 0.85% in terms of accuracy. The LSTM-RNN, with a score of 0.93%, addresses the issue of overfitting and enhances generalization by integrating many models. Likewise, it resulted in improved predicted accuracy. The LSTM-RNN model demonstrated a good accuracy of 0.93%.

Paper organization: The paper is organized in a manner that thoroughly examines the study. The study begins with an introduction that provides an overview of the context and importance of the research. The following parts then provide a detailed explanation of the suggested approach. Subsequently, the study presents its findings, demonstrating the efficacy of

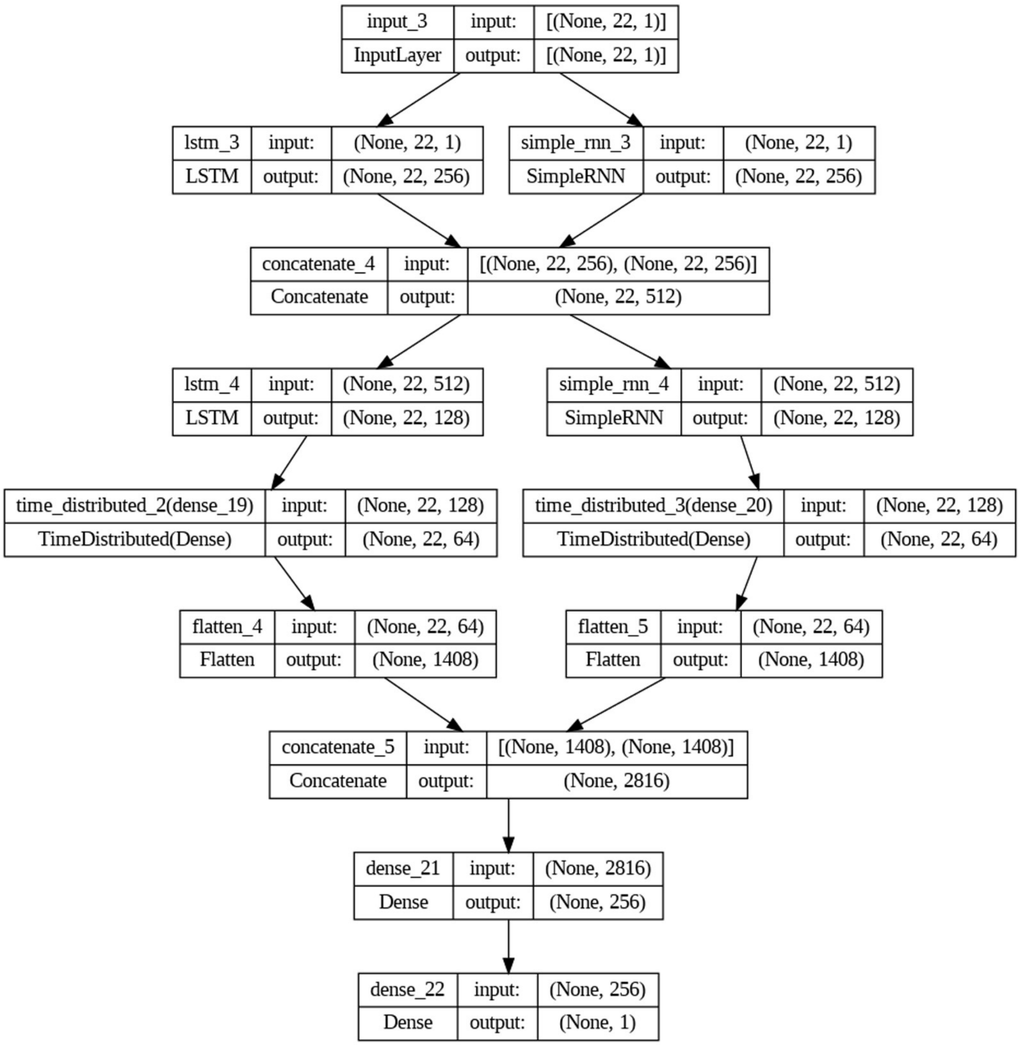


FIGURE 3
LSTM-RNN architecture.

TABLE 1 Classification reports for selected models.

Model	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
LSTM-RNN	0.94	0.94	0.94	0.94
DNN	0.86	0.86	0.84	0.86
DT	0.85	0.86	0.85	0.86
LR	0.92	0.92	0.92	0.92
SVM	0.84	0.80	0.74	0.80
GNB	0.87	0.73	0.75	0.73

the proposed methodology. Similarly, the study backs up its conclusions with visual aids and statistical measurements to help people understand better. The study will conduct a detailed assessment of its advantages and constraints and analyze its practical ramifications for future endeavors. Finally, the report provides a concise overview of the research's accomplishments and the study's prospective implications.

2 Related work

SS-DRPL has become a popular and interesting method in machine learning over the past few years (Ericsson et al., 2022). It may be better than traditional guided methods in some ways (Feng et al., 2019). However, when looking at how it can be used for more challenging tasks like predicting Parkinson's disease, the theories

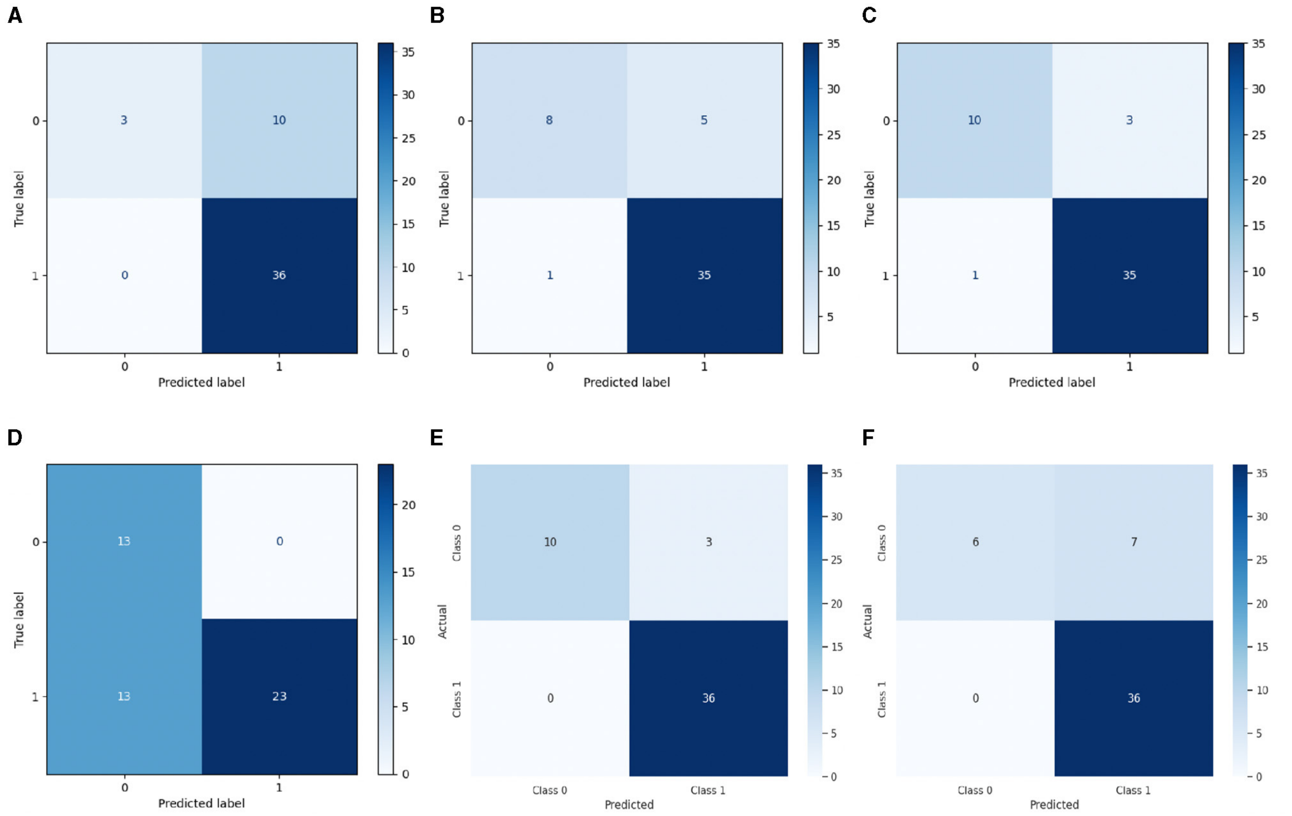


FIGURE 4
Confusion matrix of **(A)** SVM, **(B)** DT, **(C)** LR, **(D)** Gaussian NB, **(E)** LSTM-RNN, and **(F)** Deep DNN models.

behind SS-DRPL seem less developed than those behind traditional guided learning paradigms (Zhang et al., 2017). Likewise, the goal of this review of the literature is to look into the pros and cons of using SS-DRPL to predict voice-based PD, emphasizing new research and evaluations that have been done recently. The basic idea behind normal supervised learning is to figure out how well a model should do on data it has yet to see (Jiang et al., 2020; Sen et al., 2020). On the other hand, SS-DRPL methods are hard because the training loss is usually optimized for a pretext task, but the downstream task's success metric is different. In addition, this misalignment brings up important questions about how learned models in SS-DRPL can be transferred and used in other situations. Because SS-DRPL has some problems when it comes to predicting Parkinson's disease, experts have done much real-world testing using different neural network architectures and datasets. For example, convolutional neural networks have been used in recent studies to sort large amounts of spectrogram images that contain gait patterns (Cetin, 2023; Nafiz et al., 2023). Deeply dense artificial neural networks have also been used to look at voice recordings and guess when Parkinson's disease will start or get worse.

The investigations conducted have yielded empirical findings that indicate encouraging results. The proposed models consistently surpass the current advanced methods in terms of categorization accuracy. Similarly, the VGFR spectrogram detector had a notable accuracy rate of 0.88, while the voice impairment classifier displayed an even better accuracy rate of 0.89 (Johri et al., 2019). This suggests the possibility of utilizing SS-DRPL approaches to improve predictive accuracy in the diagnosis and monitoring of PD. In addition, researchers have investigated different approaches to enhance the representation of features and the accuracy of classification in tasks related to predicting PD. Moreover, one way to do this is to use correlation maps made by principal component analysis, information gain, and other methods for feature selection to make the feature space bigger (Sharma and Mishra, 2022). It is important to note that the classification results obtained by using extended feature sets have been better than those obtained by using original features. This underscores the significance of feature engineering in augmenting prediction performance. More precisely, Researchers have investigated various supervised AI algorithms, such as Decision Trees, K-Nearest Neighbors, Random Forests, Bagging, AdaBoosting, and Gradient Boosting, in order to improve classifier accuracy and predictive performance in PD classification (Sabu et al., 2022; Shastry, 2023). These techniques utilize permutation computations to optimize model performance and improve the accuracy of receiver operating characteristic (ROC) curves, which is a crucial parameter for evaluating classifier performance.

In order to enhance the accuracy of feature representation and the results of classification, authors in Malekroodi et al. (2024) have implemented three separate component selection procedures. These strategies enable each of the 23 features to identify and pick the top 10 most effective features. The

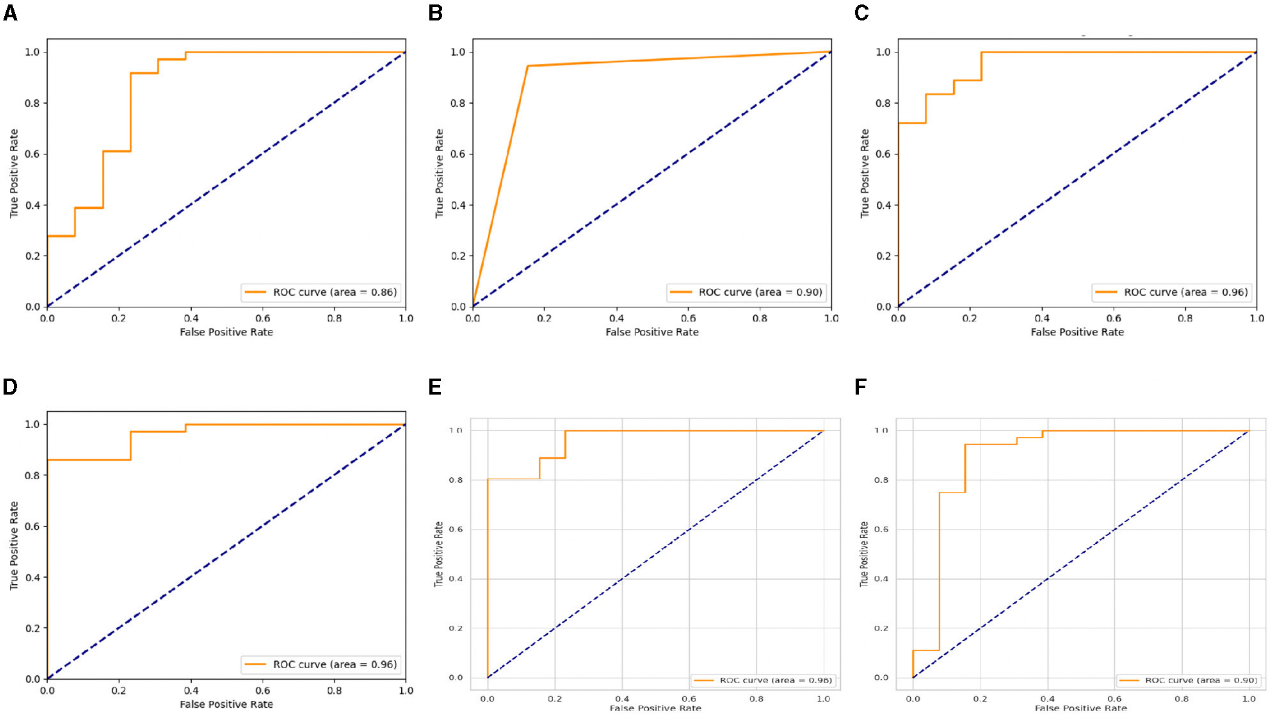


FIGURE 5
ROC curves of **(A)** SVM, **(B)** DT, **(C)** LR, **(D)** Gaussian NB, **(E)** LSTM-RNN, and **(F)** Deep DNN models.

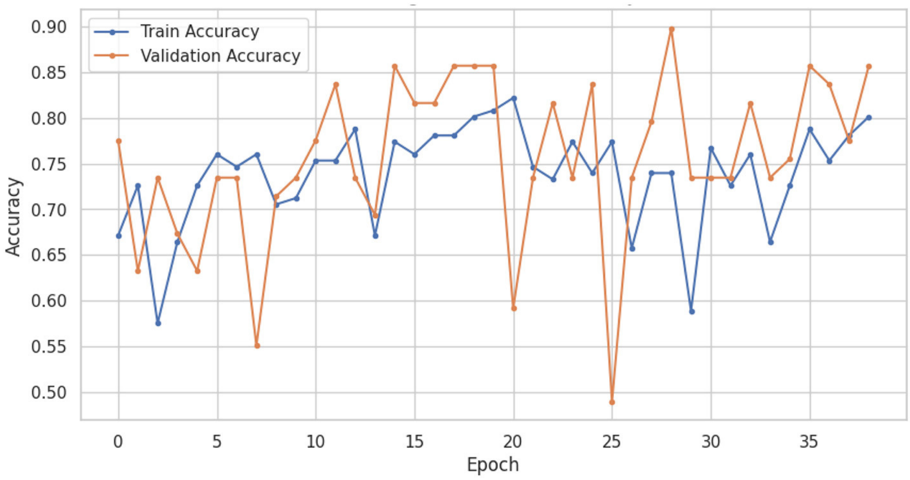


FIGURE 6
DNN training and validation accuracy curves.

DT classifier has shown remarkable accuracy, reaching 0.94%, even in datasets with 23 features, similar to those with only 11 features. Specifically, the exceptional performance has been validated through the examination of the ROC curve, which has demonstrated a noteworthy area under the curve of 0.98. The implication of these findings are of great importance for clinical practice, as they highlight the effectiveness of computer-based algorithms in reliably differentiating people with PD from those without PD at the individual level. Clinicians can utilize advanced AI techniques and careful feature selection strategies to utilize computer-based findings. This enables them to improve patient care and outcomes in the management of PD by facilitating early diagnosis, personalized treatment planning, and ongoing disease monitoring. SS-DRPL opens up new ways to improve the prediction of voice-based PD. However, there are still some problems to solve before the theoretical foundations of SS-DRPL can be used in real-life medical tasks (Tripathi et al., 2024). Overall, by tackling these problems, SS-DRPL can be fully useful for better

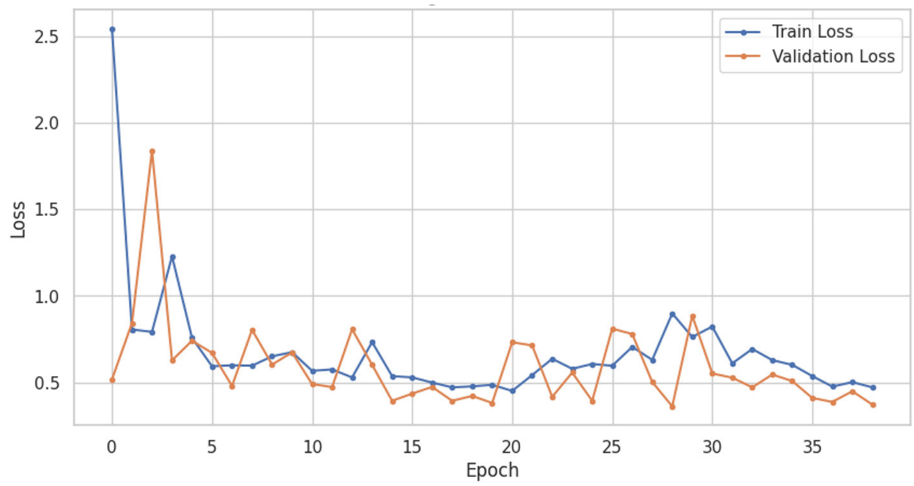


FIGURE 7
DNN training and validation loss.

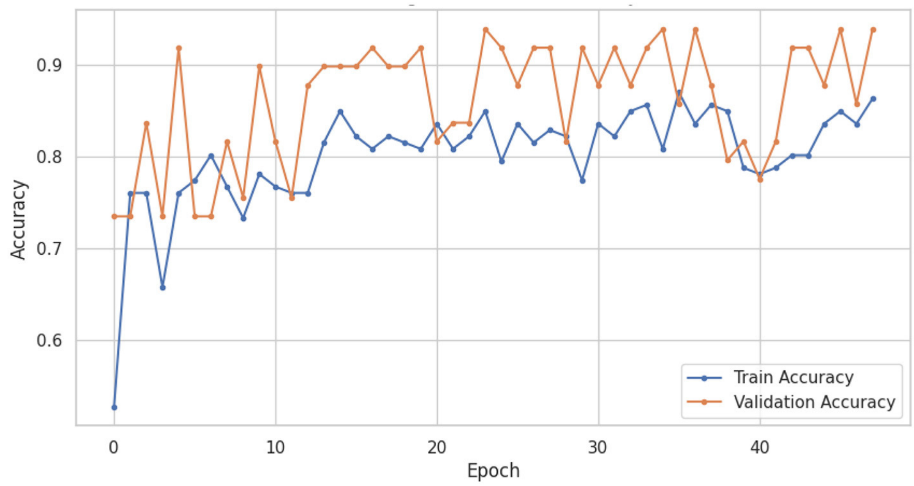


FIGURE 8
LSTM-RNN training and validation accuracy.

diagnosing, treating, and predicting Parkinson's disease, which will eventually improve patient outcomes and quality of life.

3 Proposed methodology

This section will provide an overview of the research design, paradigm, data collection methods, and analysis tools employed in the study. Ensuring accuracy in the approach is crucial for establishing the reliability and validity of the findings. Following that, this section provides a brief yet thorough summary of the strategic method, establishing the foundation for a thorough empirical investigation and significant contributions to the area. Figure 1 illustrates the methodology employed. The proposed approach starts with dataset collection, then pre-processing, then label encoding and the employing SSRL-based deep learning models for training and testing, and finally, results analysis.

The dataset used in this study consists of 707 voice samples obtained from individuals with various traumatic, biological, and neurological disorders, including Parkinson's disease, along with 53 samples from healthy individuals serving as controls. Each sample, averaging 2 s in duration, was recorded in controlled, quiet acoustic settings. The dataset encompasses multiple voice features such as fundamental frequency (MDVP:Fo), frequency variations (MDVP:Fhi, MDVP:Flo), jitter (MDVP:Jitter), shimmer (MDVP:Shimmer), noise-to-harmonic ratio (NHR), and harmonics-to-noise ratio (HNR), among others. These features provide comprehensive information about voice signals. Likewise, to ensure uniformity in feature scales, the data underwent preprocessing steps, including downsampling from 50 to 25 kHz and scaling using Min Max scaling (Ali et al., 2014). By bringing all features into a common numerical range, this preprocessing helps improve the convergence of machine learning models (Little et al., 2007). Algorithm 1 presents an

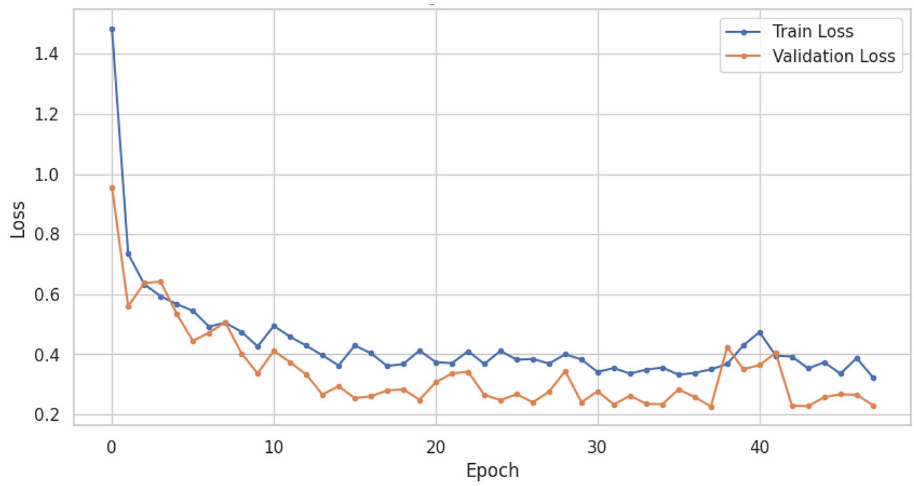


FIGURE 9
LSTM-RNN training and validation loss.

overview of the entire pipeline of the proposed approach, which begins with collecting data, preprocessing it and evaluating results.

The utilization of conventional machine learning models, including Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), and Gaussian Naive Bayes (GNB), did not incorporate SS-DRPL. Following this, the research employed SS-DRPL techniques, specifically DNN and LSTM-RNN. The classification of Parkinson's disease was performed by training these models using preprocessed voice samples, utilizing the retrieved features. The study utilizes a DNN structure consisting of numerous dense layers for representation learning and fully linked layers for self-supervised learning and transfer learning. To extract varied representations from input voice data, the model employs parallel, thick layers consisting of 256 neurons each. The representations are refined using separate, fully connected layers with 64 neurons and then flattened and concatenated for transfer learning. In order to do binary classification, a sigmoid activation function is employed in a final, fully linked layer consisting of 256 neurons. Subsequently, the model was made with the Adam optimizer and the binary cross-entropy loss function. Early stopping techniques were used to lower the risk of overfitting during training.

The study also includes an LSTM-RNN architecture for learning representations. The suggested model has two layers, LSTM and RNN, running in parallel. Each layer has 256 neurons, and they work together to capture the temporal dependencies in the voice data. The outputs of the layers listed above are combined and then sent to the extra LSTM and RNN layers, each of which has 128 neurons, improving the accuracy of the representations. Self-supervised learning is implemented using fully connected layers consisting of 64 neurons each. These layers are applied using a time-distributed wrapper to process each time step separately. In the context of transfer learning, the outputs undergo a process of flattening and concatenation. Subsequently, fully connected layers with 256 neurons are employed for binary classification, utilizing a sigmoid activation function. The experimental setup involves training the model using the Adam optimizer and binary cross-entropy loss function. To mitigate the risk of overfitting, early halting is employed. Figures 2, 3 show the architectures of DNN and LSTM-RNN. These models possess the ability to independently acquire significant representations straight from unprocessed voice signals, obviating the necessity for manually designed features. In addition, the preprocessed samples were used to train the DNN and LSTM-RNN models. Throughout the training process, the models acquired the ability to extract complex characteristics from the unprocessed voice signals, effectively capturing detailed patterns that are known to be indicative of Parkinson's disease. The inherent characteristic of self-supervised learning allows the models to acquire representations independently without the need for explicitly labeled input to extract features.

Both classical machine learning models and self-supervised deep representation learning models were tested to see how well they worked. These tests included accuracy, precision, recall, F1-score, ROC curve, and confusion matrix analysis. This set of measures offers significant insights into the outcomes of the algorithms to detect Parkinson's disease based on speech samples precisely. SS-DRPL algorithms can assist in early diagnosis by improving the accuracy and reliability of detection methods. This enables healthcare practitioners to intervene during the early stages of the disease. By taking a proactive approach, not only are patient outcomes improved, but effective treatment techniques may be implemented promptly, eventually boosting the overall management of the condition.

4 Results and experimentation

The outcomes of the study are presented and interpreted in the results and experimentation section, taking into consideration the research objectives. Similarly, this part provides the results and offers valuable perspectives, explanations, and ramifications of the findings. The dataset was split into 80% for training and 20% for testing. Likewise, voice-based PD poses a complicated diagnostic dilemma owing to its diverse range of manifestations. The objective of this study is to utilize SS-DRPL in order to create reliable classification models that can effectively detect voice-based PD using a wide range of features derived from patient data. The study utilized several machine learning algorithms, such as DT, LR, SVM, and GNB, as well as deep learning algorithms, such as LSTM-RNN and DNN, each possessing distinct strengths and capabilities. The LSTM-RNN models exhibited outstanding performance, showcasing impressive precision, recall, and F1-score metrics, all reaching a value of 0.94. Using recurrent neural networks to find complex linkages in the data is a good way to classify Parkinson's disease cases accurately. The improved models demonstrate strong performance, which shows great potential for clinical applications. In these contexts, early and precise diagnosis plays a critical role in effectively managing diseases and implementing intervention measures. Table 1 shows the performance metrics of employed models.

Nevertheless, the LSTM-RNN demonstrated better performance; the DNN model displayed somewhat poorer precision, recall, and F1-score metrics, but it still achieved satisfactory levels of accuracy. Figures 4, 5 show the confusion matrix and ROC curves of the employed models. Subsequently, the observed disparity can be attributed to the model's proficiency in extracting pertinent features from the data or its aptitude for generalizing to unfamiliar examples. However, the performance of DNN highlights its potential usefulness in voice-based PD classification tasks, although there is still some opportunity for enhancement. Moreover, our examination of conventional machine learning algorithms showed various degrees of effectiveness. The Logistic Regression and Decision Tree models demonstrated high accuracy rates of 0.92 and 0.86, respectively, indicating their suitability for voice-based PD classification tasks. In contrast, the SVM and Gaussian NB models had somewhat lower percentages of accuracy, specifically 0.80 and 0.73, respectively. This indicates the necessity for further optimization or feature engineering in order to enhance their diagnostic capabilities for Parkinson's disease. These findings provide useful insights into the advantages and constraints of various machine learning methods for classifying Parkinson's disease.

The LSTM-RNN performs strongly, highlighting their potential usefulness in clinical environments where precise and prompt diagnosis is crucial for patient care. Also, comparing different types of traditional machine learning models gives us much useful information about how well they work and where we can find new research and development opportunities. In summary, this research showcases the considerable potential of machine learning in the categorization of voice-based PD, with sophisticated models like the LSTM-RNN exhibiting significant potential. Utilizing these models' advantageous aspects and acknowledging their constraints can provide a pathway toward enhanced voice-based PD, thereby leading to enhanced patient outcomes and improved quality of life. Figures 6, 7 shows the training and validation accuracy and loss curves provide a visual representation of the DNN performance during training, offering insights into its learning progress and potential overfitting or underfitting and assessing the convergence and generalization capabilities of the model, aiding in the optimization of hyperparameters and the identification of training issues.

TABLE 2 Comparison with base study.

Study	Accuracy (%)
LSTM-RNN	0.94
Arjmandi et al. (2011)	0.91

Figures 8, 9 show the training and validation accuracy and loss curves of LSTM-RNN performance during training, offering insights into its learning progress and potential overfitting or underfitting and assessing the convergence and generalization capabilities of the model, aiding in the optimization of hyperparameters and the identification of training issues.

Table 2 provides the comparison of the base study with our proposed approach SS-DRPL-based LSTM-RNN approach. The results reveal that the proposed approach provides the best accuracy of 0.94% in comparison with a base study suggesting our approach for voice-based PD detection.

5 Conclusion and future work

This study explored the potential of deep learning techniques and SS-DRPL in voice-based PD classification. This study demonstrated the effectiveness of these models in identifying patterns within the dataset, thereby enabling accurate classification of PD individuals. Furthermore, the experiment of deep learning models yields significant findings regarding their relative effectiveness and possible applicability in clinical environments. Results reveal that the hybrid LSTM-RNN method effectively identifies voice-based PD using different sets of patient data features. Although this study offers useful insights into the classification of PD, there are still various areas for further research and development that need to be investigated. Similarly, combining different types of data sources, such as genetic markers, imaging data, and patient demographics, could improve the accuracy of classification models and give a fuller picture of the risk factors connected to PD. Furthermore, using longitudinal studies that track the progress of diseases over a long period could reveal important information about how voice-based PD develops and help the development of more personalized treatment methods. The establishment of collaborative partnerships among data scientists, physicians, and healthcare providers will play a crucial role in the translation of research findings into practical insights, thereby enhancing patient outcomes and advancing our comprehension of PD. This will ultimately result in a more precise diagnosis, tailored treatment strategies, and enhanced quality of life for individuals affected by this condition.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

TK: Writing – original draft. MK: Writing – original draft. SO: Writing – original draft. GS: Writing – original draft. MA: Writing – original draft.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R503), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Ali, P. J. M., Faraj, R. H., Koya, E., Ali, P. J. M., and Faraj, R. H. (2014). Data normalization and standardization: a technical report. *Mach. Learn. Tech. Rep.* 1, 1–6.

Ali, S., Hashmi, A., Hamza, A., Hayat, U., and Younis, H. (2023). Dynamic and static handwriting assessment in Parkinson's disease: a synergistic approach with C-Bi-GRU and VGG19. *J. Comput. Theor. Appl.* 1, 151–162. doi: 10.33633/jcta.v1i2.9469

Arjmandi, M. K., Pooyan, M., Mikaili, M., Vali, M., and Moqarehzadeh, A. (2011). Identification of voice disorders using long-time features and support vector machine with different feature reduction methods. *J. Voice* 25, e275–e289. doi: 10.1016/j.jvoice.2010.08.003

Bhat, S., Acharya, U. R., Hagiwara, Y., Dadmehr, N., and Adeli, H. (2018). Parkinson's disease: cause factors, measurable indicators, and early diagnosis. *Comput. Biol. Med.* 102, 234–241. doi: 10.1016/j.compbiomed.2018.09.008

Cetin, O. (2023). Accent recognition using a spectrogram image feature-based convolutional neural network. *Arab. J. Sci. Eng.* 48, 1973–1990. doi: 10.1007/s13369-022-07086-9

Ericsson, L., Gouk, H., Loy, C. C., and Hospedales, T. M. (2021). Self-supervised representation learning: introduction, advances and challenges. *arXiv:2110.09327*. doi: 10.48550/arXiv.2110.09327

Ericsson, L., Gouk, H., Loy, C. C., and Hospedales, T. M. (2022). Self-supervised representation learning: introduction, advances, and challenges. *IEEE Sign. Process. Mag.* 39, 42–62. doi: 10.1109/MSP.2021.3134634

Feng, Z., Xu, C., and Tao, D. (2019). "Self-supervised representation learning by rotation feature decoupling," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE), 10364–10374.

Güvenç Paltun, B., Kaski, S., and Mamitsuka, H. (2021). Machine learning approaches for drug combination therapies. *Brief. Bioinformat.* 22:bbab293. doi: 10.1093/bib/bbab293

Jiang, H., Lim, W. Y. B., Ng, J. S., Wang, Y., Chi, Y., and Miao, C. (2021). "Towards Parkinson's disease prognosis using self-supervised learning and anomaly detection," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Toronto, ON: IEEE), 3960–3964.

Jiang, T., Gradus, J. L., and Rosellini, A. J. (2020). Supervised machine learning: a brief primer. *Behav. Ther.* 51, 675–87. doi: 10.1016/j.beth.2020.05.002

Johri, A., Tripathi, A., Johri, A., and Tripathi, A. (2019). "Parkinson's disease detection using deep neural networks," in *2019 Twelfth International Conference on Contemporary Computing (IC3)* (Noida: IEEE), 1–4.

Kanagaraj, S., Hema, M., and Guptha, M. N. (2024). An improved approach for early diagnosis of Parkinson's disease using advanced dl models and image alignment. *Automatika* 65, 911–924. doi: 10.1080/00051144.2023.2284030

Lang, A. E., and Espay, A. J. (2018). Disease modification in Parkinson's disease: current approaches, challenges, and future considerations. *Mov. Disord.* 33, 660–677. doi: 10.1002/mds.27360

Little, M., Mcsharry, P., Roberts, S., Costello, D., and Moroz, I. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Nat. Prec.* 326:1. doi: 10.1038/npre.2007.326.1

Liu, T. A., Koseoglu, N. D., and Jones, C. (2024). Self-supervised deep learning—the next frontier. *J. Am. Med. Assoc. Ophthalmol.* 2023:6650. doi: 10.1001/jamaophthalmol.2023.6650

Malekroodi, H. S., Madusanka, N., Lee, B. I., and Yi, M. (2024). Leveraging deep learning for fine-grained categorization of Parkinson's disease progression levels through analysis of vocal acoustic patterns. *Bioengineering* 11:295. doi: 10.3390/bioengineering11030295

Mei, J., Desrosiers, C., and Frasnelli, J. (2021). Machine learning for the diagnosis of Parkinson's disease: a review of literature. *Front. Aging Neurosci.* 13:633752. doi: 10.3389/fnagi.2021.633752

Mohamed, A., Lee, H. Y., Borgholt, L., Havtorn, J. D., Edin, J., Igel, C., et al. (2022). Self-supervised speech representation learning: a review. *IEEE J. Select. Top. Sign. Process.* 16, 1179–1210. doi: 10.1109/JSTSP.2022.3207050

Muangpaisan, W., Mathews, A., Hori, H., and Seidel, D. (2011). A systematic review of the worldwide prevalence and incidence of Parkinson's disease. *J. Med. Assoc. Thailand* 94:749.

Nafiz, M. F., Kartini, D., Faisal, M. R., Indriani, F., and Hamonangan, T. (2023). Automated detection of COVID-19 cough sound using MEL-spectrogram images and convolutional neural network. *J. Ilm. Tek. Elektro Komput. Dan Inform.* 9, 535–548.

Nekoui, M., and Cheng, L. (2021). "Enhancing human motion assessment by self-supervised representation learning," in *BMVC*, 322.

Öksüz, N., Öztürk, Ş., and Doğu, O. (2022). Future prospects in Parkinson's disease diagnosis and treatment. *Archiv.Neuropsychiat.* 59(Suppl.1):S36. doi: 10.29399/npa.28169

Olanrewaju, R. F., Sahari, N. S., Musa, A. A., and Hakiem, N. (2014). "Application of neural networks in early detection and diagnosis of Parkinson's disease," in *2014 International Conference on Cyber and IT Service Management (CITSM)* (South Tangerang: IEEE), 78–82.

Painuli, D., Bhardwaj, S., and Köse, U. (2022). Recent advancement in cancer diagnosis using machine learning and deep learning techniques: a comprehensive review. *Comput. Biol. Med.* 146:105580. doi: 10.1016/j.compbiomed.2022.105580

Rocca, W. A. (2018). The burden of Parkinson's disease: a worldwide perspective. *Lancet Neurol.* 17, 928–929. doi: 10.1016/S1474-4422(18)30355-7

Sabu, K., Ramnath, M., Choudhary, A., Raj, G., and Prakash Agrawal, A. (2022). "A comparison of traditional and ensemble machine learning approaches for Parkinson's disease classification," in *Machine Intelligence and Data Science Applications: Proceedings of MIDAS 2021* (Berlin: Springer), 25–33.

Sardar, S., and Pahari, S. (2024). "A study to find affordable ai techniques for early Parkinson's disease detection," in *Intelligent Technologies and Parkinson's Disease: Prediction and Diagnosis* (Hershey, PA: IGI Global), 20–51.

Sen, P. C., Hajra, M., and Ghosh, M. (2020). "Supervised classification algorithms in machine learning: a survey and review," in *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018* (Berlin: Springer), 99–111.

Seppi, K., Ray Chaudhuri, K., Coelho, M., Fox, S. H., Katzenschlager, R., Perez Lloret, S., et al. (2019). Update on treatments for nonmotor symptoms of

Parkinson's disease—an evidence-based medicine review. *Mov. Disord.* 34, 180–198. doi: 10.1002/mds.27602

Sharma, A., and Mishra, P. K. (2022). Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis. *Int. J. Inform. Technol.* 14, 1949–1960. doi: 10.1007/s41870-021-00671-5

Shastry, K. A. (2023). An ensemble nearest neighbor boosting technique for prediction of Parkinson's disease. *Healthc. Analyt.* 3:100181. doi: 10.1016/j.health.2023.100181

Tripathi, S., Acien, A., Holmes, A. A., Arroyo-Gallego, T., and Giancardo, L. (2024). Generalizing Parkinson's disease detection using keystroke dynamics: a self-supervised approach. *J. Am. Med. Inform. Assoc.* 2024:ocae050. doi: 10.1093/jamia/ocae050

Xia, Y., Sun, H., Zhang, B., Xu, Y., and Ye, Q. (2024). Prediction of freezing of gait based on self-supervised pretraining via contrastive learning. *Biomed. Sign. Process. Contr.* 89:105765. doi: 10.1016/j.bspc.2023.105765

Zhang, P., Wang, F., and Zheng, Y. (2017). "Self supervised deep representation learning for fine-grained body part recognition," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (Melbourne, VIC: IEEE), 578–582.

frontiers | Frontiers in Computational Neuroscience

TYPE Original Research
PUBLISHED 24 June 2024
DOI 10.3389/fncom.2024.1423051

EDITED BY
Ridha Ejbali,
Gabes University, Tunisia

REVIEWED BY
Qasim Ali,
Beijing Institute of Technology, China
Sushil Kumar Singh,
Marwadi University, India
Muhammad Wasim,
Islamia College University, Pakistan
Wei Zhan,
Yangtze University, China

*CORRESPONDENCE
Muhammad Shahid Anwar
✉ shahidanwar786@gachon.ac.kr

RECEIVED 25 April 2024
ACCEPTED 06 June 2024
PUBLISHED 24 June 2024

CITATION
Iqbal S, Qureshi AN, Alhussein M, Aurangzeb K, Choudhry IA and Anwar MS (2024) Hybrid deep spatial and statistical feature fusion for accurate MRI brain tumor classification. *Front. Comput. Neurosci.* 18:1423051. doi: 10.3389/fncom.2024.1423051

Hybrid deep spatial and statistical feature fusion for accurate MRI brain tumor classification

Saeed Iqbal[1], Adnan N. Qureshi[2], Musaed Alhussein[3], Khursheed Aurangzeb[3], Imran Arshad Choudhry[1] and Muhammad Shahid Anwar[4]*

[1]Department of Computer Science, Faculty of Information Technology and Computer Science, University of Central Punjab, Lahore, Pakistan, [2]Faculty of Arts, Society, and Professional Studies, Newman University, Birmingham, United Kingdom, [3]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia, [4]Department of AI and Software, Gachon University, Seongnam-si, Republic of Korea

The classification of medical images is crucial in the biomedical field, and despite attempts to address the issue, significant challenges persist. To effectively categorize medical images, collecting and integrating statistical information that accurately describes the image is essential. This study proposes a unique method for feature extraction that combines deep spatial characteristics with handmade statistical features. The approach involves extracting statistical radiomics features using advanced techniques, followed by a novel handcrafted feature fusion method inspired by the ResNet deep learning model. A new feature fusion framework (FusionNet) is then used to reduce image dimensionality and simplify computation. The proposed approach is tested on MRI images of brain tumors from the BraTS dataset, and the results show that it outperforms existing methods regarding classification accuracy. The study presents three models, including a handcrafted-based model and two CNN models, which completed the binary classification task. The recommended hybrid approach achieved a high F1 score of 96.12 ± 0.41, precision of 97.77 ± 0.32, and accuracy of 97.53 ± 0.24, indicating that it has the potential to serve as a valuable tool for pathologists.

KEYWORDS

feature fusion, convolutional neural network, medical imaging, radiomics feature, deep feature

1 Introduction

In medical image analysis, object identification, detection, and recognition are essential skills that are applied in several settings, such as research, treatment planning, and illness diagnosis. Significant duties in this discipline include image registration, medical condition categorization, and tumor segmentation. In medical imaging, for instance, object detection entails locating and classifying anomalies, such as tumors, within an image. This is especially difficult because different medical problems present differently based on several circumstances, such as the patient's demographics and the imaging modalities used (Alruwaili et al., 2024; Wang et al., 2024). Conventional techniques for medical image analysis frequently depended on manually created features or textural attributes. The textural elements that Haralick et al. (1973) provided for image classification have been essential in understanding the textures observed in medical images. The Scale-Invariant Feature Transform (SIFT) is developed and Lowe (1999) finds unique local features in an image that are resistant to rotation, illumination, and scale changes. By producing a histogram of gradient orientations surrounding each key point, SIFT generates a descriptor which is then used to match key points across other images. Image registration and tumor localization in medical imaging have been made easier by the use of these descriptors for the recognition and alignment of anatomical components.

Convolutional neural network (CNN)-based object identification algorithms have significantly outperformed conventional object recognition algorithms in recent years as a result of the tremendous advancement in deep learning applications. These significant improvements have been noted by the medical image analysis field (Brahimi et al., 2017; Iqbal et al., 2024). Machine that understands features from raw images have gradually replaced algorithms that employ handcrafted features. Before AlexNet's innovation, numerous alternative methods for learning features were widely used (Bengio et al., 2013).

Three basic problems are specifically encountered while analyzing medical data, and these problems are briefly discussed below. Medical data may take on a variety of forms, from images to text values, and each type requires a particular approach. This presents many challenges when it comes to combining various types of data to, for instance, make a medical diagnosis (Yue et al., 2020; Xu Z. et al., 2023). The second issue is that conventional machine learning algorithms have performed poorly when used to analyze large amounts of data, particularly when it comes to the evaluation of medical data, which includes text notes and diagnostic images which is one of the primary medical tools used to estimate and inhibit human infection. Last but not least, a lot of medical data, including genetic expressions and bio-signals, have significant levels of noise and fluctuation, which enable information retrieval particularly challenging (Brahimi et al., 2018; Wang et al., 2023). Medical images are the primary diagnostic and prognostic tool utilized by physicians, which illustrates why statistical models are required for the interpretation of such material. Deep Learning might be viewed as the new approach in medical image interpretation. Medical images are generally utilized most frequently in the medical area of neuroscience for both brain-related studies and screening practices referred to neurological illnesses (Li et al., 2019). Such images are often distinguished by a significant level of diversity that may be controlled by complicated deep learning frameworks. These applications have demonstrated highly encouraging outcomes in the evaluation of such images in this context (Lin et al., 2020; Azam et al., 2022). Radiological examinations of the chest X-ray are often performed, and huge datasets have been used extensively in research to train algorithms that combine Recurrent Neural Networks (RNNs) for text interpretation and Convolutional Neural Networks (CNNs) for image analysis (Mavakala et al., 2023). Research on nodule recognition, description, and classification in radiography and thoracic computed tomography (CT) is still ongoing. Many strategies are being investigated (Xu Q. et al., 2023), including CNNs with conventional machine learning techniques and the integration of features produced from deep neural networks. These developments are intended to help organizations more effectively identify a variety of cancers on chest X-rays. Furthermore, the field of CT research endeavors to discern textural patterns suggestive of pulmonary ailments (Chen et al., 2023; Hao et al., 2023).

The clinical contemporary healthcare system has benefited greatly from the use of medical imaging. The ability of image segmentation technology to handle massive volumes of medical images has made it a crucial tool for computer-assisted medical evaluation and therapy. The medical image classification process is more difficult than conventional object recognition because of the subtle differences and higher levels of complexity in medical images. Additionally, the medical image contains a wealth of semantic data that is essential to clinical and pathological feature representation (Mohan and Subashini, 2018; Zhang et al., 2022). Low-level collection of features, mid-level feature visualization, and deep feature learning are the three main types of visual feature research used in medical image categorization. The low-level feature extraction methodologies, such as radiomics, SIFT (Lowe, 1999), Local Binary Patterns (LBP) (Liao et al., 2009), and Color Vector Patterns (Häfner et al., 2012) often characterize the image content primarily in terms of texture, form, color, and local pixel density. These approaches have a straightforward computation and a univocal notion but often typically omit semantic detail and are unable to deliver sufficient efficiency for a single feature. The low-level feature separation techniques are further used in the mid-level feature encoding approaches to provide statistics or acquisition that can do up to some extent and convey semantic significance. The widely used Bag-of-Visual-Words (BoVW) model, for instance, is still a useful feature representation method for environment image categorization (Nosaka et al., 2011).

Radiomics is a new branch of applied research that aims to extract high-dimensional data that may be mined from clinical imaging. The radiomic procedure may be broken down into discrete phases with defined sources and outcomes, such as image capture and reconstruction, image segmentation, feature collection and validation, assessment, and model creation. To build strong and trustworthy models that may be used in clinical applications for prediction, non-invasive disease monitoring, and assessment of disease reaction to therapy, each stage must be carefully assessed (Manthe et al., 2024).

A particular medical imaging technique cannot provide important precise and accurate results. This fact drives researchers to develop new imaging technologies or suggest fusion techniques to combine data from several manual feature extraction methodologies and acquire complementary information that may be present in one or more feature extraction methods (El-Gamal et al., 2016). Even if there are several medical image fusion methods, the clarity of the combined medical image may still be raised. To improve image resolution and expand the therapeutic application of images for medical concerns, the technique known as "image fusion" combines the more important information of numerous images from one or more feature extraction methodologies (James and Dasarathy, 2014; AlEisa et al., 2022). Image fusion techniques may be used at three distinct levels: feature, outcome-based, and pixel level. The goal of feature-level fusion techniques is to take the most important aspects or prominent features from the input images such as edges, direction, shape, and size. These prominent features are combined with other extracted features. A high level of fusion that determines the real target is called outcome-based fusion. It combines the output from many methods to get a final fusion assessment. With pixel-level fusion techniques, the actual data from the input images or their multi-resolution modifications is combined immediately (Liu et al., 2015).

According to a report by The International Agency for Research on Cancer (IARC) and The Global Cancer Observatory (GCO), it is measured that in 2020, there will be 308,102 new cases of brain tumor and 251,329 deaths worldwide (Sung et al., 2021). Among the various types of brain cancer, Glioblastoma Multiforme (GBM) is considered one of the most severe and challenging forms, which is classified as a grade IV brain tumor by the World Health Organization. It accounts for 48% of all initial malignant brain tumors and is expected to affect more than 13,000 Americans annually, with over 10,000 deaths each year, as reported by Bray in 2018. Treatment options for GBM are limited (Bray et al., 2018).

The goal of our study is to ascertain whether essential radiomics features might be present in various body cells. We evaluated the radiomics features of tumors in the brain organ. We created a Fusion model containing radiomics features and CNN features (high-level semantic features) extracted from available datasets for survival evaluation. The separate experimental dataset of brain cancers was used to examine the possibility of leveraging the chosen features to separate high-risk from low-risk groups. The study proposes using CNN and radiomics-based features to enhance the effectiveness of combined findings in medical image feature fusion. The main contributions of this study are as follows:

- A new residual block called Sequential-ResNet (Seq-ResNet) is proposed, which includes five 3 × 3 convolutional layers to examine high-level semantic information. The proposed Seq-ResNet deepens the CNN network while maintaining a manageable parameter number and adds an approach to preserve moderate-level features in addition to shortcut connections.
- A FusionNet architecture with 37 layers is designed specifically for detecting smaller or low-magnification tumors, which allows for the combination of high-level semantic information with moderate-level pertinent features.
- Radiomics features can be obtained from the tumor region that has been segmented using image processing techniques. Various feature extraction methods, including gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), and gray-level size zone matrix (GLSZM), are available for this purpose.
- Feature fusion is used to extract the most discriminating information from the source feature sets and eliminate duplicate information produced by association across different feature sets.
- The statistical and spatiotemporal features in each of the various source images are extracted using an inter-extraction method, combined, and then divided into malignant and benign.

This research employs the medical imaging modality MRI, to detect brain tumors. Brain MRI is widely used for diagnosing critical diseases globally, and Section 2 describes the details and prevalence of such diseases. Section 3 reviews the most up-to-date state-of-the-art research on brain MRI. Section 4 discusses the importance and significance of this article, while Section 5 elaborates on our proposed fusion model's detailed design and implementation. The experimental setup is presented in Section 5.3. Finally, Section 6 details the results obtained from training and testing on the Brain MRI dataset, and Section 8 concludes the study.

2 Background

According to the American Cancer Society, negative consequences of GBM therapy may include peripheral neuropathy, which includes symptoms that cause effects on the central nervous structure and limit bodily activities, lowering the quality of life for patients severely. As a result, it is vital to determine if chemotherapy will be beneficial in slowing the progression of the disease before the patient begins treatment.

Surgical excision of the tumor, followed by radiation and chemotherapy, is the treatment option for GBM patients. Individual patients who get standard care have an average survival period of 15 months, relative to only 4 months if they are left untreated once identified (Bleeker et al., 2012). Chemotherapy, a common and effective therapeutic option, kills rapidly proliferating cells but cannot always ensure the difference between tumor and normal cells. This may have unfavorable consequences (Taal et al., 2015).

One of the most challenging jobs in medical imaging analysis is the automatic classification of glioma. It would be very helpful for healthcare professionals if a computational framework could be created that could detect diseases, plan treatments, and evaluate their effectiveness better than a trained and qualified one could. Such a framework would also enable a more distinct, uniform, and exchangeable method for diagnosis of diseases, care planning, and measurement. Gliomas are the most prevalent type of brain tumor in people. The appropriate classification of medical image data is a provocative medical image analysis job because of their complex structure and composition in multi-modal Magnetic Resonance Imaging (MRI). Such gliomas require feature extraction, which requires a high level of specialist knowledge, requires time, and is inclined to human misconception. The typical approach also deficiency of coherence and reproducibility, which has a negative effect on the results and may result in incorrect diagnosis and treatment.

Due to rapid advancements in machine learning and deep learning (DL) techniques, deep neural networks (DNNs) hold a lot of potential for application in Computer Assisted Diagnosis (CAD) semi-automatic systems for healthcare data interpretation. Convolutional neural networks (CNNs) have significantly advanced, enabling models to match or surpass human performance in a variety of fields, including, among others, image analysis and microscopy segmentation (Russakovsky et al., 2015).

Despite the apparent effectiveness of Deep learning models in a variety of problem scenarios, designing well-functioning deep learning models is not easy in reality. The success of a deep learning model is strongly dependent on a circumstantially appropriate selection of design factors, such as the number of hidden layers in a model, the number of units in a layer, and the kind of unit, which are referred to as hyperparameters. Different elements of the deep learning model's behavior are governed by hyper-parameters, including the model's capacity for learning patterns from images, its degree of generalization in performance when given with fresh data, and the memory consumption cost of building the classifier (Nazir et al., 2022).

As we all know, deep learning models are black boxes and we do not know about the pertinent feature extractions and also these models are data hungry. As long as enough training data are provided, deep learning models are strong contenders for brain malignancy segmentation. The Brain Tumor Segmentation Challenge (BraTS) offers a huge, eminent-quality dataset that includes MRI brain images and segmentation masks. Tumor segmentation and MGMT methylation prognosis from pretreatment magnetic resonance (MR) images are two challenges in the RSNA BraTS 2021 competition. The organizers of the challenge have published large datasets to facilitate technique assessment and advance state-of-the-art approaches in various fields (Baid et al., 2021).

Due to the tiny size of medical image segmentation datasets (typically ∼100 samples) and the lack of a universal baseline for comparing the effects of different architectural adjustments, such analyses are frequently incorrect. Nevertheless, the dataset published for BraTS21 contains 2,040 samples (in the training, validation, and test sets, respectively, 1251, 219, and 570 examples), making it the leading dataset for medical image analysis at the instant and an ideal individual for comparing performance improvements for different UNet variants.

There has subsequently been a surge in the field of information mining and artificial intelligence (AI) usage in medicine. The topic of radiomics encompasses a set of approaches for automatically extracting huge quantities of statistical data from medical images using gray-level pixel assessment, potentially paving the way for discoveries into pathophysiological processes underpinning various medical disorders (Lambin et al., 2012). One of the key fields of radiomics is texture characterization, which evaluates gray-level value variations in images that are not discernible by a human reader's aesthetic judgment. As a result, it is useful in radiography for assessing the characteristics of various tissues or organs, perhaps leading to the discovery of novel biomarkers (Scalco and Rizzo, 2017). Texture features and characteristics may have clinical and pathological associations that might aid in the assessment of patient prognosis (Lubner et al., 2017).

Deep Learning (DL) has proven to be potentially effective in a variety of healthcare sub-specialties in recent years, and many of these techniques have now been licensed for clinical usage (Cuocolo et al., 2019, 2020; Tsuneta et al., 2021). Radiology is one of the most promising domains for radiomics and machine learning applications, since they may be used to detect and characterize lesions automatically or divide medical images (Ugga et al., 2021; Spadarella et al., 2022). There have been an increasing number of research studies that indicate DL to be a valuable technique in visualizing malignant disorders (Haq et al., 2021). It might, for example, reduce the time it takes to acquire and rebuild images (Sermesant et al., 2021). The have also shown encouraging results in digital anatomical structure segmentation and illness categorization (Bruse et al., 2017; Ghorbani et al., 2020). Finally, the capacity of DL to find hidden patterns in data may bring fresh insights into well-known illnesses, boosting future management (Bagheri et al., 2021).

3 Related work

Quan et al. (2021) presents FusionNet, a deep neural network that segments neuronal structures in connectomics data obtained from high-throughput, nano-scale electron microscopy. The primary challenge of developing scalable algorithms with minimal user input is addressed with deep learning. FusionNet combines recent machine learning advancements to improve segmentation accuracy and performs well when compared with other electron microscopy segmentation techniques. The versatility of FusionNet is also demonstrated in two segmentation tasks: cell membrane segmentation and cell nucleus segmentation.

Guo et al. (2020) proposed framework for multi-modal medical image integration, which aims to maximize physiological information, improve visual clarity, and reduce computation. It consists of four parts and captures all medical information features in the input image, calculates the weight of each feature graph, and reduces information loss. The algorithm was tested on three sets of investigations with medical images, showing better performance than other algorithms in terms of detail and structure recognition, visual features, and time complexity.

The deepSeg was discovered by Zeineldin et al. (2020). They created two fundamental components that are linked by an encryption and decoding connection. To extract features, they employed a Convolutional Neural Network (CNN) as an encoder. With CNN layers, they employed dropout and Batch Normalization (BN). Then, using the SoftMax activation function, enter the result into the decoding section to generate a prediction map. They employed a Batch Normalization layer between each convolution and ReLU in the deciphering section and a modest kernel size of 32 for the base filter. They also examined the revised UNet to other CNN models including NASNet, DenseNet, and ResNet. They used FLAIR MRI images from the BraTS 2019 competition, which enclosed 336 training instances and 125 validation cases for data size of 244 × 244. Hausdroff and Dice's Distances increased from 0.81 to 0.84 and 9.8 to 19.7, respectively.

For brain tumor segmentation, Lachinov et al. (2018) identified two frames of classification techniques from the same UNet (Multiple Encoders UNet and Cascaded Multiple Encoders UNet). For brain segmentation, they employed a customized 3D UNet CNN Model. With its cost function, the proposed Cascaded UNet utilized three UNets. To improve the dataset, they employed z-score normalization as a pre-processing strategy. To expand the range of instances of the source data, they employed b-spline transformation as data augmentation. They tested the proposed two-frame classification techniques using the BraTS 2018 dataset and found that they performed well on test data. The Dice score increased from 0.901/0.779/0.837 to 0.908/0.784/0.884 for total tumor, enhanced tumor, and tumor core segmentation when the base UNet was evaluated to the Cascaded UNet.

Based on the number of references, encoder–decoder architectures, in particular UNet, are among the most often used deep learning models for medical image analysis in the field of brain tumor segmentation (Ronneberger et al., 2015). UNet-like topologies have been among the most popular BraTS competition proposals in recent years. For example, in 2018, Myronenko added a variational autoencoder branch to a UNet model for generalization (Myronenko, 2018). Jiang et al. (2019) used a two-stage UNet pipeline to partition brain tumor structural components from rough to granular in 2019. Isensee et al. (2021) reported the nnUNet architecture in 2020, with particular BraTS-designed improvements to data post-processing, region-based training, data augmentation, and minor nnUNet flow improvements (Isensee et al., 2020). These results demonstrate that well-designed UNet-based networks may execute well enough on challenges such as brain tumor segmentation. To develop a suitable solution for problems such as BraTS21, the optimum neural network design, and the training procedure must be adopted. There are many other types of UNets, such as Attention UNet (Oktay et al., 2018), Residual UNet (He et al., 2016), Dense UNet (Huang et al., 2016), Inception UNet (Szegedy et al., 2017), UNet++ (Zhou et al., 2018), SegResNetVAE (Isensee et al., 2020), or UNETR (Hatamizadeh et al., 2022), to mention a few.

We examine feature matching encoder–decoder systems from two angles: reconstructing spatial features and utilizing hierarchical semantics. The pooling mechanism in encoder–decoder networks is notorious for inducing significant systematic errors and overlooking the connection between parts and wholes. In convolutional neural networks (CNNs), max-pooling is frequently used for downsampling. The greatest value from each region is generated by max-pooling, which divides feature maps into non-overlapping parts. This results in the loss of potentially significant geographical information. Several existing strategies have attempted to modify crude high-level semantics through the use of high-level spatial resolution information. In combination with multiresolution fusion, stacked hourglass networks perform continuous bottom-up and top-down computation (Newell et al., 2016). Recent approaches append the characteristics of various layers before prediction calculation to retrieve spatial information employing encoder–decoder networks (Bell et al., 2016; Kong et al., 2016). As the input to other concurrent sub-networks, HRNet integrates the representations created by sub-networks with high-level resolution (Sun et al., 2019). Deeply fused networks employ shallow layer interim outputs as input to deeper layers (Chen et al., 2021). The global convolutional network uses skip connections with massive kernels to encode rich spatial information from input images (Peng et al., 2017).

The high-level interpretations heavily influence the outcome of an encoder–decoder network. However, feature merging is necessary to restore low-level semantics in addition to high-level spatial characteristics. To prevent unnecessary failed states that can result from increasing depth, ResNet adds low-level semantic input feature maps to high-level semantic output feature maps (He et al., 2016). In contrast, DenseNet combines hierarchical semantics with spatial information at the same level, thereby improving classification rules (Huang et al., 2017). H-DenseUNet showcases how the optimized flow of information and parameters can reduce the complexity of training encoder–decoder networks for biomedical image segmentation (Li et al., 2018).

Badrinarayanan et al. (2017) discovered a convolutional encoder–decoder network for image analysis in their study known as SegNet. The SegNet is a fundamental trainable segmentation engine that includes an encoder network, which is structurally similar to the VGG-16 network's 13 convolutional layers, and a corresponding decoder network, along with a pixel-wise classification layer akin to the deconvolution network. What sets SegNet apart is its innovative approach to non-linear upsampling in the decoder, where it employs the pooling indices obtained during the associated encoder's max-pooling phase. This eliminates the need for learning how to up-sample. To produce dense feature maps, trainable filters are used to convolve the up-sampled (sparse) maps. SegNet outperforms many of its competitors while requiring a considerably smaller number of learnable parameters. The same authors proposed a Bayesian version of SegNet to model the uncertainty in the convolutional encoder–decoder network for scene segmentation (Kendall et al., 2015).

In their groundbreaking study, Aerts et al. (2014) reported predictive ability in separate data sets of individuals with head-and-neck and lung cancer. It showed that frequently obtained CT scans may contain diagnostic and biological information. As a result, a significant amount of radiomics research has concentrated on this topic after acknowledging that tumor diversity has prognostic value and may affect therapy response (McGranahan and Swanton, 2015). The relevance of radiomics for diagnosis and prognosis and evaluating therapeutic outcomes is highlighted by research that proves radiomics characteristics and patterns mirror the cancer micro-environment in terms of behavior and progression. For patients with non-small cell lung cancer, Ganeshan et al. (2012) discovered that tumor variability may be evaluated by non-contrast CT scan texture analysis and can offer an unbiased predictor of survival (NSCLC). The same researchers' texture characteristics found pertinent relationships in a different investigation and showed that they might function as imaging correlates for cancer hypoxia and angiogenesis (Ganeshan et al., 2013). Win et al. (2013) used pretreatment Positron Emission Tomography (PET)/CT scans to study cancer heterogeneity and permeability throughout this time. According to their research, the only variable associated with survival in the group receiving drastic therapy was textural heterogeneity assessed from CT scans. Textural heterogeneity, tumor stage, and permeability were all linked to survival outcomes in the palliative treatment group. In a similar setting, Fried et al. (2014) retrieved texture characteristics from preoperative CT images before receiving final chemo-radiation therapy and discovered that radiomics features may offer predictive information further than that gained from standard prognostic markers in NSCLC patients. Based on the widely accepted idea that tumors are diverse and the degree of diversity may aid in determining the malignancy and severity of tumors, Cherezov et al. (2019) discovered a method for discovering tumor habitats using textural data. These findings showed that lung cancer patients' long-term and short-term survival rates could be distinguished with an AUC of 0.9 and an accuracy of 85% (Cherezov et al., 2019). Furthermore, previous research has shown correlations between prognosis and therapeutic response for radiomics characteristics derived from preoperative fluorodeoxyglucose (18F-FDG) PET scans. For instance, in a previous study, textural characteristics of PET scans were linked to worse prognostications and lack of response to chemo-radiotherapy by Response Evaluation Criteria in Solid Tumors (Cook et al., 2013). Decreased diversity on PET was linked to erlotinib reaction in a different research by Cook et al. (2015), and changes in first-order entropy were significantly linked to both patient survival and effective manner and response in NSCLC patients. They explain the clustering strategy using FDG-PET and CT to detect intra-tumor diversity in lung adenocarcinomas before and after therapy. To quantify the lesion structure, strength, diversity, and other characteristics in several frequencies, 583 radiomics characteristics from 127 preoperative lung nodules were retrieved in a different research. With equal-sized benign or malignant tumors, patients were randomly assigned to one of ten categories. The random forest approach was then used to run a diagnostic model. This radiomics classification algorithm successfully achieved 80.0% sensitivity, 85.5% specificity, and 82.7% accuracy in separating cancerous primary lesions from benign ones. In contrast, the sensitivity of the conventional knowledgeable radiologists' annotations was only 56.9% with identical precision (Ma et al., 2016). Another study reported how radiomics may identify the eventual development of cancer by doing quantitative analysis on preliminary low-dose CT chest lesions and analyzing images from the well-known National Lung Screening Trial (NLST). There were two lineages: one included 104 cases and 92 individuals with lung malignancies found by screening, while the other included a similar group of 208 incidents and 196 individuals with harmless pulmonary lesions found through screening. Such findings are comparable to the precision of the McWilliams framework for analysis and outperformed the accuracies of the Lung-RADS and malignancy volume methodologies (Kim et al., 2019). In total, 23 reliable radiomics features chosen by the Random Forest (RF) algorithm accurately predicted malignancy that became malignant in 1 or 2 years, with accuracies of 80% (Area Under the Curve, AUC 0.83) and 79% (AUC 0.75), respectively (Hawkins et al., 2016). Using peritumoral and intra-tumoral radiomic characteristics, Pérez-Morales et al. (2020) created an independent lung cancer prognostic prediction model (Pérez-Morales et al., 2020). This algorithm might pinpoint a subset of individuals with initial lung cancer who are at significant risk and have a bad outcome. The lung cancer screening system allowed doctors to customize clinical care for these high-risk individuals who were diagnosed with lung cancer in its early stages. Furthermore, Horeweg et al. (2014) identified that determining the radiomic capacity doubling time for medium lesions helped direct lung cancer care and perhaps forecast the likelihood of lung cancer (Horeweg et al., 2014). Lesion treatment strategies that use volumetric or volume-based diameter boundaries (ranging from 9 to 295 mm^3 or 6–11 mm in diameter) have demonstrated enhanced sensitivity of up to 92.4% and specificity of 90.0% as opposed to the ACCP lesion handling procedure using low-dose CT scans in target populations. By lowering false positives and negatives in lung cancer analysis and misdiagnoses, more research into radiomics applications could improve lung cancer screening. Constanzo et al. (2017) has generally concentrated on hand-crafted Radiomics, while deep learning-based Radiomics is only briefly discussed without addressing various topologies, interpret-ability, and hybrid models. While Parmar et al. (2018) discovered both forms of Radiomics, the combination of hand-crafted and deep learning-based characteristics is not taken into account. Additionally, the difficulties with radiomics and the connection between radiomics and gene expression (radio-genomics) are not fully covered. Finally, only deep learning-based radiomics features are covered in the study by Litjens et al. (2017), leaving out hand-crafted features, their stability, hybrid radiomics, and radio-genomics. All of them necessitate a quick and timely effort to expose radiology to your community, as image processing is one of the fundamental elements of radiology.

The entropy of metastatic disease was proven to be a relevant measure in an MRI-based radiomics investigation; greater entropy values were discovered in tumor tissues relative to mild tumors, indicating the tumor's diversity and vascular state (Parekh and Jacobs, 2017). Using dynamic contrast-enhanced MRI, another research (Whitney et al., 2019) attempted to develop a collection of quantitative parameters that might be retrieved from MR images to differentiate luminal breast tumors from mild breast tumors.

4 Our contributions

When Deep Convolutional Neural Networks (DCNNs) utilize feature fusion for retrieving spatial information and leveraging multi-layer semantics, two issues arise. First, the deep convolution layer feature maps provide lower-level spatial information required for reconstructing the merged feature maps. Second, feature-matching methods only provide feature maps' semantics at the same level of resolution. These problems are challenging to address as element-wise addition and channel concatenation result in a fusion method that is overly restrictive and only aggregates extracted features of the same scale. The encoder and decoder convolution layers are the only ones at the same level in encoder–decoder networks because downsampling reduces the scale of feature maps, and upsampling increases it. UNet lacks multi-layer semantics and global spatial information, which results in analyzing images pixel by pixel and distinguishing objects using color contrast. However, using contrasting colors to make objects stand out may not necessarily improve tumor borders.

The purpose of our investigation is to determine whether fundamental radiomics traits may be found throughout several human tissues. In the brain organs, we assessed the radiomics characteristics of malignancies. Using the dataset for survival analysis, we built a radiomics model and selected characteristics. The potential for employing the desired traits to distinguish between high and low-risk groups was investigated using an independent test dataset BraTS brain tumors.

To achieve automation with manual medical image feature fusion, we use CNN and radiomics-based features to upgrade the effectiveness of the combined findings. The main contributions of this study are as follows:

1. Sequential-ResNet (Seq-ResNet) is a new residual block, which is proposed. Five 3 × 3 convolutional layers are included in a single Sequential-ResNet to examine high-level semantic

```
Require: input_image, radiomics_features
▷ Seq-ResNet Layer
function Seq_ResNet(input_image):
    ▷ Sequential Residual Architecture(x)
    ▷ Extract feature subspaces using Seq-ResNet layer
    return feature_map with multiple subspaces
▷ Down-sampling Operation
function down_sampling(input_high_res_feature_map):
    ▷ Transpose Convolution to obtain low-resolution features
    low_res_feature_map ← Transpose_Convolution (input_high_res_feature _map)
    ▷ Deconvolution to produce high-resolution features
    high_res_feature_map ← Deconvolution(low_res_feature _map)
    ▷ Fusion of high-resolution features with residual output
    final_high_res_feature_map ← Concatenate (input_high_res_feature_map, high_res_feature_map)
    return final_high_res_feature_map
▷ Up-sampling Operation
function up_sampling(input_low_res_feature_map):
    ▷ Deconvolution to produce high-resolution features
    high_res_feature_map ← Deconvolution(input_low_res_feature_map)
    ▷ Transpose Convolution for low-resolution features
    low_res_feature_map ← Transpose_Convolution (high_res_feature_map)
    ▷ Fusion of low-resolution features with residual output
    final_low_res_feature_map ← Concatenate (input_low_res_feature_map, low_res_feature_map)
    return final_low_res_feature_map
▷ Main FusionNet Process
seq_resnet_features ← Seq_ResNet(input_image)
down_sampled_features ← down_sampling (seq_resnet_features)
concatenated_features ← Concatenate (down_sampled_features, radiomics_features)
fc_layer_output ← Fully_Connected_Layer (concatenated_features)
classification_result ← SVM_Classification(fc_layer_output)
return classification_result
```

Algorithm 1. Proposed FusionNet.

information. A hierarchical framework made up of these five layers generates features that are then analyzed by different numbers of convolutional layers. The proposed Seq-ResNet deepens the CNN network while maintaining a manageable parameter number as compared with the original ResNet. The proposed Seq-ResNet adds an approach to preserve the moderate-level features in addition to the shortcut connection increasing the outcomes.

2. We design a Seq-ResNet for the detection of a smaller or low-magnification tumor. This CNN architecture comprises 37 Seq-ResNet layers. The proposed model having a deeper architecture allows the model to capture more intricate and complex high-level semantic features from the input data. Additionally, the pairing of numerous Seq-ResNet preserves the moderate-level features necessary for smaller tumor recognition while allowing the resulting features to be digested by varied numbers of convolution layers. As a result, the feature produced by the proposed model combines high-level semantic information with moderate-level pertinent features.
3. The essence of feature fusion lies in its ability to extract the most discriminating information from the source feature sets engaged in fusion and get rid of the duplicate information produced by association across different feature sets. The statistical and spatiotemporal features in each of the various source images are extracted using an inter-extraction method. The source images' features are combined, and the combined result is divided into malignant and benign.

5 Methodology

5.1 FusionNet

The natural contour of the body edge is changeable in medical images because distinct features of the human body are visible. To handle the separation of microorganisms, a structure has been developed that uses sophisticated models which amalgamate a high-level representation of depth-wise convolution layers and residual blocks with a description of the presence of upsampling and downsampling levels to achieve detailed segmentation. The original residual block is suggested as a solution to the vanishing gradient issue that arises as network complexity grows. Networks with different depths have been developed to effectively investigate high-level contextual information by layering numerous residual blocks. Low-level precise information is observed as being as significant as high-level contextual information in the case of tiny feature extraction. Although the auxiliary link in the initial residual block aids in deepening the network, convolutions still lead to low extensive information vanishing. The 3×3 convolution layer in the original residual block is replaced with a sequential ResNet convolution structure to allow for simultaneous acquisition of low-level specific information and high-level contextual information. Proposed Seq-ResNet is the term given to the res-block since the input feature is separated and analyzed hierarchically, which is similar to mounting a mountain.

The Proposed FusionNet algorithm is shown in Algorithm 1 and Figure 1 and use of Seq-ResNet is shown in Figure 2 to obtain

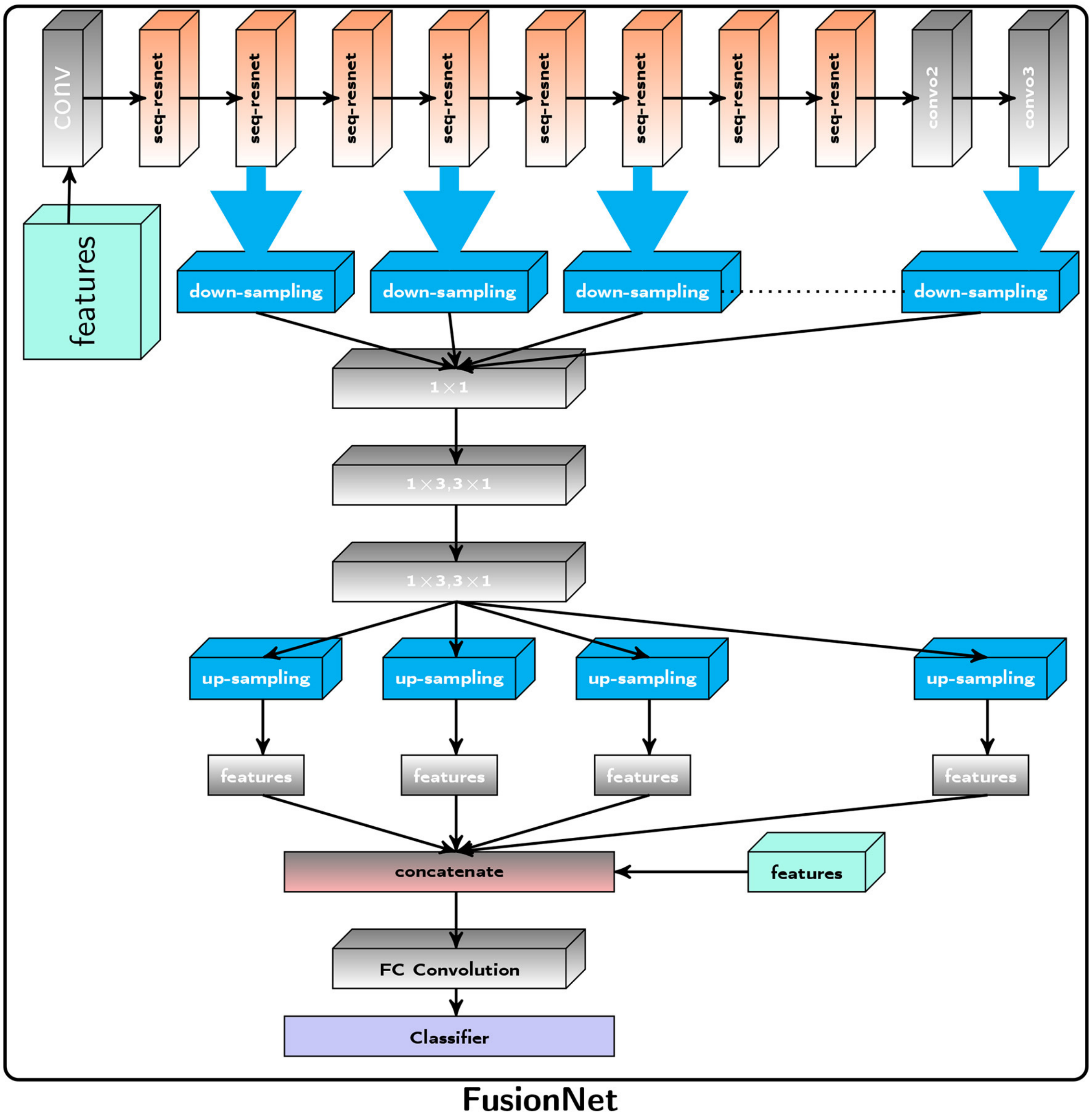


FIGURE 1
Proposed Framework: The Proposed FusionNet, where seq-resnet layer is shown in Figure 2, extracts feature subspace (3 × 3), represents the number of convolution layers, the subspace has gone through, and j represents the subspace number. The extracted features via Seq-ResNet are forwarded for downsampling steps and 1 × 1 convolution to 1 × 3, 3 × 1 depth-wise separable convolution, respectively. After that, we apply upsampling to get the pertinent features, and these features are concatenated with radiomics features and apply the FC layer. These concatenated features are passed to the Support Vector Machine (SVM) for classification.

feature subspace, where j is the subspace number and (3 × 3) is the size of convolution layers the subspace has undergone. The Seq-ResNet-extracted features are then sent forward for downsampling steps and 1 × 1 depth-wise separable convolution with 1 × 3 and 3 × 1, respectively. The relevant characteristics are then obtained by using upsampling, and these features are combined with radiomics features before applying an FC layer. Support Vector Machine (SVM) is given these concatenated characteristics to classify the data. The proposed CNN model FusionNet comprises Seq-ResNet and is made up of two parts:

5.1.1 Sequential residual network

Figure 2 depicts the proposed Seq-ResNet, where $\mathbb{F}_i^j$ stands for a feature subspace (i [3 × 3] for the number of convolution layers the subspace has undergone and j for the subspace number). $\mathbb{F}_i$ stands for a feature map constructed from feature subspace (i indicates the number of this feature map). The symbol $\mathbb{C}_i^j$ denotes a convolution layer with j depicting the kernel value and i presenting the number of layers. The input image of Seq-ResNet is convolved with kernel 1 × 1 to decrease the number of channels into $\frac{1}{5}$. These number of channel outputs such as $\mathbb{F}_1^1$, $\mathbb{F}_2^1$, $\mathbb{F}_3^1$, and $\mathbb{F}_4^1$ are convolved with kernel value 3 × 1 and 1 × 3, except the $\mathbb{F}_0^1$. The remaining feature ($\mathbb{F}_0^1$) is processed-free overlaid to final feature map ($\mathbb{F}_i$). We then pass $\mathbb{F}_2^1$, $\mathbb{F}_3^1$, and $\mathbb{F}_4^1$ to depth-wise seperable convolution with kernel value 3 × 1 and 1 × 3, and the remaining feature $\mathbb{F}_0^1$ and $\mathbb{F}_1^1$ forward to final feature map ($\mathbb{F}_i$). The detailed operation is shown in Figure 2 Seq-ResNet block. At the end, we concatenate all the processed and processed-free features into a final feature map $\mathbb{F}_i$. To restore the original channel number, we

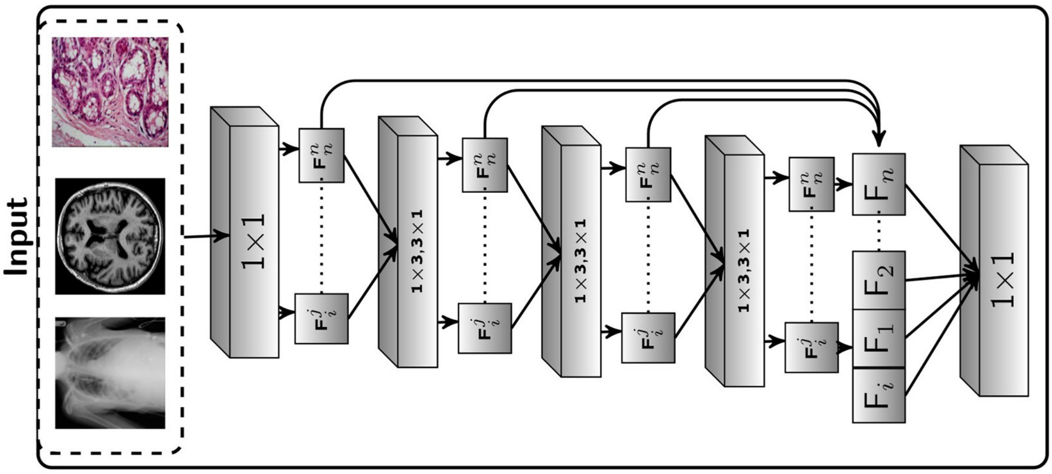


FIGURE 2
Proposed Framework: The suggested Seq-ResNet, where $\mathbb{F}_i^j$ represents a feature subspace, i(3 × 3) represents the number of convolution layers the subspace has gone through, and j represents the subspace number. $\mathbb{F}_i$ denotes a feature map built from feature subspace (i indicate the number of this feature map). The symbol $\mathbb{C}_i^j$ represents a convolution layer, with *j* representing the kernel value and *i* representing the number of layers. The Seq-ResNet input image is convolved with kernel 1 × 1 to reduce the number of channels to $\frac{1}{5}$. These numbers of channel output such as $\mathbb{F}_1^1$, $\mathbb{F}_2^1$, $\mathbb{F}_3^1$, and $\mathbb{F}_4^1$, are convolved with kernel value 3 × 1 and 1 × 3 except the $\mathbb{F}_0^1$.

process the concatenated feature map with convolution with a kernel value is 1 × 1. Multiple subspaces of the result are analyzed using various convolutional layers, resulting in various subspaces that include features with various receptive fields. Smaller receptive field subspace, such as $\mathbb{F}_0^1$ and $\mathbb{F}_1^1$, comprise more relevant details and undergo smaller convolution stages, which is crucial for smaller tumor recognition. Large receptive field subsets, such as $\mathbb{F}_2^2$, $\mathbb{F}_3^3$, and $\mathbb{F}_4^4$, blur particular information while exploring deep contextual features is equally crucial for detection. At the start and conclusion of the suggested block, two $\mathbb{C}_1^3$ and $\mathbb{C}_2^3$ are utilized to automatically choose the appropriate features and apply a bottleneck to lower the parameter count.

5.1.2 Upsampling and downsampling

Typically, a compilation of feature maps is created by sampling and concatenating the information from various levels of the proposed Seq-ResNet. To construct a feature selection procedure, several 1 × 1 convolutions are applied during the gathering of feature maps. After feature extraction, the gathered features are merged to create feature maps with various resolutions, which are then put through distinct upsampling and downsampling procedures. Throughout the upsampling and downsampling operations, both high and low contextual information from the proposed feature fusion network can be used. However, because of these processes, information impurity might happen. Our suggested feature fusion network uses a deep upsampling and downsampling deConvolution layer to address this problem. This lessens the effect of imperfect information. This methodology was motivated by the deep upsampling and downsampling units found in super-resolution image reconstruction methods. Figure 3 depicts the proposed deep upsampling and downsampling deConvolution module's structure. The two main components, upsampling and downsampling, are used to extract low-level and high-level contextual features. The upper layer of Figure 3 presents the upsampling methodology. The low-resolution feature maps (LowRes_i) are passed to deConvolution to produce the high-resolution features (HiRes_i), as shown in Figure 3. The Transpose Convolution (Tconv) is used to extract pertinent features from low resolution (LowRes_i) to convert it into high resolution (HiRes_i). The low resolution (LowRes_2) feature map residual outcome and initial low resolution (LowRes_1) feature map are passed to deConvolution to obtain high resolution (HiRes_i) feature maps. The final high resolution residual output is obtained from the initial high resolution feature map (HiRes_1), and the output of the Transpose Convolution is concatenated in final high-resolution feature maps (HiRes_n). The lower layer of Figure 3 portrays the downsampling technique. Initially, we pass high resolution (HiRes_i) feature maps to Transpose Convolution to acquire low resolution (LowRes_i). These residual outcomes are passed to deConvolution to get the feature maps of high resolution (HiRes_i). The initial high resolution and the residual output of the second Transpose Convolution which are applied on second high-resolution feature maps are fused. After concatenation, we apply Transpose Convolution to generate the low-resolution feature maps (LowRes_i). These low resolution feature maps are convolved with Transpose Convolution and then fused with the initial low resolution feature maps. After fusion, we apply Transpose Convolution to achieve low-resolution residual feature maps.

The term "handcrafted features" in our study refers to certain traits or attributes that are retrieved from raw data by following predetermined rules or algorithms. Other names for these characteristics include "handmade features" and "hand-engineered features." They are thoughtfully created and defined by researchers or domain specialists based on their comprehension of the data and the particular issue being addressed. In our study, we use a set of statistical and textural metrics that we extract from medical tomography images as these "handcrafted features." The choice of these metrics is an important step since they help obtain vital

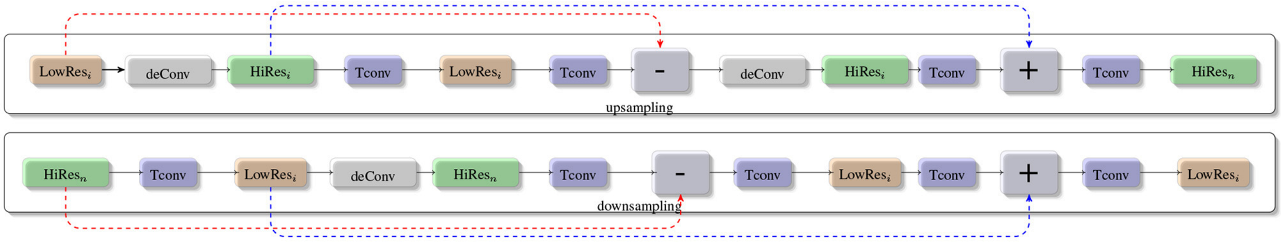


FIGURE 3
The low resolution feature maps (LowRes*i*) are passed to deConvolution to produce the high resolution features (HiRes*i*). The Transpose Convolution (Tconv) algorithm is used to extract relevant features from low resolution (LowRes*i*) images and convert them to high resolution (HiRes*i*). The residual outcome of the low resolution (LowRes2) feature map and the initial low resolution (LowRes1) feature map is passed to deConvolution to produce high resolution (HiRes*i*) feature maps. The final high resolution residual output is obtained by concatenating the output of the Transpose Convolution with the output of the initial high resolution feature map (HiRes1). In the downsampling technique, we send high resolution (HiRes*i*) feature maps to Transpose Convolution to obtain low resolution (LowRes*i*). These residual results are deconvolved to produce high resolution feature maps (HiRes*i*). The preliminary high resolution feature maps are fused with the residual output of the second Transpose Convolution, which is applied to the second high resolution feature maps. Following concatenation, we use Transpose Convolution to create low resolution feature maps (LowRes*i*).

information about the texture and characteristics contained in the images.

These hand-crafted features include metrics derived from various statistical approaches and texture analysis methods, gradient-based features, and other characteristics. The Gray Level Co-occurrence Matrix (GLCM), the Gray Level Run Length Matrix (GLRLM), and the Gray Level Size Zone Matrix (GLSZM) are three well-known matrices that we specifically add to our features. With the aid of these matrices, we can record vital details on the connections between pixel values and their spatial distribution within the images.

5.1.3 Radiomics features

The qualities of radiomics features, such as histograms, textures, shapes, transformations, and models, may be categorized numerically (Benoit-Cattin, 2006). Radiomics features may be retrieved from either 3-dimensional (3D) Volumes of Interest (VOIs) or 2-dimensional (2D) Region of Interest (ROIs). We used ROI as a catch-all word for both to make the text easier to read. Additionally, gray-level brightness that has not been changed or discretized can have numerical features estimated. Gray-level discretization and feature importance aggregation are not covered in this article since they fall outside of its purview. Gray-level discretization limits the range of gray levels to a predetermined number to improve reliability and manageability, whereas feature significance aggregation obtains a single value when the same feature is recognized in various forms and simplifies it using average values. Gray-level variance, minimum, maximum, mean, and percentiles are the most basic statistical variables that are derived on the global gray-level histogram (Zwanenburg et al., 2020). First-order features are those that are dependent on single-voxel or single-pixel analysis. Skewness and kurtosis are more complex characteristics as depicted in Equation (1) which describe the brightness distribution of the data. Skewness must represent the leftward or rightward asymmetry of the data distribution curve (negative skew, below the mean) (positive skew, above the mean). Kurtosis tends to reflect the tailedness of a distribution of the data compared with a Gaussian distribution as a result of anomalies. Other characteristics include energy also called homogeneity and histogram entropy. These are distinct from their corresponding co-occurrence matrix models of the same name.

By measuring the intensity variations in gray levels throughout an image, studying the absolute gradient provides a straightforward method for characterizing genuine radiomics textures. When two adjacent pixels or voxels have the same color, the gradient stays at zero, and it reaches its maximum when one is black and the other is white or vice versa. Similar to histograms, gradients are subjected to statistical features including variance, skewness, kurtosis, and mean, regardless of the direction of the gray-level transition (Benoit-Cattin, 2006; Zwanenburg et al., 2020). To balance the ratio of the dataset, we apply preprocessing techniques and data augmentation methodology before retrieving radiomics features.

$$
\begin{aligned}
\mathrm{Sk} &= \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)^3 \\
\mathrm{K} &= \frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}
\end{aligned}
\tag{1}
$$

In their research, Haralick et al. (1973) established the second-order gray-level histogram feature known as the Gray Level Co-occurrence Matrix (GLCM), as shown in Equation (2). The spatial associations between pairs of pixels or voxels with defined distances between them, predetermined gray-level intensities, and numerous directions such as vertical, horizontal, or diagonal-for a 2D analysis and 13 directions for a 3D analysis are captured by GLCM. Entropy, which shows gray-level inhomogeneity or randomness, angular second moment, uniformity or energy, and contrast, draws attention to the gray-level differences between adjacent pixels or voxels that are some of the qualities that make up GLCM.

$$
\mathrm{GLCM}(i,j,d,\theta) = \sum_{x=1}^{N}\sum_{y=1}^{M}\begin{cases}1, & \text{if } I(x,y)=i \text{ and } I(x+d,y+\theta)=j \\ 0, & \text{otherwise}\end{cases}
\tag{2}
$$

Galloway (1975) proposed the Gray Level Run Length Matrix (GLRLM), as shown in Equation (3), which is intended to record the spatial distribution of succeeding pixels in one or more directions, as well as in two or three dimensions. Many elements of GLRLM are included, such as fraction which examines the percentage of pixels in the Area of Interest. The availability of short and long runs is shown, respectively, by the weighted measurements known as short- and long-run emphasis inverse moments. Measures that evaluate the dispersion of runs over various gray levels and run lengths, respectively, are run-length non-uniformity and gray-level non-uniformity.

$$\mathrm{GLRLM}(i,j,d,\theta) = \sum_{x=1}^{N}\sum_{y=1}^{M}\begin{cases}1, & \text{if } I(x,y) = i \text{ and } I(x+d, y+\theta) = j\\ 0, & \text{otherwise}\end{cases} \tag{3}$$

An effective statistical method for characterizing textures is the gray-level size zone matrix (SZM), as shown in Equation (4), which was first presented by Thibault et al. (2013). The Gray Level Size Zone Matrix (GLSZM) uses the infinity norm to quantify the gray-level zones in an image. These are the areas where linked voxels have the same gray-level intensity within a given distance, usually 1. Texture homogeneity increases SZM's size and smoothness. While SZM does not require multidimensional calculations such as RLM and COM do, its efficacy is dependent on gray-level compression, therefore the best way to use it is to test out various compression techniques on training datasets.

$$\mathrm{GLSZM}(i,j) = \sum_{x=1}^{N}\sum_{y=1}^{M}\begin{cases}1, & \text{if } I(x,y) = i \text{ and the size of the zone containing } i \text{ is } j\\ 0, & \text{otherwise}\end{cases} \tag{4}$$

Based on their importance in collecting various facets of textural and structural information in medical tomography images, we chose these characteristics. Our objective was to achieve a balance between removing useful characteristics and avoiding too much dimensionality, which might result in overfitting and more complicated computations. The use of these elements enhances the overall efficacy of our suggested approach for image classification and enables us to provide a meaningful representation of the textures contained in the images.

5.2 Dataset

For our study, we utilized the MRI dataset provided by the Brain Tumor Segmentation (BraTS) Challenge in 2020 and 2021. The BraTS 2020 Challenge dataset is used to produce a separation model that could detect the malignancy region. The dataset comprised 369 MRI images captured in four distinct modalities, T2-weighted (T2w), T1-weighted (T1w), fluid-attenuated inversion recovery (FLAIR), and T1-weighted contrast-enhanced (T1wCE). These images and extraction patterns were provided in NIfTI format with coronal orientation. The masks provided four classifications, including non-tumor, non-enhancing tumor core, peritumoral edema, and enhancing tumor. As we were only interested in the broad tumor area, we combined the last three types for our analysis.

The Radiological Society of North America (RSNA) and the Medical Image Computing and Computer Assisted Intervention (MICCAI). Society expanded the MGMT promoter methylation detection component of BraTS challenge in 2021. A pre-selected collection of 585 MRI images from 2020 in almost the same four modalities was made available. To represent a broad spectrum of healthcare practices used across the world, the images were collected from several institutions utilizing a range of tools. For the classification task, we used these datasets, which were in DICOM format and annotated with their methylation status. The methylation pattern of MGMT was confirmed by laboratory analysis of surgical brain tumor tissues. The four modalities were kept the same as in the earlier datasets, but the T1w scans were not utilized since the diameters were continuously enormous, resulting in much more chaos than information.

5.3 Experimental setup

The proposed feature fusion model uses the Keras, PyTorch package with the TensorFlow backend and is entirely written in Python. Experiments are carried out utilizing MRI slices with a resolution of 256 × 256 to evaluate all of the suggested feature instrument networks. The cost function's relationship to its parameters must be optimized using stochastic scaling for the BraTS dataset of convolutional neural networks (CNNs). We utilized the adaptive moment (Adam) estimator for parameter estimation. Adam typically uses the first and second moments of the gradients to update and fix the linear trend derived from the real gradients. The settings for our Adam optimizer are as follows: 150 epochs is the maximum allowed, and the learning rate is 0.0001. With all biases set to 0, all weights are normally distributed with a mean of 0 and a variance of 0.01.

5.4 Dataset preprocessing

For each sample in the BraTS21 dataset, four NIfTI files containing different MRI modalities are provided. During the initial phase of data pre-processing, these modalities were stacked, resulting in an input tensor of shape (4, 240, 240, 155) in the (C, H, W, D) layout, where C represents the channels, H represents the height, W represents the width, and D represents the depth. Next, the redundant background voxels (with a voxel value of zero) at the edges of each volume were removed, as they did not provide any useful information and could be eliminated by the neural network.

The standard deviation for each channel was then calculated independently within the non-zero zone for each image. To normalize all volumes, the mean was first subtracted, and then, the standard deviation was divided. The background voxels were left unnormalized, and their measure remained at the cardinal. To differentiate between normalized voxels, which had numbers close to zero, and background voxels, an extra source channel was created using one-hot encoding for foreground voxels and then combined with the input data.

Image segmentation is frequently completed as part of the image enhancement activity. The initial step in comprehending an image is to improve it. We used three datasets in the empirical results section: TCIA, FIGSHARE, and BraTS 2019. As a result, we

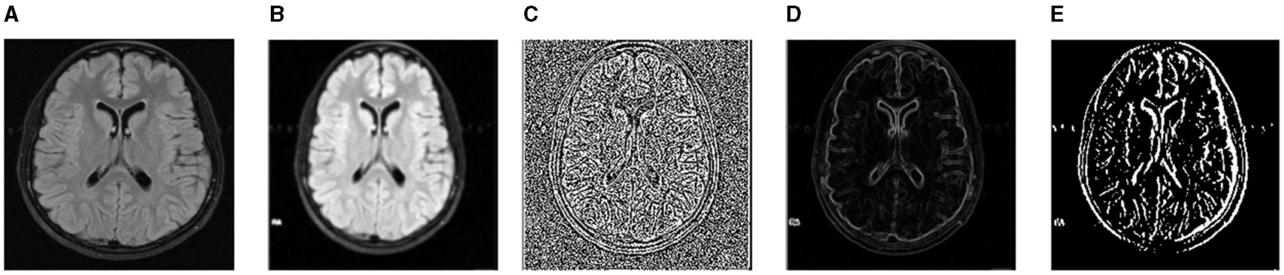


FIGURE 4
For noise removal, we apply different preprocessing techniques and acquire noise-free images of the BraTS dataset. **(A)** Input image. **(B)** SWMF Filter. **(C)** Laplacian. **(D)** Sobel. **(E)** Gabor.

will go over the pretreatment steps for each dataset in this section. We used initial image improvement approaches in the TCIA dataset: noise reduction and contrast adjustment. The artifacts caused by the imaging approach are attempted to be reduced using preliminary processing procedures. Additionally, noise reduction is a competent way to enhance outcomes before analysis (e.g., edge detection on image). It should be noted that the images used are gray-scale. We utilized two filters to remove noise: a median filter and a soft filter. After that, we will show you how to use the two filters. The median filter is a type of non-linear digital filter. It is a technique for removing unwanted signals or noise from an image. The Soft Weighted Median Filter (SWMF) is a novel image-processing approach for removing noise. Two noisy images are processed using this filter. The first is constant value noise, which is similar to salt and pepper noise in that its value does not vary. The second type is Random Value Noise (RVN), which is a sort of arbitrary value noise that has a variable value, similar to Gaussian and Speckle noise. The outcomes of the preprocessing steps are shown in Figure 4. The preprocessing techniques were only applied to the BraTS dataset.

The batch normalization approach is used on the FIGSHARE and BraTS datasets. Convolutional neural network (CNN) training is a complex task with several issues. Batch normalization is one of the most used methods for dealing with this problem. It is a widely used strategy in the field of deep learning. Batch normalization accelerates neural network learning. Moreover, batch normalization provides regularization, preventing over-fitting.

Data augmentation is a technique for avoiding over-fitting by artificially growing datasets during the learning stage. The essential data augmentations were used throughout the learning phase to strengthen our method:

1. Flips: the volume of each axis was reversed with a probability of 0.5 for the x, y, and z axes individually.
2. Gaussian Blur: The deviation of the Gaussian Kernel is obtained periodically from (0.5, 1.5) with a probability of 0.15, subjecting the reference volume to Gaussian fading.
3. Brightness: at a frequency of 0.15, a random number is regularly selected at random from the range (0.7, 1.3), after which source volume voxels are increased by it.
4. Zoom: the image size is increased to its initial dimensions twice the chosen value using cubic interpolation, and the source data are scaled using nearest neighbors interpolation. An arbitrary value is regularly gathered from (1.0, 1.4) with a frequency of 0.15.
5. Gaussian noise: every voxel is captured, and the source volume is then filled with randomized Gaussian noise with an average of zero and variance regularly selected from the range of (0, 0.33) with a probability of 0.15.
6. Contrast: at a frequency of 0.15, an arbitrary value is uniformly obtained from (0.65, 1.5). It is then enhanced, and the source volume voxels are trimmed to the value of the original range.
7. Biased crop: From the source volume, a piece with the dimensions (5, 128, 128, 128) was arbitrarily selected. Moreover, a probability of 0.4 ensures that some prominent voxels (with true positives in the underpinning data) will be retained in the trimmed area of the patch chosen using arbitrary-biased crop.

5.5 Multiple kernel for classification

For classification techniques, monitored learners such as the Support Vector Machine (SVM) method are utilized. We used an expanded form of SVM that includes various kernel training. The fundamental SVM operates as such. The content is first divided into binary categories. Next, a hyper-plane is located which distinguishes between both categories. Support vectors seem to be the parameters that seem to be close to the hyper-plane and are scientifically described in Equations (5–7):

$$(m, n), ...(m_i, n_i), ...(m_j, n_j);\ m_i \in R^N, n_j \in \{-1, 1\} \tag{5}$$

The below expression can be used to represent the hyper-plane that categorizes a particular set of information as being linearly distinguishable. The Maximum Dividing hyper-plane is the name of this hyper-plane.

$$f(x) = \sum_{i=1}^{j} (\alpha_i n_j (X_i^T X) + b) \tag{6}$$

Provided below is a depiction of the ideal hyperplane accompanied by the support vectors m_1, m_2, and m_3 which are located on its edge.

$$g(\bar{X}) = \bar{W}^T \bar{X} + b \tag{7}$$

A variety of techniques for traditional machine learning are used by numerous kernel systems. A predetermined collection of kernels is included in each approach. The kernel function allows the SVM to transform the data into a higher-dimensional space, facilitating the separation of data points through both linear and nonlinear decision boundaries. This strategy lessens the influence of bias throughout the training experience.

For the overall classification of BraTS, the precision and accuracy statistics of the best kernel operations are accumulated for all characteristics. Whereas the polynomial kernel is reliable for classifying radiomics-selected features, three different feature configurations (GLCM, GLRLM, and GLSZM) are better identified using the linear kernel function in SVM. Table 1 shows the top-chosen kernel technique for the BraTS sample together with assessment precision of the kernel function.

By using the SVM classifier on the BraTS, TCIA, and FIGSHARE datasets, we subsequently trained and validated UNet, VNet, and UNet^{++} on a classification task that incorporated.

In five-fold cross-validation, the classification algorithm we created using a Support Vector Machine with different Kernels yielded an estimated ROC AUC of 96.4 ± 0.43%. On a fusion features classification assignment, we trained different pre-trained CNN models such as AlexNet, SqueezeNet, VGG16, InceptionV3, Xception, UNet, VNet, and UNet^{++}, as well as a workflow using the categorization predictions of these models against a professional investigator. The pipeline increased recall (93.4 ± 0.5% vs. 95.4 ± 0.7%) without noticeably reducing accuracy (81.5 ± 0.4% to 95.7 ± 0.5%).

Our findings show that UNet and other similar networks such as VNet and UNet^{++} continue to have large false positive rates, which may preclude their application in healthcare situations. We showed that adding a different classifier significantly increases accuracy. We are aware that this is simply one of many viable options; future advancements to UNet could eliminate the requirement for a different classifier.

6 Result and discussion

This is the first study that, as far as we know, combines radiomic analysis for medical imaging with deep neural network installation. Our findings show that utilizing transpose convolution for both up and downsampling, the integration of Seq-ResNet architecture greatly enhances the ability to identify cancerous slices in brain MRI images. This new mix of customized imaging biomarkers and powerful deep learning approaches boosts model performance on the BraTS dataset, even with a smaller patient cohort compared with earlier research. We carried out effective image preprocessing to guarantee reliable and repeatable deep learning execution. To be more precise, we scaled each image to a 256 × 256 grid and normalized it to 256 gray levels. The repeatability of our findings is aided by these common digital image processing procedures. Furthermore, we carried out a thorough assessment of the pre-trained CNN models' dependability. The promise for enhanced medical image analysis is highlighted by this fusion of complex neural network topologies with exact image preprocessing techniques, opening the door to more accurate brain tumor diagnosis.

Average Precision (AP), mean Average Precision (mAP), and F1 are utilized in the trials to assess the effectiveness of the suggested framework. Finding objects and classifying them into multiple classes is the basic goal of image analysis. The assessment metrics for these two tasks are recall (abbreviated as R) which may be stated as "the percentage of the appropriate tumors identified to all tumors" and precision, "the accurate rate of the categorization of identified tumors" (abbreviated as P). The terms False Positive (FP), False Negative (FN), True Positive (TP), and True Negative (TN) are used to characterize these metrics. The calculation for Precision (P) and Recall (R) is shown in Equation (8):

$$\begin{aligned} R &= \frac{TP}{TP+FN} \\ P &= \frac{TP}{TP+FP} \end{aligned} \tag{8}$$

The accuracy and recall metrics might tend to be in conflict when measuring the results of several architectures. Additionally, a single index must be used to determine accuracy of a classifier. Precision and Recall create a rectangle-coordinated graph using Precision and Recall as the coordinates after being ordered by grading value. Precision-Recall curve is the name of this rectangle-shaped graph. The area underneath the Precision-Recall curve or AP is the average of APs across many classes or mean Average Precision (mAP). The Equation (9) mAP evaluates the classifier's performance across all classes, AP evaluates the classifier's performance across each category. As a result, in this study, AP is employed when goals come from a single class, whereas mAP is utilized when criteria come from a variety of categories. Another often-used indication for object detectors is F1-Measure (sometimes called F1-Score). Recall (R) and Precision (P) are weighted averaged to get the F1-Measure. F1-Measure is referred to as F1 when it is equal to 1:

$$\begin{aligned} F1 &= \frac{(\alpha^2+1)\times(P\times R)}{\alpha^2(P+R)} \\ F1 &= \frac{2\times(P\times R)}{P+R} \\ mAP &= \frac{1}{N}\sum_{i=1}^{N} AP_i \end{aligned} \tag{9}$$

A higher F1 score denotes greater classifier efficiency. The F1 score is calculated by combining the Precision and Recall values. We also offer the Recall and Precision values of the suggested approach on several datasets to give a thorough evaluation. Within a transfer learning framework, this study uses eight popular Convolutional Neural Network (CNN) pre-trained models. Since VGG16 has a relatively modest number of learnable parameters under the transfer learning framework and is widely used in medical image evaluation tasks, it was first selected because it required less computing power for network training than other well-known models.

For the proposed radiomic-based FusionNet architecture, we also looked at AlexNet, SqueezeNet, VGG16, InceptionV3, Xception, UNet, VNet, and UNet^{++}. In contrast, the efficiency after the feature fusion architecture was greater in the UNet pre-trained model than in the VNet model. This difference can be

TABLE 1 We report the results of employing a Support Vector Machine (SVM) on the BraTS, TCIA, and FIGSHARE datasets to classify particular radiomics characteristics (GLCM, GLRLM, and GLSZM) in terms of sensitivity, specificity, precision, accuracy, and F1-score.

Dataset	Feature set	Sensitivity	Specificity	Precision	Accuracy	F1-Score
BraTS	GLCM	73.16	74.31	75.91	75.14	74.93
	GLRLM	71.36	71.85	72.34	72.76	73.44
	GLSZM	73.69	71.94	73.49	72.35	73.11
	Average	72.74	72.7	73.91	73.43	73.83
TCIA	GLCM	75.20	73.80	74.50	74.75	74.25
	GLRLM	74.40	72.90	73.80	73.15	73.60
	GLSZM	73.60	75.10	74.90	74.35	74.75
	Average	74.40	73.60	74.07	74.08	74.20
FIGSHARE	GLCM	74.71	76.05	77.13	76.58	76.81
	GLRLM	72.85	73.42	74.12	73.64	73.98
	GLSZM	76.07	74.91	75.94	75.49	75.72
	Average	74.04	74.46	75.06	75.24	75.50

Included are the average performance metrics for every dataset.

TABLE 2 Training accuracy, F1-Score, specificity, precision, and sensitivity on the BraTS, TCIA, and FIGSHARE datasets of the proposed model (FusionNet with ResNet, FusionNet with Seq-ResNet, and FusionNet with Seq-ResNet and radiomics features) and other pre-trained CNN models such as AlexNet, SqueezeNet, VGG16, InceptionV3, Xception, UNet, VNet, and UNet^{++}.

Dataset	Model	Sensitivity	Specificity	Precision	Accuracy	F1-Score
BraTS	AlexNet	81.39	81.43	82.04	81.93	81.73
	SqueezNet	82.19	83.29	82.95	83.07	82.78
	VGG16	82.63	82.14	83.09	82.79	82.54
	InceptionV3	84.95	85.23	85.31	85.49	85.19
	Xception	85.36	85.94	85.97	85.43	85.27
	UNet	86.49	87.09	87.12	86.71	86.34
	VNet	86.91	86.97	87.04	87.12	87.01
	UNet^{++}	88.13	88.27	87.38	87.51	87.21
	FusionNet	**90.19**	**89.77**	**89.84**	**90.07**	**89.92**
	FusionNet-Seq-ResNet	**93.06**	**93.34**	**93.41**	**93.79**	**93.81**
	FusionNet-Seq-ResNet+Radiomics	**95.19**	**95.37**	**95.46**	**95.83**	**95.79**
TCIA	UNet	82.34	82.78	83.02	82.69	82.88
	VNet	83.12	83.45	83.21	83.34	83.17
	UNet^{++}	81.87	81.98	82.45	82.13	82.28
	FusionNet	**91.56**	**91.23**	**91.92**	**91.64**	**91.74**
	FusionNet-Seq-ResNet	**93.02**	**93.71**	**93.19**	**93.93**	**94.09**
	FusionNet-Seq-ResNet + Radiomics	**95.07**	**95.84**	**94.98**	**94.93**	**95.16**
FIGSHARE	UNet	86.21	86.45	86.32	86.39	86.28
	VNet	87.03	86.89	87.12	87.08	87.17
	UNet^{++}	87.45	87.51	87.38	87.62	87.49
	FusionNet	**91.67**	**92.78**	**92.45**	**91.72**	**92.59**
	FusionNet-Seq-ResNet	**94.21**	**94.34**	**94.47**	**93.15**	**93.28**
	FusionNet-Seq-ResNet+Radiomics	**95.17**	**95.81**	**95.67**	**95.72**	**95.86**

attributed to the superior classification results of UNet utilizing solely MRI slices compared with VNet. The empirical analyses (precision, specificity, recall, sensitivity, F1, and accuracy) of experimental pre-trained CNN models with radiomics features are shown in Table 2. The radiomics feature fusion with deep learning model produced substantial gains in all parameters with $p \leq 0.05$ for the pre-trained CNN architecture and proposed FusionNet model. Furthermore, the data with lower standard deviations demonstrated the radiomics feature fusion with deep learning architecture's improved resilience. These quantifiable findings demonstrate the efficiency of the proposed feature fusion with the deep learning model.

A significant degree of resilience is shown by the variance, which is <3% after 50 epochs. When comparing the pre-trained CNN models with the pilot model, the average values of all parameters for the suggested FusionNet architecture are superior. Only a few statistically significant improvements are observed in the MRI tumor and healthy class data, and the reported mathematical benefits are minor. Further evidence of enhanced resilience is provided by the FusionNet model's decreased variances.

On the BraTS dataset, the performance and efficacy of the novel suggested framework have been tested and verified. Table 2 displays the Average Precision of the proposed Fusion CNN model framework together with other cutting-edge and pre-trained CNN models. The novel and proposed framework, as shown in Figure 2, obtains an AP of 97.53%, which is ∼3.7% higher than the APs attained by other pretraind CNN models such as AlexNet, SqueezeNet, VGG16, InceptionV3, Xception, UNet, VNet, and UNet^{++}. The proposed FusionNet classifier's Recall and Precision curve and specific performance are shown in Table 2 and Figure 5.

To exhibit the performance and efficacy of the novel proposed Seq-ResNet with FusionNet, upsampling and downsampling have been evaluated and the results are shown in Figure 5 and Tables 2, 3.

FusionNet with radiomics features uses Sequential Residual Network layers to create its backbone in contrast to the planned FusionNet (which uses conventional ResNet layers), although other components are the same. In other respects, the Seq-ResNet employed in their backbones is the only distinction between simple FusionNet and FusionNet with Seq-ResNet. By contrasting the assessment outcomes of FusionNet and FusionNet with Seq-ResNet, it is possible to show the usefulness of the suggested novel Seq-ResNet layers. FusionNet without Seq-ResNet obtains an AP of 94.73%, which is 3.31% lower than the AP achieved by FusionNet with Seq-ResNet, according to the assessment findings shown in Tables 2, 3. To put it another way, Sequential Residual Network (Seq-ResNet) layers outperform simple ResNet in the evaluation by 3.31% AP.

Another major difference between simple FusionNet and FusionNet with Seq-ResNet is using upsampling and downsampling modules. By contrasting the assessment outcomes of the proposed simple FusionNet and FusionNet with Seq-ResNet, it is possible to show the usefulness of the proposed FusionNet with Seq-ResNet. Table 2 shows that the feature fusion with upsampling and downsampling modules delivers a strong performance advantage of 3.18% AP when compared with normal simple FusionNet. To have a thorough grasp of the proposed novel FusionNet with Seq-ResNet performance, the classification results are carefully examined. The majority of tumors may be appropriately located in the classification outcomes irrespective of their orientations, colors, and scales, proving the usefulness of the suggested framework. However, the suggested technique occasionally fails when portions of the tumors make up more than half of the original kernel values.

The BraTS dataset has also assessed the proposed framework. Table 2 displays the assessment outcomes of the proposed FusionNet with Seq-RestNet and other pre-trained CNN models such as AlexNet, SqueezeNet, VGG16, InceptionV3, Xception, UNet, VNet, and UNet^{++}. Generally speaking, the suggested novel framework outperforms existing state-of-the-art algorithms, achieving an F1 Score of 96.72%.

Figure 5 shows the Precision and Recall curves for the proposed FusionNet with Seq-ResNet framework based on the BraTS dataset. The comprehensive assessment findings for the suggested novel framework are shown in Table 2. Table 4 shows that the proposed FusionNet with Seq-ResNet performs well for identifying tumors in MRI and achieves plausible results when detecting tumors in MRI images with F1 Score of 95.88 and 96.72%, respectively. As compared with other algorithms, the performance of FusionNet with Seq-ResNet can be steady even as the tumor's scales change quickly.

The accuracy, precision-recall, and F1-measure calculations are used to examine the quantifiable efficiency. The performance and efficiency measures for tumor detection (Accuracy, Sensitivity, Specificity, Precision, and F1 measure) for the dataset BraTS are shown in Table 2. The table shows that the suggested technique performs better than previous methods in terms of Precision-Recall and F1 measure. The suggested technique combines many aspects that are better able to depict image changes and hence execute both statistically and aesthetically better. It is frequently used as a plot indicative of the effectiveness of the classifier. A ROC curve is a useful tool for visualizing, organizing, and choosing learners based on their efficiency. A classifier's ability to determine outcomes is measured by the Receiver Operating Characteristic (ROC), which contrasts and illustrates the trade-off between the model's specificity and sensitivity. The result for the ROC plot is the region beneath the ROC curve, and a large value denotes a successful algorithm. An excellent predictor produces a value in the top left corner of the ROC space, or coordinate (0,1), signifying 100% specificity (zero false positives) and 100% sensitivity (zero false negatives). The ROC plot for the BraTS dataset for the various change detection methods is shown in Figure 7. It is evident from the chart that for the dataset, the suggested strategy outperforms conventional approaches. The area under the ROC curve or AUC is used to calculate the ROC plot's mathematical value. A high AUC value means the applied tumor detection framework is good and can successfully differentiate between benign and malignant patches. By examining the proximity of the curve to the top left corner of the image in Figure 7, it is evident that the suggested strategy works better than the established tumor detection methods for the BraTS dataset.

In Figure 5, the novel proposed method's precision-recall outcomes are shown. For MRI images, the first fusion experiment

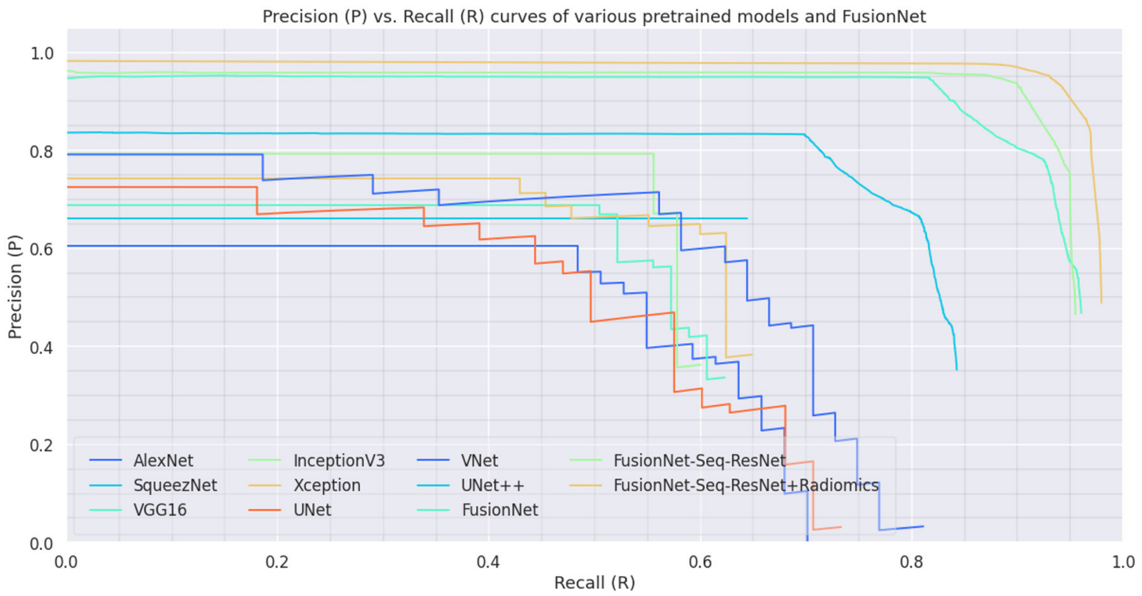


FIGURE 5
Precision-Recall Curve on BraTS dataset: Precision-Recall curve of the proposed model (FusionNet with ResNet, FusionNet with Seq-ResNet, and FusionNet with Seq-ResNet and Radiomics features) and other pre-trained CNN models such as AlexNet, SqueezeNet, VGG16, InceptionV3, Xception, UNet, VNet, and UNet^{++}.

TABLE 3 Training accuracy, F1-Score, specificity, precision, and sensitivity on the BraTS dataset of the proposed model (FusionNet with ResNet, FusionNet with Seq-ResNet, and FusionNet with Seq-ResNet and radiomics features) and other state-of-the-arts methodologies.

Model	Sensitivity	Specificity	Precision	Accuracy	F1-Score
Zare et al. (2018)	81.39 ±	–	–	82.11 ±	–
Zhu et al. (2023)	92.00 ±	93.11 ±	92.01 ±	92.03 ±	–
Jie et al. (2023)	91.17 ±	–	–	92.04 ±	–
Wen et al. (2023)	86.15 ±	–	–	87.12 ±	86.77 ±
Singh and Anand (2019)	–	87.15	–	87.11	88.15
Qin et al. (2018)	–	–	–	77.80	–
Dogra and Kumar (2022)	87.16	–	–	87.08	88.15
Wang et al. (2021)	89.13	88.71	87.19	87.11	88.15
FusionNet	**90.19**	**89.77**	**89.84**	**90.07**	**89.92**
FusionNet-Seq-ResNet	**93.06**	**93.34**	**93.41**	**93.79**	**93.81**
FusionNet-Seq-ResNet+Radiomics	**95.19**	**95.37**	**95.46**	**95.83**	**95.79**

TABLE 4 FusionNet training accuracy, F1-Score, specificity, precision, and sensitivity on the BraTS dataset.

Fold	Sensitivity	Specificity	Precision	Accuracy	F1-Score
Fold-1	93.29	94.71	95.81	95.57	94.38
Fold-2	94.37	95.13	95.77	96.51	95.04
Fold-3	95.91	95.31	96.85	96.17	95.64
Fold-4	94.88	95.57	96.15	97.13	95.89
Fold-5	95.89	95.21	97.77	97.53	96.12
Average	94.88	95.18	96.47	96.58	95.41

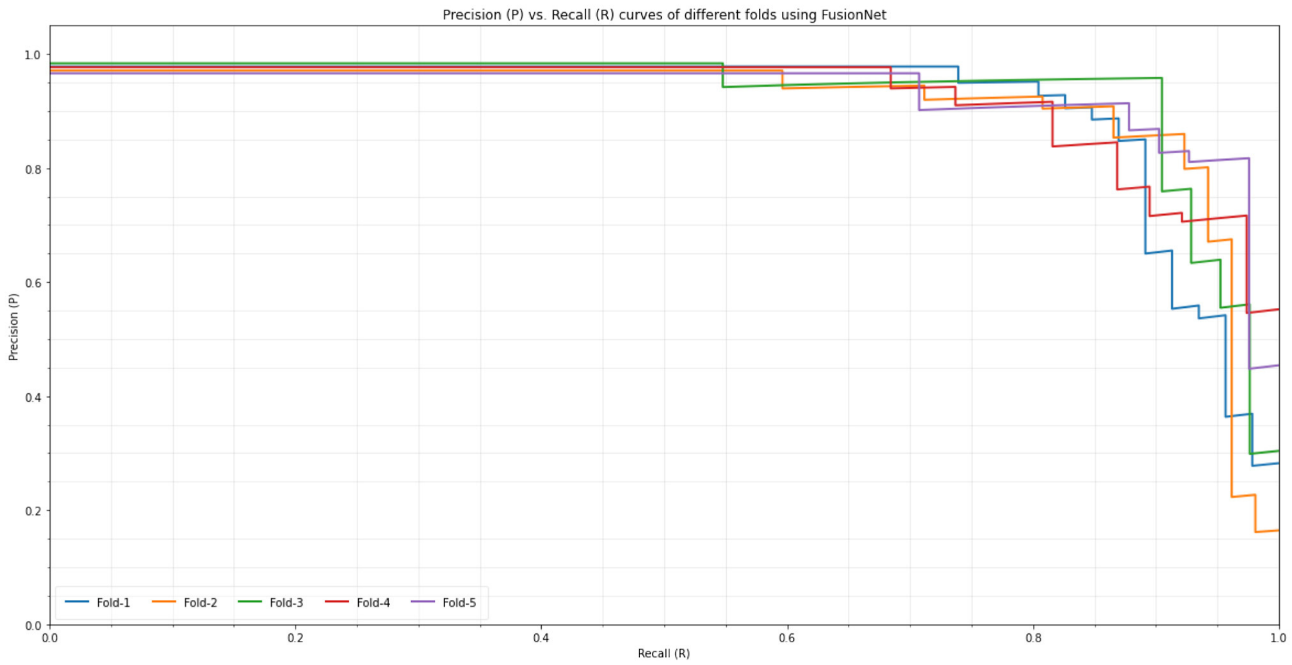


FIGURE 6
Precision (P) vs. Recall (R) curves of different folds using proposed model (FusionNet) on the BraTS dataset.

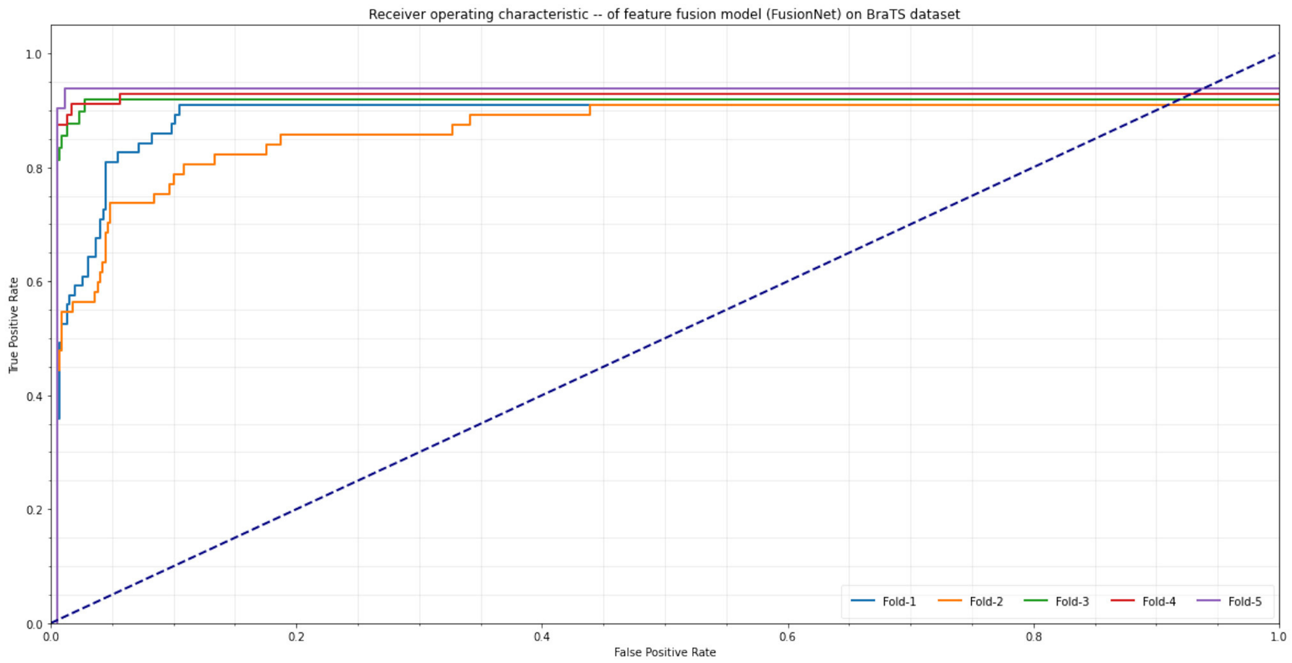


FIGURE 7
Receiver Operating Characteristic (ROC) curves of different folds using proposed model (FusionNet) on the BraTS dataset.

was conducted. The radiomics feature fusion with deep learning features approach offers more accurate anatomical details in MRI images separately. Figure 5 shows how the suggested strategy maintains the measurement elements of MRI images in the fused images. The study that performed on the proposed FusionNet model in consideration of various metrics under various numbers of folds is shown in Table 4 and Figure 6. The proposed FusionNet model's specificity and sensitivity assessments for various fold numbers are shown in Figure 6. The greatest specificity and sensitivity outcomes for the novel proposed FusionNet model under fold-1 is 94.71 and 93.29%, respectively. The presented feature fusion model also produced higher specificity and sensitivity results under fold-2, with corresponding outcomes of 95.13 and 94.37%, respectively. The presented feature fusion model approach also produced the highest specificity and sensitivity outcomes under fold-3, which were 95.31 and 95.91%, respectively. Additionally, the reported FusionNet model strategy had higher specificity and sensitivity under fold-4 95.57 and 94.88%, respectively. The suggested feature fusion approach also showed ideal specificity and sensitivity results of 95.89 and 95.21%, respectively, under fold-5.

The ROC curves of the proposed FusionNet model are shown in Figure 7, demonstrating its accuracy and precision at different fold counts during cross-validation. The model performs better than expected, with fold-1 obtaining 95.57% accuracy and 95.81% precision by combining feature fusion with up and downsampling methods. The feature fusion method yields the best accuracy and precision in fold-2, with 96.51 and 95.77%, respectively. Transpose Convolution, when used for up and downsampling in fold-3, produces even better results, with accuracy and precision reaching 96.17 and 96.85%, respectively. The model achieves significantly greater accuracy and precision under fold-4, 97.13 and 96.15%, respectively. The feature fusion model under fold-5 performs best overall, with a precision of 97.77% and an accuracy of 97.53%. This shows how well the suggested FusionNet integrates deep learning methods for improved medical image interpretation.

A novel framework for the CAD solution may be established by the proposed architecture FusionNet comprises Seq-ResNet and up-down sampling with Transpose Convolution, which combines radiomics analysis with Convolutional Neural Network implementation. The established Seq-ResNet calculation method may potentially improve the efficiency of neural networks in other applications, especially those requiring the intake of multi-channel imaging images. The suggested technique also offers a radiomics viewpoint on the interpretability of deep learning. Since the neural network's hyperparameters were developed without explicit personal expertise interaction, actual interpretation of them is challenging. The black box world of deep learning-based CAD systems hindered their medical and clinical implementations without the support of medical practitioners and radiologists. We explored neural network information using a radiomics-based approach as a first milestone toward deep learning interpretability. It has been shown that radiomic feature regions may be computationally fragmented to produce interpretation. Radiomics have been extensively researched as computational imaging biomarkers for illness identification and performance assessment. Complicating factors, such as anatomical outlines from radiation therapy and histopathology samples from biopsy, can be employed to improve deep learning interpretability after the saliency analytical method in this study. Future research will examine these topics once suitable datasets are obtained.

7 Ablation studies and future works

We performed several sets of ablation experiments to demonstrate the effectiveness of the suggested vascular segmentation method. The purpose of these tests was to investigate the effects of different loss parameter settings and the efficacy of different techniques. We create a baseline using an X-shaped network that contains an encoder with modified sequential residual blocks and a feature decoder (influenced by Szegedy et al., 2017), to validate the efficiency of the proposed methodologies. We develop our network based on this foundation. Tables 2, 3 and Figure 5 show the results, from which we derive several inferences.

Table 2 shows the quantitative experiment data for the baseline and the three proposed models. When compared with the baseline, Model 1 (FusionNet) improves Accuracy, F1-Score, Sensitivity, Specificity, and Precision by 2.15%, 1.89%, 1.96%, 1.73%, and 2.11% percent, demonstrating the benefits of using both up and downsampling strategies. Furthermore, when compared with FusionNet, FusionNet-Seq-ResNet achieves a 2.94% improvement in Accuracy, while FusionNet-Seq-ResNet + Radiomics achieves a 4.81% improvement, demonstrating the efficacy of the Sequential Residual Network and radiomics features. FusionNet exceeds UNet by 2.79% in F1-Score, demonstrating that combining up and downsampling approaches improves the model's performance even further. Similarly, when compared with UNet++, FusionNet-Seq-ResNet shows a 2.98% gain in F1-Score, highlighting the importance of Transpose Convolution and deconvolution in improving brain MRI classification. Finally, when compared with VNet, FusionNet-Seq-ResNet + Radiomics improves accuracy by 3.97%, highlighting the effectiveness of radiomics features and the impact of Transpose Convolution and deConvolution in up and downsampling. The proposed approaches exhibit significant improvements over the baseline, with increments of 3.98% in accuracy, 4.15% in F1-score, 3.89% in Sensitivity, 3.91% in Specificity, and 3.79% in Precision, demonstrating the proposed network's superior segmentation performance.

Our new local convolution-based network, FusionNet, performs well on publicly accessible BraTS datasets. We do, however, find a constraint on its ability to generalize to other data distributions. This restriction results from the fundamental characteristics of local convolution, which prioritize local data above important global data. Consequently, long-term dependencies are difficult for CNNs to capture and are crucial for improving the model's summarization ability. We investigate the possibility of using the Transformer design, which has demonstrated efficacy in creating global dependencies via a self-attention mechanism, to tackle this problem. Transformers have several drawbacks, including a high parameter count and a heavy dependency on large amounts of training data, even if they provide dynamic attention and global context fusion.

We track the variations in accuracy as the epochs proceed during the training process of various networks. The results are shown in Figures 6, 7. When comparing FusionNet with standard pre-trained CNN networks and a network created specifically for BraTS classification, it is clear that FusionNet has easier convergence and training. This gain can be ascribed to FusionNet's Seq-ResNet module, which improves learning ability while utilizing much fewer parameters than typical convolutional layers. When trained on 1,315 samples, the ROC of FusionNet reaches 0.79 at the 13th epoch, 0.86 at the 27th epoch, and 0.885 at the 35th epoch, as shown in Figure 7. On an NVIDIA 1050TI, training for one epoch takes 297 s. This suggests that by applying FusionNet, a high-performing BraTS MRI classification model may be produced in 137 min (80 epochs). These findings confirm that the proposed FusionNet is simple to train and capable of giving satisfactory outcomes.

Table 2 shows how incorporating the up and downsampling module into FusionNet improves performance. FusionNet beats pre-trained CNN models, improving accuracy from 87.11 to 90.07%, precision from 88.15 to 89.92%, and F1-Score from 87.38 to 89.84%. These improvements are ascribed to the proposed up and downsampling module's dynamic merging of multi-scale context information. Two experiments were carried out to further validate this. First, we introduced a Seq-ResNet module, FusionNet-Seq-ResNet, into the baseline with no point-wise convolution rate. It improved accuracy by 4.21% over the baseline but decreased by 2.18% when compared with FusionNet-Seq-ResNet + Radiomics. This result emphasizes the need to obtain multi-scale contextual information. Second, we incorporated radiomics features with various Transposed convolution and deconvolution settings into FusionNet's parallel branches. While these two models outperformed the baseline and FusionNet-Seq-ResNet, they fell just short of FusionNet. This finding demonstrates that the FusionNet-Seq-ResNet + Radiomics transposed convolution and deConvolution configuration is best, and the addition of multi-scale context information is especially useful for up and downsampling. As shown in Figure 6, FusionNet-Seq-ResNet +radiomics outperforms FusionNet in the analysis of under-segmented patches, particularly those with relatively small scales. This finding adds to the evidence that dynamic selection of multi-scale contextual information promotes more successful MRI analysis.

In the future, we suggest a hybrid network design that combines the best features of Transformers and CNNs. This method entails sandwiching a thin transformer module between the encoder and decoder of the CNN. By combining multi-scale information, the lightweight transformer will facilitate the effective integration of multi-scale information and global channel attention, spatial attention, and scale attention. The objective is to create a network that combines the benefits of Transformers (dynamic attention and improved generalization) with CNNs (local receptive fields, shared weights, shift, and scale invariance). By combining these elements, we hope to create a segmentation technique that preserves the advantages of both designs, enhancing FusionNet's overall robustness and speed.

8 Conclusion

According to the results of this comprehensive research, radiomics feature fusion with deep learning features in medical image analysis is a nascent but promising topic that supports medical practice in medical imaging interpretation across all disciplines. We have honed in on important insights, described unanswered problems and summarized essential terminology, approaches, and appraised the state of the art for radiomics feature fusion with deep learning features in medical imaging. Different preprocessing methodologies are carried out initially to continue improving the accuracy of the diseased patch in the proposed fusion framework (FusionNet), and the training dataset is subsequently employed to expand the training dataset. Various pre-trained learning models are used to design and train on the BraTS dataset. Additionally, a FusioneNet deep model is improved with radiomic features, and this manual feature is carried out with other models. Subsequently, the proposed fusion technique (FusionNet) is employed to better integrate the information rather than the initial serial-based methodology. A novel feature simplification approach is presented as a result of the examination of the fused feature space, which shows several duplicate characteristics. The proposed FusionNet model captures the structural, textural, and statistical aspects of brain tumors with an F1 score of 96.72, sensitivity and specificity of 96.31, and AUC of 96.93.

The topic of feature fusion for deep learning in medical imaging is growing, and it is anticipated that new fusion techniques will be created. The upcoming study should concentrate on common nomenclature and measurements for particular and appropriate direct evaluation of various radiomics fusion models. We discovered that radiomics feature fusion with deep learning features for automated medical imaging tasks significantly outperforms in single modality models, and further research may provide insights to guide the most effective methods. We will monitor the effectiveness of the presented methodology based on distinct handcrafted feature fusion by including spatial information and other medical datasets using deep learning and Seq-ResNet in the future.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://www.med.upenn.edu/cbica/brats2020/data.html.

Author contributions

SI: Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. AQ: Writing – review & editing, Supervision, Software, Investigation, Formal analysis. MAl: Writing – review & editing, Visualization, Resources, Project administration, Funding acquisition. KA: Writing – original draft, Methodology, Formal analysis, Data curation. IC: Writing – original draft, Investigation, Formal analysis, Data curation, Conceptualization. MAn: Writing – original draft, Resources, Methodology, Formal analysis, Data curation.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research is funded by Researchers Supporting Project Number (RSPD2024R553), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aerts, H. J., Velazquez, E. R., Leijenaar, R. T., Parmar, C., Grossmann, P., Carvalho, S., et al. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* 5, 1–9. doi: 10.1038/ncomms5006

AlEisa, H. N., Touiti, W., ALHussan, A. A., Aoun, N. B., Ejbali, R., Zaied, M., et al. (2022). Breast cancer classification using fcn and beta wavelet autoencoder. *Comput. Intell. Neurosci.* 2022:8044887. doi: 10.1155/2022/8044887

Alruwaili, M., Alsayat, A., Idris, M., Alanazi, S., and Aurangzeb, K. (2024). Integration and analysis of diverse healthcare data sources: a novel solution. *Comput. Human Behav.* 157:108221. doi: 10.1016/j.chb.2024.108221

Azam, M. A., Khan, K. B., Salahuddin, S., Rehman, E., Khan, S. A., Khan, M. A., et al. (2022). A review on multimodal medical image fusion: compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics. *Comput. Biol. Med.* 144:105253. doi: 10.1016/j.compbiomed.2022.105253

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. doi: 10.1109/TPAMI.2016.2644615

Bagheri, A., Groenhof, T. K. J., Asselbergs, F. W., Haitjema, S., Bots, M. L., Veldhuis, W. B., et al. (2021). Automatic prediction of recurrence of major cardiovascular events: a text mining study using chest X-ray reports. *J. Healthc. Eng.* 2021:6663884. doi: 10.1155/2021/6663884

Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., et al. (2021). The RSNA-ASNR-MICCAI brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv* [preprint]. arXiv:2107.02314. doi: 10.048550/arXiv.2107.02314

Bell, S., Zitnick, C. L., Bala, K., and Girshick, R. (2016). "Inside-outside net: detecting objects in context with skip pooling and recurrent neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Las Vegas, NV: IEEE), 2874–2883. doi: 10.1109/CVPR.2016.314

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 1798–1828. doi: 10.1109/TPAMI.2013.50

Benoit-Cattin, H. (2006). *Texture Analysis for Magnetic Resonance Imaging.* Croatia: IntechOpen.

Bleeker, F. E., Molenaar, R. J., and Leenstra, S. (2012). Recent advances in the molecular understanding of glioblastoma. *J. Neurooncol.* 108, 11–27. doi: 10.1007/s11060-011-0793-0

Brahimi, S., Ben Aoun, N., Ben Amar, C., Benoit, A., and Lambert, P. (2018). Multiscale fully convolutional densenet for semantic segmentation. *J. WSCG* 26:104. doi: 10.24132/JWSCG.2018.26.2.5

Brahimi, S., Aoun, N. B., and Amar, C. B. (2017). "Very deep recurrent convolutional neural network for object recognition," in *Ninth International Conference on Machine Vision (Icmv 2016), Vol. 10341* (Plzen: SPIE), 28–32. doi: 10.1117/12.2268672

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., Jemal, A., et al. (2018). Global cancer statistics 2018: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 68, 394–424. doi: 10.3322/caac.21492

Bruse, J. L., Zuluaga, M. A., Khushnood, A., McLeod, K., Ntsinjana, H. N., Hsia, T.-Y., et al. (2017). Detecting clinically meaningful shape clusters in medical image data: metrics analysis for hierarchical clustering applied to healthy and pathological aortic arches. *IEEE Trans. Biomed. Eng.* 64, 2373–2383. doi: 10.1109/TBME.2017.2655364

Chen, C. Z., Wu, Q., Li, Z. Y., Xiao, L., and Hu, Z. Y. (2021). Diagnosis of Alzheimer's disease based on deeply-fused nets. *Comb. Chem. High Throughput Screen.* 24, 781–789. doi: 10.2174/1386207323666200825092649

Chen, J., Cai, Z., Heidari, A. A., Liu, L., Chen, H., Pan, J., et al. (2023). Dynamic mechanism-assisted artificial bee colony optimization for image segmentation of COVID-19 chest X-ray. *Displays* 118:102485. doi: 10.1016/j.displa.2023.102485

Cherezov, D., Goldgof, D., Hall, L., Gillies, R., Schabath, M., Müller, H., et al. (2019). Revealing tumor habitats from texture heterogeneity analysis for classification of lung cancer malignancy and aggressiveness. *Sci. Rep.* 9, 1–9. doi: 10.1038/s41598-019-38831-0

Constanzo, J., Wei, L., Tseng, H.-H., and El Naqa, I. (2017). Radiomics in precision medicine for lung cancer. *Transl. Lung Cancer Res.* 6:635. doi: 10.21037/tlcr.2017.09.07

Cook, G. J., Yip, C., Siddique, M., Goh, V., Chicklore, S., Roy, A., et al. (2013). Are pretreatment 18f-fdg pet tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? *J. Nucl. Med.* 54, 19–26. doi: 10.2967/jnumed.112.107375

Cook, G. J. O'Brien, M. E., Siddique, M., Chicklore, S., Loi, H. Y., Sharma, B., Punwani, R., Bassett, P., Goh, V., Chua, S. (2015). Non-small cell lung cancer treated with erlotinib: heterogeneity of 18f-fdg uptake at pet-association with treatment response and prognosis. *Radiology* 276, 883–893. doi: 10.1148/radiol.2015141309

Cuocolo, R., Perillo, T., De Rosa, E., Ugga, L., and Petretta, M. (2019). Current applications of big data and machine learning in cardiology. *J. Geriatr. Cardiol.* 16:601.

Cuocolo, R., Caruso, M., Perillo, T., Ugga, L., and Petretta, M. (2020). Machine learning in oncology: a clinical appraisal. *Cancer Lett.* 481, 55–62. doi: 10.1016/j.canlet.2020.03.032

Dogra, A., and Kumar, S. (2022). Multi-modality medical image fusion based on guided filter and image statistics in multidirectional shearlet transform domain. *J. Ambient Intell. Humaniz. Comput.* 14, 1–15. doi: 10.1007/s12652-022-03764-6

El-Gamal, F., Elmogy, M., and Atwan, A. (2016). (2016). Current trends in medical image registration and fusion. *Egypt. Inform. J.* 17, 99–124. doi: 10.1016/j.eij.2015.09.002

Fried, D. V., Tucker, S. L., Zhou, S., Liao, Z., Mawlawi, O., Ibbott, G., et al. (2014). Prognostic value and reproducibility of pretreatment ct texture features in stage III non-small cell lung cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 90, 834–842. doi: 10.1016/j.ijrobp.2014.07.020

Galloway, M. M. (1975). Texture analysis using gray level run lengths. *Comput. Graphics Image Process.* 4, 172–179. doi: 10.1016/S0146-664X(75)80008-6

Ganeshan, B., Goh, V., Mandeville, H. C., Ng, Q. S., Hoskin, P. J., Miles, K. A., et al. (2013). Non-small cell lung cancer: histopathologic correlates for texture parameters at CT. *Radiology* 266, 326–336. doi: 10.1148/radiol.12112428

Ganeshan, B., Panayiotou, E., Burnand, K., Dizdarevic, S., and Miles, K. (2012). Tumour heterogeneity in non-small cell lung carcinoma assessed by ct texture analysis: a potential marker of survival. *Eur. Radiol.* 22, 796–802. doi: 10.1007/s00330-011-2319-8

Ghorbani, A., Ouyang, D., Abid, A., He, B., Chen, J. H., Harrington, R. A., et al. (2020). Deep learning interpretation of echocardiograms. *NPJ Digit. Med.* 3, 1–10. doi: 10.1038/s41746-019-0216-8

Guo, K., Li, X., Zang, H., and Fan, T. (2020). Multi-modal medical image fusion based on fusionnet in yiq color space. *Entropy* 22:1423. doi: 10.3390/e22121423

Häfner, M., Liedlgruber, M., Uhl, A., Vécsei, A., and Wrba, F. (2012). Color treatment in endoscopic image classification using multi-scale local color vector patterns. *Med. Image Anal.* 16, 75–86. doi: 10.1016/j.media.2011.05.006

Hao, S., Huang, C., Heidari, A. A., Xu, Z., Chen, H., Alabdulkreem, E., et al. (2023). Multi-threshold image segmentation using an enhanced fruit fly optimization for COVID-19 X-ray images. *Biomed. Signal Process. Control* 86:105147. doi: 10.1016/j.bspc.2023.105147

Haq, I.-U., Haq, I., and Xu, B. (2021). Artificial intelligence in personalized cardiovascular medicine and cardiovascular imaging. *Cardiovasc. Diagn. Ther.* 11:911. doi: 10.21037/cdt.2020.03.09

Haralick, R. M., Shanmugam, K., and Dinstein, I. H. (1973). Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* SMC-3, 610–621. doi: 10.1109/TSMC.1973.4309314

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., et al. (2022). "Unetr: transformers for 3D medical image segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (Waikoloa, HI: IEEE), 574–584. doi: 10.1109/WACV51458.2022.00181

Hawkins, S., Wang, H., Liu, Y., Garcia, A., Stringfield, O., Krewer, H., et al. (2016). Predicting malignant nodules from screening ct scans. *J. Thorac. Oncol.* 11, 2120–2128. doi: 10.1016/j.jtho.2016.07.002

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Las Vegas, NV: IEEE), 770–778. doi: 10.1109/CVPR.2016.90

Horeweg, N., van Rosmalen, J., Heuvelmans, M. A., van der Aalst, C. M., Vliegenthart, R., Scholten, E. T., et al. (2014). Lung cancer probability in patients with ct-detected pulmonary nodules: a prespecified analysis of data from the nelson trial of low-dose ct screening. *Lancet Oncol.* 15, 1332–1341. doi: 10.1016/S1470-2045(14)70389-4

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. (2016). Densely connected convolutional networks. *arXiv* [Preprint]. doi: 10.48550/arXiv.1608.06993

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Honolulu, HI: IEEE), 4700–4708. doi: 10.1109/CVPR.2017.243

Iqbal, S., Qureshi, A. N., Aurangzeb, K., Alhussein, M., Anwar, M. S., Zhang, Y., et al. (2024). Adaptive magnification network for precise tumor analysis in histopathological images. *Comput. Human Behav.* 156:108222. doi: 10.1016/j.chb.2024.108222

Isensee, F., Jäger, P. F., Full, P. M., Vollmuth, P., and Maier-Hein, K. H. (2020). "NNU-NET for brain tumor segmentation," in *International MICCAI Brainlesion Workshop* (Cham: Springer), 118–132. doi: 10.1007/978-3-030-72087-2_11

Isensee, F., Jäger, P. F., Full, P. M., Vollmuth, P., and Maier-Hein, K. H. (2021). "NNU-NET for brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II 6* (Cham: Springer), 118–132.

James, A. P., and Dasarathy, B. V. (2014). Medical image fusion: a survey of the state of the art. *Inf. Fusion* 19, 4–19. doi: 10.1016/j.inffus.2013.12.002

Jiang, Z., Ding, C., Liu, M., and Tao, D. (2019). "Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task," in *International MICCAI brainlesion workshop* (Cham: Springer), 231–241. doi: 10.1007/978-3-030-46640-4_22

Jie, Y., Li, X., Zhou, F., and Tan, H. (2023). Medical image fusion based on extended difference-of-gaussians and edge-preserving. *Expert Syst. Appl.* 2:, 120301. doi: 10.1016/j.eswa.2023.120301

Kendall, A., Badrinarayanan, V., and Cipolla, R. (2015). Bayesian segnet: model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. *arXiv* [Preprint]. arXiv:1511.02680. doi: 10.48550/arXiv.1511.02680

Kim, J., Ryu, S.-Y., Lee, S.-H., Lee, H. Y., and Park, H. (2019). Clustering approach to identify intratumour heterogeneity combining FDG pet and diffusion-weighted mri in lung adenocarcinoma. *Eur. Radiol.* 29, 468–475. doi: 10.1007/s00330-018-5590-0

Kong, T., Yao, A., Chen, Y., and Sun, F. (2016). "Hypernet: towards accurate region proposal generation and joint object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Las Vegas, NV: IEEE), 845–853. doi: 10.1109/CVPR.2016.98

Lachinov, D., Vasiliev, E., and Turlapov, V. (2018). "Glioma segmentation with cascaded UNET," in *International MICCAI Brainlesion Workshop* (Cham: Springer), 189–198. doi: 10.1007/978-3-030-11726-9_17

Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., Van Stiphout, R. G., Granton, P., et al. (2012). Radiomics: extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* 48, 441–446. doi: 10.1016/j.ejca.2011.11.036

Li, H., Li, A., and Wang, M. (2019). A novel end-to-end brain tumor segmentation method using improved fully convolutional networks. *Comput. Biol. Med.* 108, 150–160. doi: 10.1016/j.compbiomed.2019.03.014

Li, X., Chen, H., Qi, X., Dou, Q., Fu, C.-W., Heng, P.-A., et al. (2018). H-denseunet: hybrid densely connected unet for liver and tumor segmentation from ct volumes. *IEEE Trans. Med. Imaging* 37, 2663–2674. doi: 10.1109/TMI.2018.2845918

Liao, S., Law, M. W., and Chung, A. C. (2009). Dominant local binary patterns for texture classification. *IEEE Trans. Image Process.* 18, 1107–1118. doi: 10.1109/TIP.2009.2015682

Lin, E., Lin, C.-H., and Lane, H.-Y. (2020). Precision psychiatry applications with pharmacogenomics: artificial intelligence and machine learning approaches. *Int. J. Mol. Sci.* 21:969. doi: 10.3390/ijms21030969

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., et al. (2017). A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88. doi: 10.1016/j.media.2017.07.005

Liu, Y., Liu, S., and Wang, Z. (2015). A general framework for image fusion based on multi-scale transform and sparse representation. *Inf. Fusion*, 24, 147–164. doi: 10.1016/j.inffus.2014.09.004

Lowe, D. G. (1999). "Object recognition from local scale-invariant features," in *ICCV, Vol.* 99 (Kerkyra), 1150–1157. doi: 10.1109/ICCV.1999.790410

Lubner, M. G., Smith, A. D., Sandrasegaran, K., Sahani, D. V., and Pickhardt, P. J. (2017). CT texture analysis: definitions, applications, biologic correlates, and challenges. *Radiographics* 37, 1483–1503. doi: 10.1148/rg.2017170056

Ma, J., Wang, Q., Ren, Y., Hu, H., and Zhao, J. (2016). "Automatic lung nodule classification with radiomics approach," in *Medical Imaging 2016: PACS and Imaging Informatics: Next Generation and Innovations, volume 9789* (San Diego, CA: SPIE), 26–31. doi: 10.1117/12.2220768

Manthe, M., Duffner, S., and Lartizien, C. (2024). "Whole brain radiomics for clustered federated personalization in brain tumor segmentation," in *Medical Imaging with Deep Learning* (Nashville, TN: PMLR), 957–977.

Mavakala, A. W., Adoni, W. Y. H., Ben Aoun, N., Nahhal, T., Krichen, M., Alzahrani, M. Y., et al. (2023). COV19-Dijkstra: a COVID-19 propagation model based on Dijkstra's algorithm. *J. Comput. Sci.* 19, 75–86. doi: 10.3844/jcssp.2023.75.86

McGranahan, N., and Swanton, C. (2015). Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* 27, 15–26. doi: 10.1016/j.ccell.2014.12.001

Mohan, G., and Subashini, M. M. (2018). MRI based medical image analysis: survey on brain tumor grade classification. *Biomed. Signal Process. Control* 39, 139–161. doi: 10.1016/j.bspc.2017.07.007

Myronenko, A. (2018). "3D MRI brain tumor segmentation using autoencoder regularization," in *International MICCAI Brainlesion Workshop* (Cham: Springer), 311–320. doi: 10.1007/978-3-030-11726-9_28

Nazir, S., Patel, S., and Patel, D. (2022). "Model optimisation techniques for convolutional neural networks," in *Handbook of Research on New Investigations in Artificial Life, AI, and Machine Learning*, ed. M. K. Habib (Hershey, PA: IGI Global), 269–298. doi: 10.4018/978-1-7998-8686-0.ch011

Newell, A., Yang, K., and Deng, J. (2016). "Stacked hourglass networks for human pose estimation," in *European conference on computer vision* (Cham: Springer), 483–499. doi: 10.1007/978-3-319-46484-8_29

Nosaka, R., Ohkawa, Y., and Fukui, K. (2011). "Feature extraction based on co-occurrence of adjacent local binary patterns," *Pacific-RIM symposium on image and video technology* (Berlin: Springer), 82–91. doi: 10.1007/978-3-642-25346-1_8

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., et al. (2018). Attention U-NET: learning where to look for the pancreas. *arXiv* [Preprint]. arXiv:1804.03999. doi: 10.48550/arXiv.1804.03999

Parekh, V. S., and Jacobs, M. A. (2017). Integrated radiomic framework for breast cancer and tumor biology using advanced machine learning and multiparametric MRI. *NPJ Breast Cancer* 3, 1–9. doi: 10.1038/s41523-017-0045-3

Parmar, C., Barry, J. D., Hosny, A., Quackenbush, J., and Aerts, H. J. (2018). Data analysis strategies in medical imagingdata science designs in medical imaging. *Clin Cancer Res.* 24, 3492–3499. doi: 10.1158/1078-0432.CCR-18-0385

Peng, C., Zhang, X., Yu, G., Luo, G., and Sun, J. (2017). "Large kernel matters-improve semantic segmentation by global convolutional network," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Honolulu, HI: IEEE), 4353–4361. doi: 10.1109/CVPR.2017.189

Pérez-Morales, J., Tunali, I., Stringfield, O., Eschrich, S. A., Balagurunathan, Y., Gillies, R. J., et al. (2020). Peritumoral and intratumoral radiomic features predict survival outcomes among patients diagnosed in lung cancer screening. *Sci. Rep.* 10, 1–15. doi: 10.1038/s41598-020-67378-8

Qin, F., Gao, N., Peng, Y., Wu, Z., Shen, S., Grudtsin, A., et al. (2018). Fine-grained leukocyte classification with deep residual learning for microscopic images. *Comput. Methods Programs Biomed.* 162, 243–252. doi: 10.1016/j.cmpb.2018.05.024

Quan, T. M., Hildebrand, D. G. C., and Jeong, W.-K. (2021). Fusionnet: a deep fully residual convolutional neural network for image segmentation in connectomics. *Front. Comput. Sci.* 3:613981. doi: 10.3389/fcomp.2021.613981

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention* (Cham: Springer), 234–241. doi: 10.1007/978-3-319-24574-4_28

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* 115, 211–252. doi: 10.1007/s11263-015-0816-y

Scalco, E., and Rizzo, G. (2017). Texture analysis of medical images for radiotherapy applications. *Br. J. Radiol.* 90:20160642. doi: 10.1259/bjr.20160642

Sermesant, M., Delingette, H., Cochet, H., Jaïs, P., and Ayache, N. (2021). Applications of artificial intelligence in cardiovascular imaging. *Nat. Rev. Cardiol.* 18, 600–609. doi: 10.1038/s41569-021-00527-2

Singh, S., and Anand, R. S. (2019). Multimodal medical image fusion using hybrid layer decomposition with CNN-based feature mapping and structural clustering. *IEEE Trans. Instrum. Meas.* 69, 3855–3865. doi: 10.1109/TIM.2019.2933341

Spadarella, G., Perillo, T., Ugga, L., and Cuocolo, R. (2022). Radiomics in cardiovascular disease imaging: from pixels to the heart of the problem. *Curr. Cardiovasc. Imaging Rep.* 15, 1–11. doi: 10.1007/s12410-022-09563-z

Sun, K., Xiao, B., Liu, D., and Wang, J. (2019). "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Long Beach, CA: IEE), 5693–5703. doi: 10.1109/CVPR.2019.00584

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. doi: 10.3322/caac.21660

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence* (San Francisco, CA). doi: 10.1609/aaai.v31i1.11231

Taal, W., Bromberg, J. E., and van den Bent, M. J. (2015). Chemotherapy in glioma. *CNS Oncol.* 4, 179–192. doi: 10.2217/cns.15.2

Thibault, G., Angulo, J., and Meyer, F. (2013). Advanced statistical matrices for texture characterization: application to cell classification. *IEEE Trans. Biomed. Engi.* 61, 630–637. doi: 10.1109/TBME.2013.2284600

Tsuneta, S., Oyama-Manabe, N., Hirata, K., Harada, T., Aikawa, T., Manabe, O., et al. (2021). Texture analysis of delayed contrast-enhanced computed tomography to diagnose cardiac sarcoidosis. *Jpn. J. Radiol.* 39, 442–450. doi: 10.1007/s11604-020-01086-1

Ugga, L., Perillo, T., Cuocolo, R., Stanzione, A., Romeo, V., Green, R., et al. (2021). Meningioma mri radiomics and machine learning: systematic review, quality score assessment, and meta-analysis. *Neuroradiology* 63, 1293–1304. doi: 10.1007/s00234-021-02668-0

Wang, J., Cao, M., Zhang, T., Liu, B., and Li, H. (2024). Point-based visual status evaluation of worn pavement markings based on a feature-binary-pointnet network and shape descriptors using lidar point clouds: a case study of an expressway. *Transp. Res. Rec.*, 2678, 562–574. doi: 10.1177/03611981231185139

Wang, L., Dou, J., Qin, P., Lin, S., Gao, Y., Wang, R., et al. (2021). Multimodal medical image fusion based on nonsubsampled shearlet transform and convolutional sparse representation. *Multimed. Tools Appl.* 80, 36401–36421. doi: 10.1007/s11042-021-11379-w

Wang, S., Chen, H., and Zhang, Y. (2023). Bionic artificial neural networks in medical image analysis. *Biomimetics* 8:211. doi: 10.3390/biomimetics8020211

Wen, J., Qin, F., Du, J., Fang, M., Wei, X., Chen, C. P., et al. (2023). MsgFusion: medical semantic guided two-branch network for multimodal brain image fusion. *IEEE Trans. Multimedia.* 26, 944–957. doi: 10.1109/TMM.2023.3273924

Whitney, H. M., Taylor, N. S., Drukker, K., Edwards, A. V., Papaioannou, J., Schacht, D., et al. (2019). Additive benefit of radiomics over size alone in the distinction between benign lesions and luminal a cancers on a large clinical breast MRI dataset. *Acad. Radiol.* 26, 202–209. doi: 10.1016/j.acra.2018.04.019

Win, T., Miles, K. A., Janes, S. M., Ganeshan, B., Shastry, M., Endozo, R., et al. (2013). Tumor heterogeneity and permeability as measured on the ct component of pet/ct predict survival in patients with non-small cell lung cancertumor heterogeneity. *Clin. Cancer Res.* 19, 3591–3599. doi: 10.1158/1078-0432.CCR-12-1307

Xu, Q., Ma, Z., Na, H., and Duan, W. (2023). Dcsau-net: a deeper and more compact split-attention u-net for medical image segmentation. *Comput. Biol. Med.* 154:106626. doi: 10.1016/j.compbiomed.2023.106626

Xu, Z., Zhang, X., Zhang, H., Liu, Y., Zhan, Y., Lukasiewicz, T., et al. (2023). EFPN: effective medical image detection using feature pyramid fusion enhancement. *Comput. Biol. Med.* 163:107149. doi: 10.1016/j.compbiomed.2023.107149

Yue, L., Tian, D., Chen, W., Han, X., and Yin, M. (2020). Deep learning for heterogeneous medical data analysis. *World Wide Web* 23, 2715–2737. doi: 10.1007/s11280-019-00764-z

Zare, M. R., Alebiosu, D. O., and Lee, S. L. (2018). "Comparison of handcrafted features and deep learning in classification of medical X-ray images," in *2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)* (Kota Kinabalu: IEEE), 1–5. doi: 10.1109/INFRKM.2018.8464688

Zeineldin, R. A., Karar, M. E., Coburger, J., Wirtz, C. R., and Burgert, O. (2020). DEEPSEG: deep neural network framework for automatic brain tumor segmentation using magnetic resonance flair images. *Int. J. Comput. Assist. Radiol. Surg.* 15, 909–920. doi: 10.1007/s11548-020-02186-z

Zhang, J., Liu, Y., Wu, Q., Wang, Y., Liu, Y., Xu, X., et al. (2022). Swtru: star-shaped window transformer reinforced u-net for medical image segmentation. *Comput. Biol. Med.* 150:105954. doi: 10.1016/j.compbiomed.2022.105954

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. (2018). "Unet++: a nested u-net architecture for medical image segmentation," in *Deep learning in medical image analysis and multimodal learning for clinical decision support* (Chma: Springer), 3–11. doi: 10.1007/978-3-030-00889-5_1

Zhu, Z., He, X., Qi, G., Li, Y., Cong, B., Liu, Y., et al. (2023). Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal MRI. *Inf. Fusion* 91, 376–387. doi: 10.1016/j.inffus.2022.10.022

Zwanenburg, A., Vallières, M., Abdalah, M. A., Aerts, H. J., Andrearczyk, V., Apte, A., et al. (2020). The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 295:328. doi: 10.1148/radiol.2020191145

frontiers | Frontiers in Computational Neuroscience

TYPE Original Research
PUBLISHED 26 June 2024
DOI 10.3389/fncom.2024.1418280

OPEN ACCESS

EDITED BY
Najib Ben Aoun,
Al Baha University, Saudi Arabia

REVIEWED BY
Sadaqat Ur Rehman,
University of Salford, United Kingdom
Zahid Khan,
Prince Sultan University, Saudi Arabia

*CORRESPONDENCE
Asim Zeb
✉ asimzeb1@gmail.com

RECEIVED 16 April 2024
ACCEPTED 27 May 2024
PUBLISHED 26 June 2024

CITATION
Kiran L, Zeb A, Rehman QNU, Rahman T, Shehzad Khan M, Ahmad S, Irfan M, Naeem M, Huda S and Mahmoud H (2024) An enhanced pattern detection and segmentation of brain tumors in MRI images using deep learning technique.
Front. Comput. Neurosci. 18:1418280.
doi: 10.3389/fncom.2024.1418280

An enhanced pattern detection and segmentation of brain tumors in MRI images using deep learning technique

Lubna Kiran[1], Asim Zeb[2]*, Qazi Nida Ur Rehman[3], Taj Rahman[1], Muhammad Shehzad Khan[2], Shafiq Ahmad[4], Muhammad Irfan[5], Muhammad Naeem[2], Shamsul Huda[6] and Haitham Mahmoud[4]

[1]Qurtuba University of Science and Information Technology, Peshawar, Pakistan, [2]Abbottabad University of Science and Technology, Abbottabad, Pakistan, [3]Institute of Management Sciences (IMSciences), Peshawar, Pakistan, [4]Department of Industrial Engineering, College of Engineering, King Saud University, Riyadh, Saudi Arabia, [5]Department of Computer Science, Kohat University of Science and Technology, Kohat, Pakistan, [6]School of Information Technology, Deakin University, Burwood, VIC, Australia

Neuroscience is a swiftly progressing discipline that aims to unravel the intricate workings of the human brain and mind. Brain tumors, ranging from non-cancerous to malignant forms, pose a significant diagnostic challenge due to the presence of more than 100 distinct types. Effective treatment hinges on the precise detection and segmentation of these tumors early. We introduce a cutting-edge deep-learning approach employing a binary convolutional neural network (BCNN) to address this. This method is employed to segment the 10 most prevalent brain tumor types and is a significant improvement over current models restricted to only segmenting four types. Our methodology begins with acquiring MRI images, followed by a detailed preprocessing stage where images undergo binary conversion using an adaptive thresholding method and morphological operations. This prepares the data for the next step, which is segmentation. The segmentation identifies the tumor type and classifies it according to its grade (Grade I to Grade IV) and differentiates it from healthy brain tissue. We also curated a unique dataset comprising 6,600 brain MRI images specifically for this study. The overall performance achieved by our proposed model is 99.36%. The effectiveness of our model is underscored by its remarkable performance metrics, achieving 99.40% accuracy, 99.32% precision, 99.45% recall, and a 99.28% F-Measure in segmentation tasks.

KEYWORDS

brain tumor, deep learning, pattern detection, neuroscience, segmentation technique, convolution neural network, binary convolution neural network, magnetic resonance images

1 Introduction

Neuroscience is a rapidly evolving field dedicated to decoding the complex functions of the human brain and mind (Efford, 2000; Yamashita et al., 2018; Interpolation Methods, 2024).

Brain tumors represent a critical health challenge, potentially fatal at any stage of detection. A radiologist's knowledge and experience are vital when diagnosing brain tumors manually, yet it is not always accessible. Moreover, conventional diagnostic methods are time-intensive and susceptible to errors (Solanki et al., 2023). Brain cancer is a significant health concern, impacting individuals of any gender and at any life stage. A brain tumor consists of an aberrant growth of cells within the brain that proliferates differently and uncontrollably from normal brain tissues. To date, over 100 distinct types of brain tumors have been diagnosed. These tumors are broadly categorized into two main categories—primary and metastatic or secondary.

The tumor that develops inside or around the brain tissues is called a primary brain tumor. It can be benign (non-cancerous) or malignant (cancerous). Tumors that develop in other parts of the body and transfer or reach the brain are called secondary or metastatic brain tumors. Mostly, secondary brain tumors are considered malignant (cancerous). More than 84,000 persons in the United States were diagnosed with a primary brain tumor in 2021, according to the American Brain Tumor Association (ABTA). There are more than 100 subtypes of primary brain and central nervous system (CNS) cancers. Malignant tumors comprise about a third (29.7%) of all CNS malignancies. Currently, there are approximately 28,000 cases of pediatric brain tumors in the United States. Over 18,000 people lost their lives to primary malignant brain tumors in 2021. There is a significant difference in survival time after a diagnosis of a primary brain tumor based on factors such as age, region, race, tumor location, tumor type, and molecular markers (Oztek et al., 2023).

The World Health Organization (WHO) defines tumors in terms of grades—from Grade I to Grade I—based on their size and growth. Grade I and Grade II tumors are considered non-cancerous; they are slow-growing and curable. Grade III and Grade IV tumors are aggressive and grow very quickly. These types of tumors are considered malignant (cancerous) and categorized as metastatic or secondary tumors (Louis et al., 2021).

Grade I: Tumors in Grade I grow very slowly and do not spread aggressively. A patient with such a tumor can survive for a longer period; this tumor can be removed through surgery, and the patient can survive completely.

Grade II: Grade II tumors also grow slowly but can affect their neighboring tissues and progress to higher grades. After surgery, the tumor can develop again and affect the patient.

Grade III: Compared to Grade I and Grade II tumors, the rate of growth is faster in Grade III tumors, and it can affect the neighboring tissues quickly. Simple surgery is not very effective in removing this type of tumor; further post-surgery treatment is needed for survival.

Grade IV: These tumors are highly aggressive and spread to neighboring tissues. The blood vessels are the most important path for their growth. A patient diagnosed with such a tumor cannot survive for long.

Detecting brain tumors using deep learning, artificial intelligence, computer vision, and image processing has gained attention today. Automatic learning systems require different features to detect brain tumors, including shape, size, location, intensity, and growth. Researchers in the computer science field place a lot of importance on building robust and automated detection methods (Amin et al., 2021).

State-of-the-art research work has been done in the domain of brain tumor segmentation. Different methods have been proposed by researchers for the segmentation of brain tumors. One such method uses the convolutional neural network (CNN) to classify five types of tumors into three classes (Irmak, 2021). Another method based on the multiclass support vector machine (M-SVM) used meningioma, glioma, and pituitary brain tumors for segmentation. Yet another method was based on transfer learning for brain tumor segmentation (Maqsood et al., 2022). They used gliomas, meningioma, pituitary, and normal brain MRI images for brain tumor categorization. The model was trained on 75% of the images, and 25% were used for validation (Shoaib et al., 2022). However, gaps (Nida-Ur-Rehman et al., 2017; Irmak, 2021; Maqsood et al., 2022; Shoaib et al., 2022) still need to be filled while addressing brain tumor segmentation. Firstly, there are more than 100 types of brain tumors in the world that need to be segmented. Secondly, brain tumors are divided into four grades (Grade I to Grade IV) based on their size, growth, and aggressiveness. Each brain tumor must be accurately and timely classified into respective grades, which is very important. Thirdly, the image dataset used in brain tumor segmentation needs to be multivariant and multimodal to make the segmentation system more mature and accurate in brain tumor segmentation.

In this study, we propose a novel method that leverages deep learning using a binary convolutional neural network (BCNN) to classify the 10 most common types of brain tumors into their respective grades (Grade I to Grade IV); current models are limited to the detection of four brain tumor types. In our proposed model, image acquisition is followed by a comprehensive preprocessing phase, during which binary conversion using adaptive thresholding and morphological operations are executed. Secondly, segmentation is carried out to accurately classify tumor types into their respective grades. The model also accurately classifies healthy brain MRI images. Another contribution is the development of a dataset of 6,600 brain MRI images created for this research work; the dataset consists of different modalities, angles, and shapes for the entire brain model. The study is organized as follows: Section 2 includes a comprehensive discussion of brain tumor detection mechanisms. Our proposed model is explained in Section 3. Section 4 highlights the validation of the proposed technique using simulation experiments. Section 5 presents the conclusion and potential for future work.

2 Literature review

A detailed review of the latest and most relevant literature is presented in this section. An overview of the missing points in the literature is presented in a summary at the end of this section.

In Soomro et al. (2023), a detailed review of brain tumor segmentation from 1998 to 2020 is presented. It includes a complete overview of machine learning and image segmentation methods for brain tumor segmentation using MRI images. The state-of-the-art machine learning techniques and deep learning are reviewed, and at the end of their review, they establish that deep learning techniques perform better in brain tumor segmentation using MRI images.

Brain tumor segmentation and segmentation using MRI images detect many types of noise, including speckle noise, salt and pepper noise, and Gaussian noise, which may arise during the scanning process (Ramesh et al., 2021). Consequently, there may be lower accuracy rates in categorization. Therefore, the authors propose a novel noise-canceling algorithm—the iterative group median filter with modifications. Moreover, kernel principal component analysis based on maximum likelihood estimation is presented for feature extraction. The VGG16 architecture, which is based on deep learning, was used for the segmentation task. The suggested method has proven to perform better in both qualitative and quantitative experiment evaluations.

In Ahmed et al. (2016a,b), a total of 1,200 brain MRI scans of brain tissue damaged by tumors and 300 scans of normal brain tissue are included. The suggested approach is effective in detecting four different forms of brain tumors: CNS lymphoma, glioblastoma, meningioma, and metastases. This technique separates the tumor regions from healthy tissue using a 2D adaptive filter and Otsu segmentation. A combination of morphological operations and image fusion is used to highlight the tumor region so that it may be studied in detail.

In Nida-Ur-Rehman et al. (2017), brain tumors are broken down into four distinct categories. The dataset utilized in this study consists of two thousand magnetic resonance imaging (MRI) scans with a clinical and expert opinion from the FCPS neurosurgeon. Histogram differencing is utilized to segregate and detect tumor pixels from the rest of the brain tissues. The volume of data utilized in the categorization process might alter the final findings and cause them to differ between datasets.

In Le et al. (2021), the authors offer a deep learning methods-based strategy to detect and segment brain tumors. This investigation has two key phases. In the first phase, the network only pays attention to the area around the tumor to identify brain tumors using a contextual detection network. The #D atrous residual network is used in the second phase to segment tumors.

Comparative approaches of different segmentation techniques are used in Bahadure et al. (2018), and the best one is selected by comparing their segmentation score. Further, to improve the segmentation accuracy, the genetic algorithm is employed to automatically segment the tumor stage. The decision on the segmentation stage is supported by extracting relevant features and calculating the area. The experimental results of the proposed technique are evaluated and validated for performance and quality analysis on magnetic resonance brain images based on segmentation score, accuracy, sensitivity, specificity, and dice similarity index coefficient. The experimental results achieved 92.03% accuracy, 91.42% specificity, 92.36% sensitivity, and an average segmentation score between 0.82 and 0.93, demonstrating the effectiveness of the proposed technique for identifying normal and abnormal tissues from brain MRI images. The experimental results also obtained an average of 93.79% dice similarity index coefficient, which indicates better overlap between the automated extracted tumor regions and manually extracted tumor regions by radiologists.

In another study, researchers utilized a machine learning technique, specifically a Convolutional Neural Network (CNN) (Badža and Barjaktarović, 2020), for the segmentation of brain tumors. CNNs are well known performing high performance in image segmentation tasks. The authors introduced a novel CNN architecture tailored for segmenting three types of brain tumors. This new network is more straightforward compared to existing pre-trained models and tested using T1-weighted contrast-enhanced MRI scans. The network's performance is assessed through four different methods: two variations of 10-fold cross-validation and two distinct databases. Its ability to generalize is evaluated using subject-wise cross-validation, and improvements are measured with an augmented image database. The highest accuracy, 96.56%, is achieved with record-wise cross-validation on the augmented dataset. With its robust generalization and swift execution, this CNN architecture shows promise as a decision-support tool for radiologists in medical diagnostics.

In Garg and Garg (2021), a method is evaluated using a dataset of 2556 images, split 85:15 for training and testing, achieving an accuracy of 97.305%. This method involves brain tumor segmentation using a hybrid ensemble approach that combines K-Nearest Neighbors (KNN), Random Forest (RF), and Decision Tree (DT) based on the Majority Voting method, namely KNNRF-DT. The aim is to calculate the tumor region's area and classify tumors as benign or malignant. Otsu's Threshold method is used for segmentation, while feature extraction is performed using Stationary Wavelet Transform (SWT), Principal Component Analysis (PCA), and Gray Level Co-occurrence Matrix (GLCM), providing thirteen features for segmentation. The hybrid ensemble classifier (KNN-RFDT) based on Majority Voting aims to enhance the performance of traditional classifiers without resorting to deep learning techniques.

In Phan and ThanhHieu (2024), a combination of three different existing algorithms is proposed for segmenting brain tumors. The algorithms used are the PGDBCWMF algorithm for noise removal in the preprocessing, the SIFT (scale-invariant characteristic remodel) approach for feature extraction, and the HV region algorithm for segmenting brain tumors. A brain and pancreatic tumor dataset is used for the segmentation of tumors.

In Akter et al. (2024), a deep convolution neural network is proposed for the classification and a U-NET-based segmentation model for the segmentation of four different categories of MRI images. The four different categories consist of glioma brain tumor, pituitary brain tumor, meningioma brain tumor, and images with no tumor. Six different datasets were used to train the segmentation model and to test the classification model. The overall accuracy achieved by their proposed model is 98.7% based on the merged dataset, 98.8% accuracy in the segmentation section, and 97.7% classification accuracy with individual datasets.

TABLE 1 Dataset of brain MRI images of each tumor type and healthy brain MRI images.

S. No	Tumor types	Number of MRI images
1	CNS Lymphoma	800
2	Glioblastoma	600
3	Meningioma	600
4	Metastases	400
5	Astrocytoma	600
6	Cystic Pituitary Adenoma and Meningioma	500
7	Ependymomas	600
8	CNS Embryonal Tumor NOS	500
9	Oligodendrogliomas	500
10	Hemangioblastomas	500
11	Healthy brain MRI images	1,000
	Total	6,600

The literature review reveals numerous missing aspects and requires attention to create a more robust and accurate segmentation module. Firstly, very few brain tumor types have been considered for segmentation. Secondly, tumor types must still be classified into grades from Grade I to Grade IV. Finally, the dataset of images needs to be multivariate and multimodal and offer diverse features so that the segmentation module classifies every image easily and accurately.

3 Materials and methods

3.1 Dataset of brain MRI images

The dataset of brain MRI images used in this study is collected from Nida-Ur-Rehman et al. (2017) and Radiopaedia's (2023). The dataset contains brain MRI images of 10 tumor types and healthy brain MRI images (Table 1). The collected dataset of images was checked and verified by doctors from the medical field for its authenticity. The dataset included multimodal and multivariant brain MRI images to cover all the angles, shapes, and positions of the brain for the classification of tumors. Table 1 presents the number of MRI images used for each type of tumor and healthy brain MRI images.

CNS Lymphoma: Primary central nervous system Lymphoma is a type of brain tumor that can be primary and secondary; in this type of tumor, cells emerge in the lymphoma and/or the spinal cord region (RMH Neuropathology, 2013).

Glioblastoma: This type of tumor is considered dangerous because it grows fast and spreads quickly inside the brain. Initially, glioblastoma attacks adjacent brain tissues (Gaillard, 2018).

Meningioma: This brain tumor starts inside the brain tissues called meninges that protect the brain and spinal cord. Most meningiomas are not dangerous but can reach up to Grade III tumor levels (Di Muzio, 2023).

Metastases: These types of tumors spread from other parts of the body, such as the lungs, breasts, and kidneys, to the rest of the body. Once they spread to the brain, they can create one or more tumors inside the brain (Brusic, 2021).

Astrocytoma: These tumors can be cancerous or non-cancerous. Some grow very slowly, while others can be aggressive. They appear first in cells called astrocytes (Gaillard, 2021).

Cystic Pituitary Adenoma and Meningioma: They are generally slow-growing types and fall in the benign category of brain tumors, which are mostly considered Grade 1 or Grade II tumors. Most patients with this type of tumor are diagnosed after several years before observing any signs. It develops from pituitary tissues and grows inside the pituitary gland of the brain (Gaillard, 2016).

Ependymomas: This type of brain tumor develops inside the brain or spinal cord area. It can reach Grade 3 from Grade 1. It initially begins in ependymal cells that help to maintain and improve brain streams (Schubert, 2011).

CNS Embryonal Tumor NOS: It is the most common type of brain tumor found in children <3 years of age. By nature, this type of tumor is malignant, and it exists in the area of the cerebellum as a solid mass (Jones, 2021).

Oligodendrogliomas: They emerge around the brain and cortex, the brighter white portion of the brain. They are most commonly considered the middle-aged adult's tumor (Gaillard, 2010).

Hemangioblastomas: They are benign brain tumors that mostly rise around the brain, spinal cord, and behind the eye tissues (retina). They normally occur in young and middle-aged people (Balachandran, 2010).

Healthy brain MRI images: An MRI image without any diagnosis of a brain tumor. These brain MRI images are considered to have no tumor signs present. The shape, nature, and performance of the brain cells are normal in the healthy brain MRI images. There is no sign of abnormality, nor are any abnormal tissues found inside the healthy brain MRI images.

Figure 1 shows a sample of images from the brain MRI dataset, including healthy brain MRI images.

3.2 Research methodology

Brain tumors exhibit distinct variations in size, intensity, and contrast compared to normal tissue in MRI scans, which are critical for tumor classification. Leveraging deep learning, which performs complex computations on vast datasets through artificial neural networks, mimics human cognitive processes for automated learning.

In this study, we propose a binary convolutional neural network (BCNN) tailored for the segmentation of various brain tumor types in MRI imagery. BCNNs, characterized by multiple processing layers, excel in extracting features from extensive data, making them ideal for detailed feature analysis within large datasets.

Our methodology unfolds in two stages. In the first stage, image acquisition and preprocessing are conducted, where images are resized and their intensity or contrast levels adjusted. Subsequent steps involve converting these preprocessed images into binary format using adaptive thresholding, followed by morphological operations to delineate and classify the tumor regions by grade.

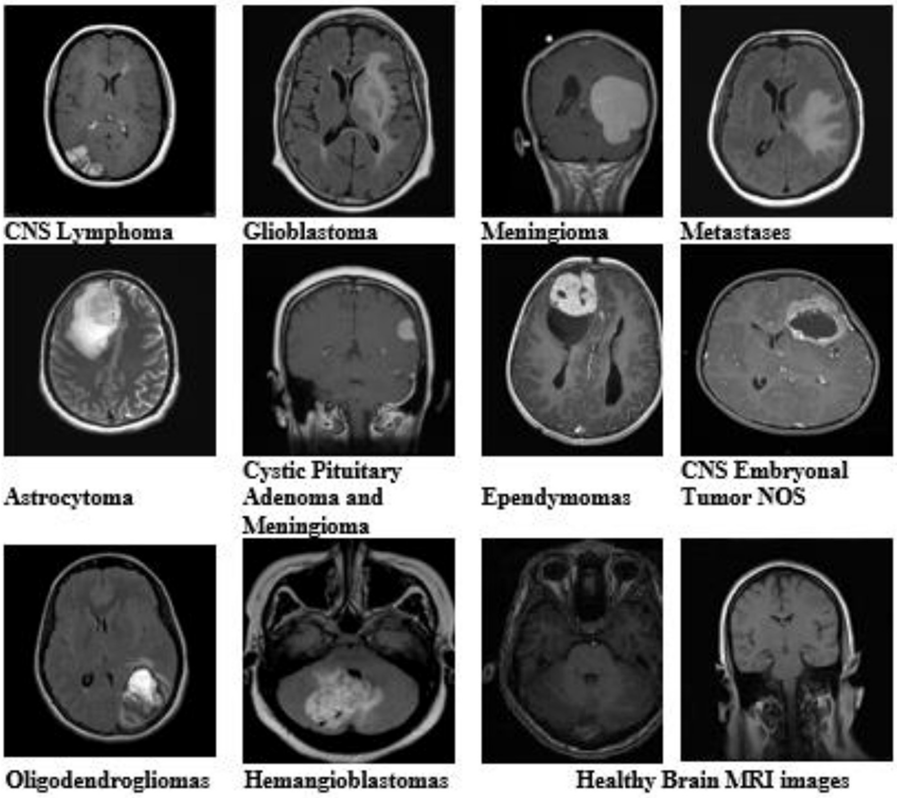


FIGURE 1
Brain MRI images of all tumor types and healthy brain MRI images from the dataset with different variants and modalities.

In the second stage, the preprocessed images, now segmented, undergo further classification into specific tumor grades, from Grade I to Grade IV, utilizing the size characteristics of the tumor regions. The data generated through preprocessing serves as the foundation for training and testing our BCNN model. Approximately 80% to 90% of our labeled dataset, encompassing both tumor-afflicted and healthy brain MRI scans, is allocated for model training. The remaining 10% to 20% of segmented but unlabeled data is reserved for model testing, assessing the BCNN's efficacy in classifying the data into correct tumor grades or identifying healthy tissue. The data is labeled initially with the support of experts from the medical field, especially a senior FCPS neurosurgeon. The labeling processing with the help of medical experts gives us a clear view and understanding of tumor size, growth, grades, and different modalities and variations of brain tumor and MRI images. Ultimately, the output will distinguish between accurate (true segmentation) and inaccurate (false segmentation) classifications, with the model striving to precisely categorize each MRI scan into the appropriate tumor grade or as a healthy brain image. The comprehensive flow of these phases is depicted in Figure 2, illustrating each step from initial preprocessing to final classification in our proposed research methodology. Using this methodology, we accurately segment brain tumors from healthy brain MRI images and also classify the ten most common types of tumor into their respective grades (Grade 1 to Grade IV).

3.2.1 Preprocessing

The MRI images are captured under different light conditions, and their sizes differ. Therefore, during preprocessing, we convert the images into equal sizes, normalize the lighting effect, or set the contrast of the image. This will help us convert an image into a perfect condition to get more accurate results during segmentation. We convert each image to an equal size of 600*600 pixels dimensions. The resizing of the image uses the nearest-neighbor interpolation. Once we give an image to the algorithm, Matlab converts the image automatically into the defined dimensions of 600*600.

3.2.1.1 Image resizing—Nearest-neighbor interpolation

Nearest-neighbor interpolation is a simple yet effective method for quickly converting grayscale images into different sizes as required. In this method, the known value of the nearest pixel is taken without paying attention to other pixels.

For instance, an image of 2 × 2 pixels (see Figure 3) can be converted and its size maximized to 4 × 4 pixels (the size of the image grid becomes 4 × 4 pixels).

For example, we have an image of the following size and pixels:

In this case, we know the value of a few pixels in 2 × 2, which is then converted into 4 × 4 to interpolate other unknown pixels (red circle) (see Figure 4).

Now, we will need to find the value of an unknown pixel in the red circle. The nearest neighbor value of the unknown pixel value in the red circle in Figure 4 is 4, which is known, so the value

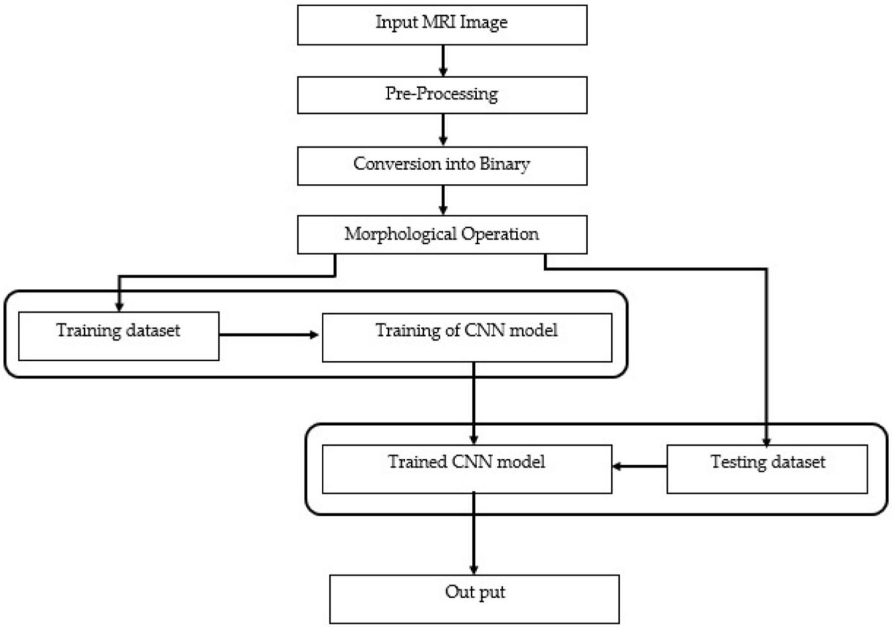


FIGURE 2
A step-by-step flow diagram of the proposed research methodology.

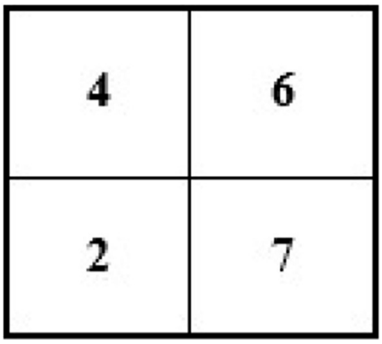


FIGURE 3
Showing 2 × 2 size original image with the pixel value.

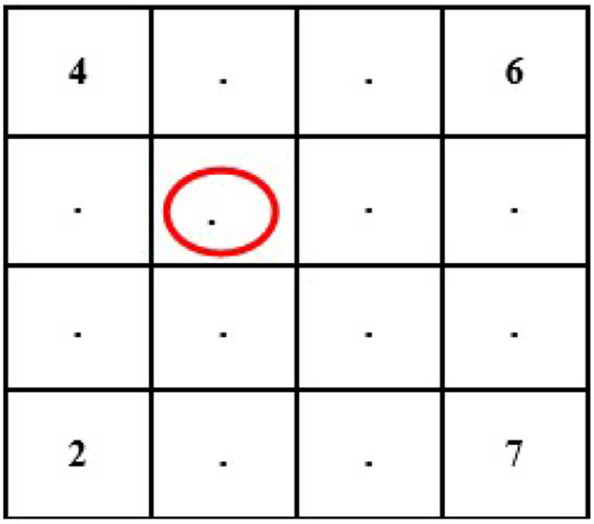


FIGURE 4
Showing 4 × 4 size images with unknown pixel values.

of the unknown pixel in the red circle will become 4. Similarly, the remaining unknown pixel values will be filled out in the same fashion (see Figure 5).

If we have an image of 4 × 4 and want to minimize its size to 2 × 2, then the new size of the image will become 2 × 2 (see Figure 6).

To reduce the size of the above image from 4 × 4 to 2 × 2, we remove every second row and second column (see Figure 7).

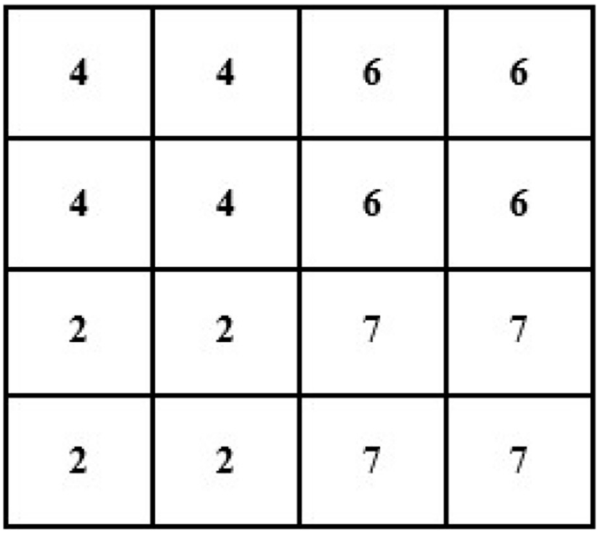
After removing the second row and second column, we will have the following image with a 2 × 2 size (see Figure 8).

3.2.1.2 Image filtering using two-dimensional adaptive filter

The two-dimensional adaptive filter estimates the local mean and variance around each pixel. It reduces the mean square error as much as possible [see Equation (1) and Equation (2)].

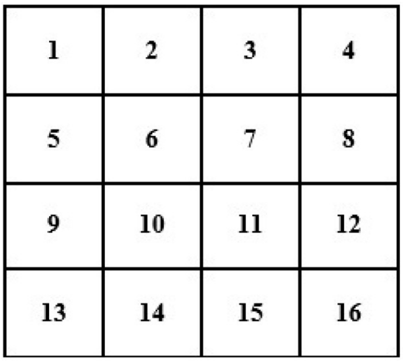
$$\mu = \frac{1}{NM} \sum_{n_1,n_2 \in \eta} a\left(n_1,n_2\right) \tag{1}$$

$$\sigma^2 = \frac{1}{NM} \sum_{n_1,\, n_2 \,\epsilon\, \eta} a^2\left(n_1,n_2\right) - \mu^2 \tag{2}$$



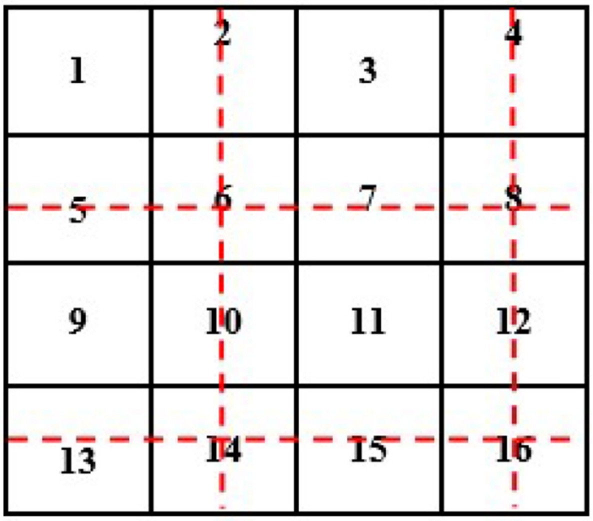
4	4	6	6
4	4	6	6
2	2	7	7
2	2	7	7

FIGURE 5
Showing 4 × 4 size images after interpolation of known pixel values.



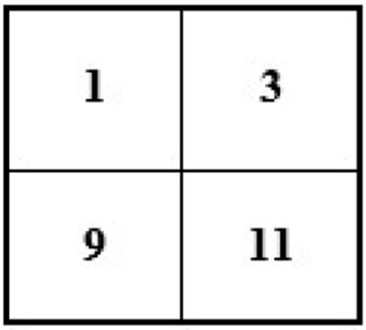
1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

FIGURE 6
Showing original 4 × 4 size image with known pixel values.



1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

FIGURE 7
Showing 4 × 4 original size image conversion into 2 × 2.



1	3
9	11

FIGURE 8
Showing 2 × 2 size image after interpolation/conversion from 4 × 4.

Where the N-by-M local neighborhood of each pixel is in image A, the two-dimensional adaptive filter creates a pixel-wise Wiener filter using these estimates [see Equation (3)].

$$b(n_1, n_2) = \mu + \frac{\sigma^2 - \nu^2}{\sigma^2} (a(n_1, n_2) - \mu) \tag{3}$$

Where ν^2 is the noise variance. If the noise variance is not given, the two-dimensional adaptive filter uses the average of all the local estimated variances.

3.2.1.3 Conversion into binary and morphological operation

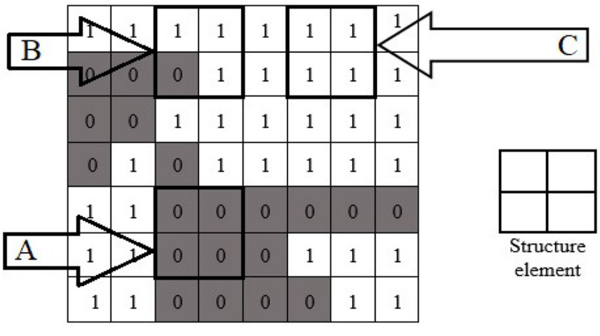
Once the preprocessing step is completed, the image is converted into binary, and the morphological operations are performed. The grayscale image is converted into binary using the adaptive thresholding method. A threshold value is calculated locally using the mean of the neighborhood pixels using a filter; if the pixel value is above the threshold, it will be considered a foreground value or one; otherwise, it will be considered a background value or zero. In this method, a mean filter around the neighborhood is subtracted from a constant value of the pixels to find the foreground pixels.

For example:

T is our threshold value for the output image, M is set to be our threshold value for the mean filter of the neighborhood, and C is our constant value to be subtracted from T. Ultimately, it will give us a new binary image with foreground values as our resultant image (Equation 4).

$$T \text{ threshold} = M \text{ mean of neighboorhood pixels} - C \text{ constant} \tag{4}$$

Morphological operations remove any noise or unwanted pixels that may cause errors or produce false segmentation results. The binary-converted image shows two types of pixels—black and



1	1	1	1	1	1	1	1
0	0	0	1	1	1	1	1
0	0	1	1	1	1	1	1
0	1	0	1	1	1	1	1
1	1	0	0	0	0	0	0
1	1	0	0	0	1	1	1
1	1	0	0	0	0	1	1

FIGURE 9
Morphological operation's structure element with fit **(A)**, hit **(B)**, and neither fit nor hit **(C)** the image.

white. The white pixels show the foreground pixels or the targeted pixels, which include the tumor area and some other extra and unwanted noisy pixels, such as the boundary of images. We need to remove the extra and unwanted noisy pixels to focus attention on the tumor pixels. This will help us keep the tumor pixels as foreground pixels and use them further during segmentation.

Morphological image processing is a collection of non-linear operations related to the shape or morphology of features in an image. Morphological techniques probe an image with a small shape or template called a structuring element. The structure element is a pre-defined matrix or binary image with values 0 and 1, which is used to probe the image. The structuring element is positioned at all possible locations in the image and compared with the corresponding neighborhood of pixels. Some operations test whether the element "fits" within the neighborhood, while others test whether it "hits" or intersects the neighborhood. The actual structure element that we used for erosion and dilation was 22*22. An example of using a structure element can be seen in Figure 9.

When a structuring element is placed in a binary image, each of its pixels is associated with the corresponding neighborhood pixel under the structuring element. The structuring element is said to fit the image if, for each pixel set to 1, the corresponding image pixel is also 1. Similarly, a structuring element is said to hit or intersect an image if, at least for one of its pixels set to 1, the corresponding image pixel is also 1. In Figure 9, the structure element is the design of 2 × 2. The structure element fits at location "A" because the structure elements of 2 × 2 all fit the position "A" of the image. The structure element hit the image at position "B." The structure element neither hit nor fit at position "C."

The erosion of a binary image f by a structuring element s (denoted $f \ominus s$) produces a new binary image $g = f \ominus s$ with ones in all locations (x, y) of a structuring element's origin at which that structuring element s fits the input image f, i.e., $g(x, y) = 1$ is s fits f and 0 otherwise, repeating for all pixel coordinates (x, y). The erosion shrinks an image by removing a layer of pixels from the inner and outer boundaries of image regions. The holes and gaps between different regions of the image become larger, and small details or noise get eliminated. Erosion removes small-scale details from the binary image but simultaneously reduces the size of regions of interest. We perform a dilation operation in the morphology to maintain the foreground pixels or the tumor pixels. It adds the region of tumor pixels removed in the erosion stage. The dilation of an image f by a structuring element s (denoted $f \oplus s$) produces a new binary image $g = f \oplus s$ with ones in all locations (x, y) of a structuring element's origin at which that structuring element s hits the input image f, i.e., $g(x, y) = 1$ if s hits f and 0 otherwise, repeating for all pixel coordinates (x, y). Dilation has the opposite effect of erosion—it adds a layer of pixels to both the inner and outer boundaries of the regions.

3.2.1.4 Binary convolution neural network—BCNN

After the binarization and morphological operations, we used BCNN as the main part of our methodology. The BCNN is used to classify the tumor of every tumor type used in this study into grades. Because BCNN works on the binary images generated from the morphological steps, all the binary images generated from the morphological steps are saved in different folders labeled with tumor grades (from Grade I to Grade IV) and the healthy brain MRI images folder. Labeling all the folders of binary images from Grade I to Grade IV tumors and healthy brain MRI images is for the training of the BCNN.

The BCNN stores values in binary formats 1 and 0. This process, known as 1-bit quantization, saves memory, increases the processing speed of the network, and reduces memory access time. Overall, it is fast in computation and uses less memory. The binary images or data used to train our neural network are most suitable for embedded and microcontroller devices.

The general weights of CNN depend on grayscale and color images, which have three values. In our BCNN, a binarization function is used to binarize those values. The two functions used are sign and stochastic [see equations (5–7)].

$$x^{b} = Sign\,(x) = \begin{cases} +1 \ \ if\ x\ \geq 0\,, \\ -1\ otherwise, \end{cases} \tag{5}$$

$$x^{b} = \begin{matrix} +1\ \ with\ probability\ p = \sigma\,(x)\,, \\ -1\ \ with\ probability\ 1 - p, \end{matrix} \tag{6}$$

$$\sigma\,(x)\ clip\left(\frac{x+1}{2},0,1\right)\max\left(0,\min\left(1,\frac{x+1}{2}\right)\right) \tag{7}$$

Our BCNN has three main convolution layers and one fully connected layer. The three convolution layers include the input convolution layer, pooling layer, and batch normalization layer. The basic structure of the neural network can be seen in Figure 10.

The three main convolution layers and one fully connected layer that are used to build our BCNN are discussed in detail below.

A. **Convolution layer**

The convolution layer is the first layer used in a BCNN. It gets the input matrix of dimensions, which includes the elements H1 x W1 x D1; H1 is the height of the matrix, W1 is the width of the matrix, and D1 is the dimension of the matrix. Next, we have kernels (structure elements or filters) in the convolution layer. A kernel is a matrix with dimensions H2 x W2 x D2. A

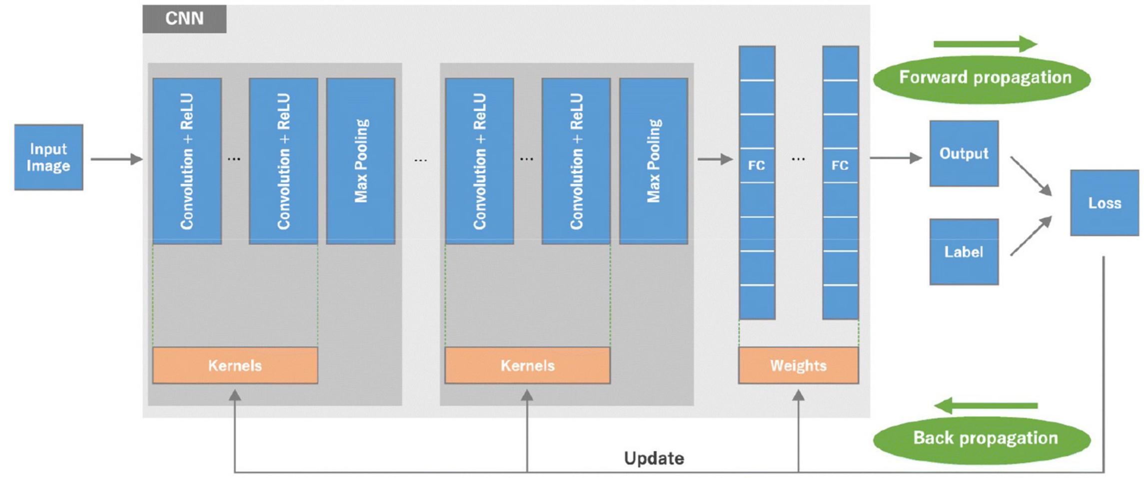


FIGURE 10
The basic structure of the neural network.

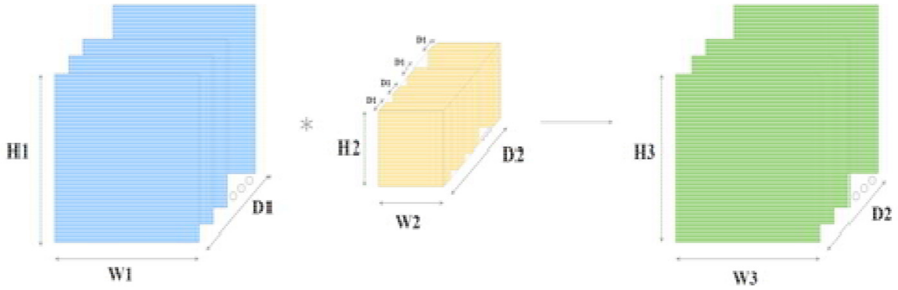


FIGURE 11
Convolution layer with input, kernel, and output layer; blue is the input to the convolution layer, yellow is the kernel (filter), and green is the output of the convolution layer.

convolution layer has multiple kernels placed on top of each other in a sequence. These multiple kernels above each other create a 3-dimensional matrix; D2 is the number of dimensions. At the end of the convolution layer is the output layer with the dimensions H3 x W3 x D2. A detailed representation of the three sections of the convolution layer is presented in Figure 11.

B. **Pooling layer**

The pooling layer minimizes features of the in-plane dimensionality to make a new invariance to small changes and misrepresentations and minimize the upcoming parameters. The advantage of using a max pooling layer is that it minimizes the number of parameters of the input plot and minimizes overfitting, extracts important features from the input plot, minimizes computation, and, therefore, introduces maximum efficiency. In the pooling layer, there are no such learnable parameters. Filters, padding, and strides are used as hyperparameters in the pooling layer, similar to convolution layers.

The max pooling method is used in our pooling layer in BCNN. The max pooling operation is used to extract patches from the input features plot, and an output produces a new plot extracting the maximum number in each plot and discards the rest of the values. Other than height and width, this will not change the depth dimension of the feature map. Figure 12 presents a 4 × 4 input plot while extracting a new 2 × 2 plot, extracting the maximum value in each plot by using a 2 × 2 filter on the input plot.

C. **Batch normalization layer**

Before training, batch normalization is used as a preprocessing step in the neural network. It improves the learning capability of the network and also avoids overfitting. It converts the data into a standard format before the training phase of the network to make the training phase easy. Since we have different types of MRI images in terms of size, shape, and intensity, it would be challenging for the network to train on such diverse data. Moreover, this would make the network more complicated and less efficient, and we may not be able to learn 100% as per our target. Consequently, the overall capability of the network would be decreased.

The approach we use inside the batch normalization layer is to scale it to a selection from 0 to 1. In Equation 7, x is the facts factor to normalize, m is the mean of the data set, x_{max} is the maximum value, and x_{min} is the minimum fee. This technique is normally used in data input. The non-normalized data inputs with huge variations can produce instability in neural networks. The tremendously big inputs can cascade down to the layers, inflicting problems that include exploding gradients [see Equation (8)].

$$x_{normilization} = \frac{x - m}{x_{\max} - x_{\min}} \tag{8}$$

D. **One fully connected layer**

It works as a feed-forward layer in the neural network. It receives input from the previous pooling and batch normalization and is forwarded for further processing. It is a unidirectional layer that receives input from one direction and forwards in the same direction without using any repetition or loop. The input received by the fully connected layer from the previous layers is in a vector format. The fully connected layer has hidden layers, which are a combination of affine and non-linear functions. The one affine and one non-linear layer is called one fully connected layer or one hidden layer. We can add additional fully connected layers or hidden layers as per the requirement of our segmentation model.

The calculation in Equation (9) is used for every fully connected layer of the neural network. In Equation (9), x represents the input from the previous layer as the input vector, w is the weight matrix with dimensions, b is the bias vector, and g is the activation function, usually ReLU [see Equation (9)].

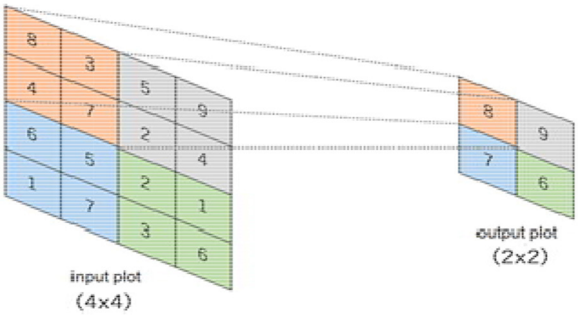


FIGURE 12
Showing pooling method functionality.

$$g\,(wx + b) \tag{9}$$

After the process completion of the fully connected layer, and once it passes from the last layer, it is used to calculate the probability and classify the values into their respective class. Finally, we get the probability of the object or input data in the class to which it belongs. That is how the overall neural network works. The mechanism of the neural network is displayed in Figure 13, where the pooling and batch normalization work as feature selection layers, and the rest of the section of the neural network is used as a part of the segmentation layer, which includes a flattening layer, a fully connected layer, and the final output layer.

4 Results and discussion

This section discusses the complete step-by-step results of our proposed research work. The software used for the implementation and analysis of our proposed research methodology is also explained in this section.

We used Matlab version 2021 software in our research to implement our proposed methodology and analysis. Matlab provides state-of-the-art functionality and facilitation to easily implement and analyze complex and tricky methodologies. We used an HP Core-i5 7th generation computer/laptop to run the software and generate the overall results. The detailed implementation and the results of our proposed research work are discussed below.

The proposed methodology is implemented in two phases. In phase one, preprocessing, conversion into binary, and morphological operations are applied to the input images. In the second phase, the tumor types are segmented into grades.

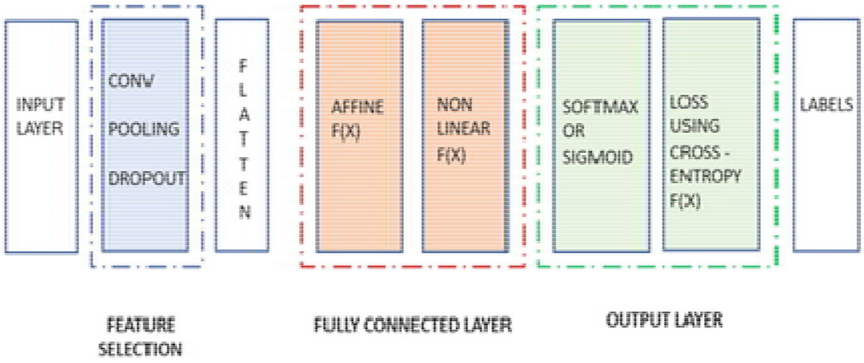


FIGURE 13
Different sections of the neural network.

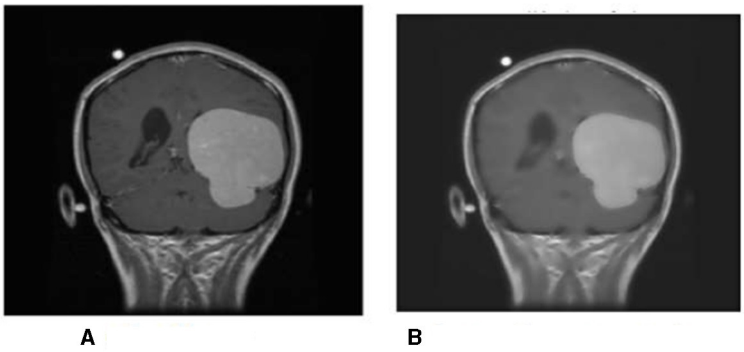


FIGURE 14
(A) The original image from the data set; **(B)** The resultant image after preprocessing.

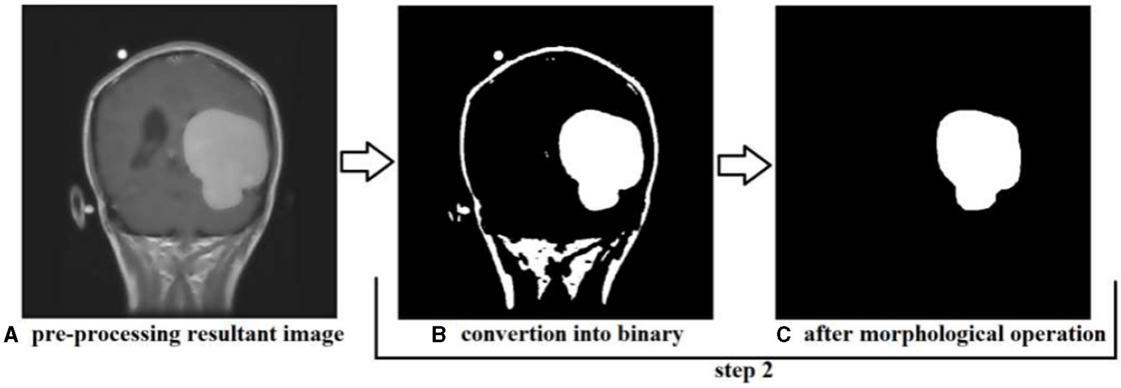
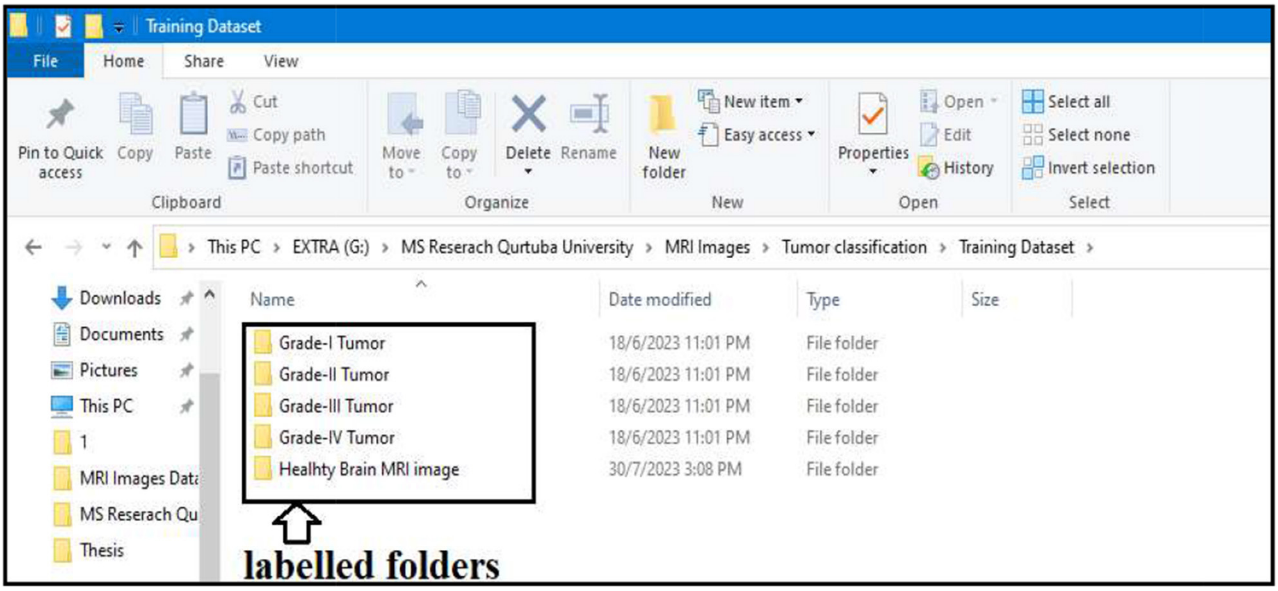


FIGURE 15
Showing the results produced after performing step 2, **(A)** is the resultant image of step 1, **(B)** is the resultant image converted into binary, and **(C)** is the image after performing morphological operation erosion and dilation.



This PC › EXTRA (G:) › MS Reserach Qurtuba University › MRI Images › Tumor classification › Training Dataset ›

Name	Date modified	Type	Size
Grade-I Tumor	18/6/2023 11:01 PM	File folder	
Grade-II Tumor	18/6/2023 11:01 PM	File folder	
Grade-III Tumor	18/6/2023 11:01 PM	File folder	
Grade-IV Tumor	18/6/2023 11:01 PM	File folder	
Healhty Brain MRI image	30/7/2023 3:08 PM	File folder	

FIGURE 16
Showing labeled folders.

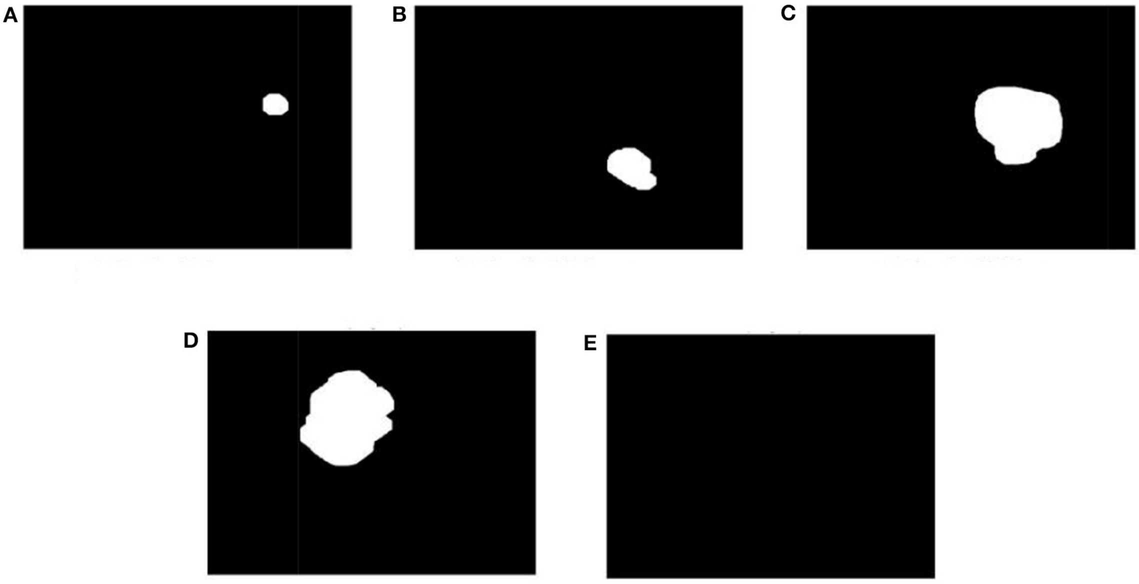
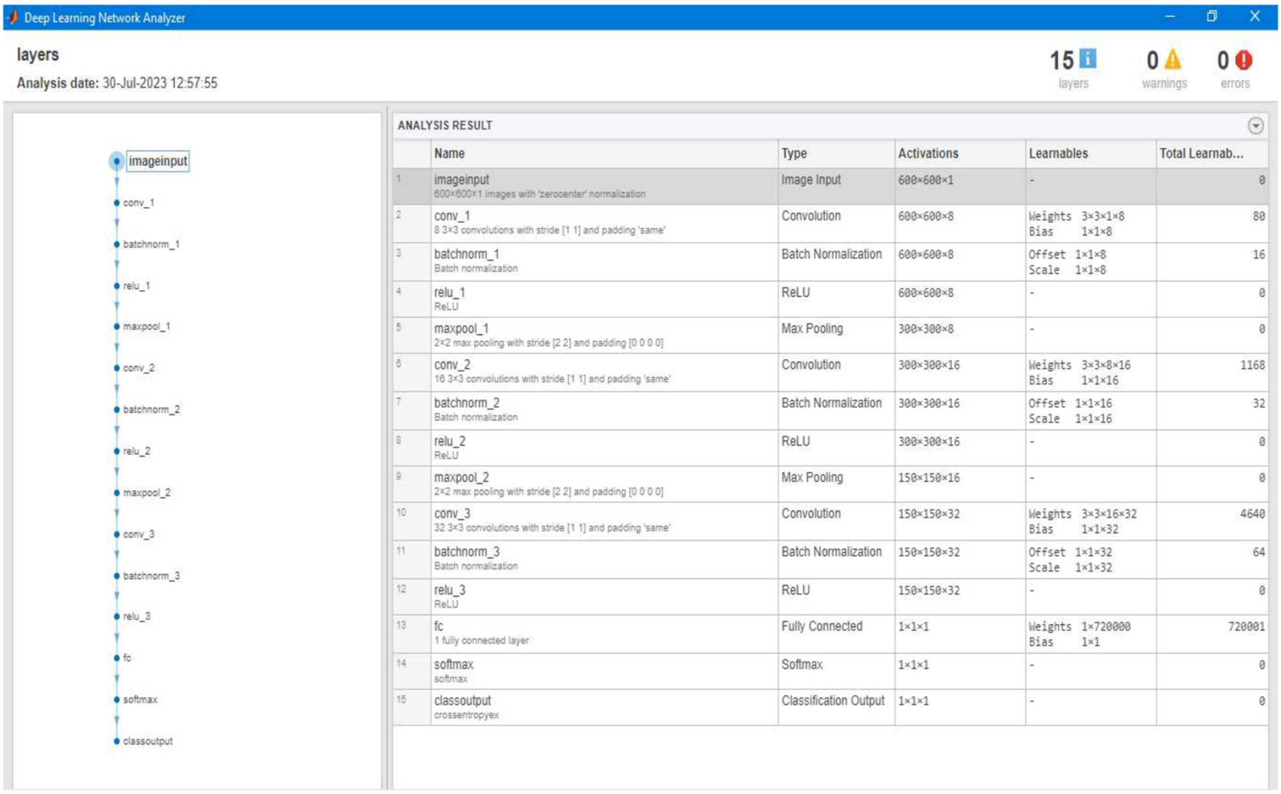


FIGURE 17
Display a few images from all the labeled folders. **(A)** Grade I tumor. **(B)** Grade II tumor. **(C)** Grade III tumor. **(D)** Grade IV tumor. **(E)** Healthy brain MRI image.

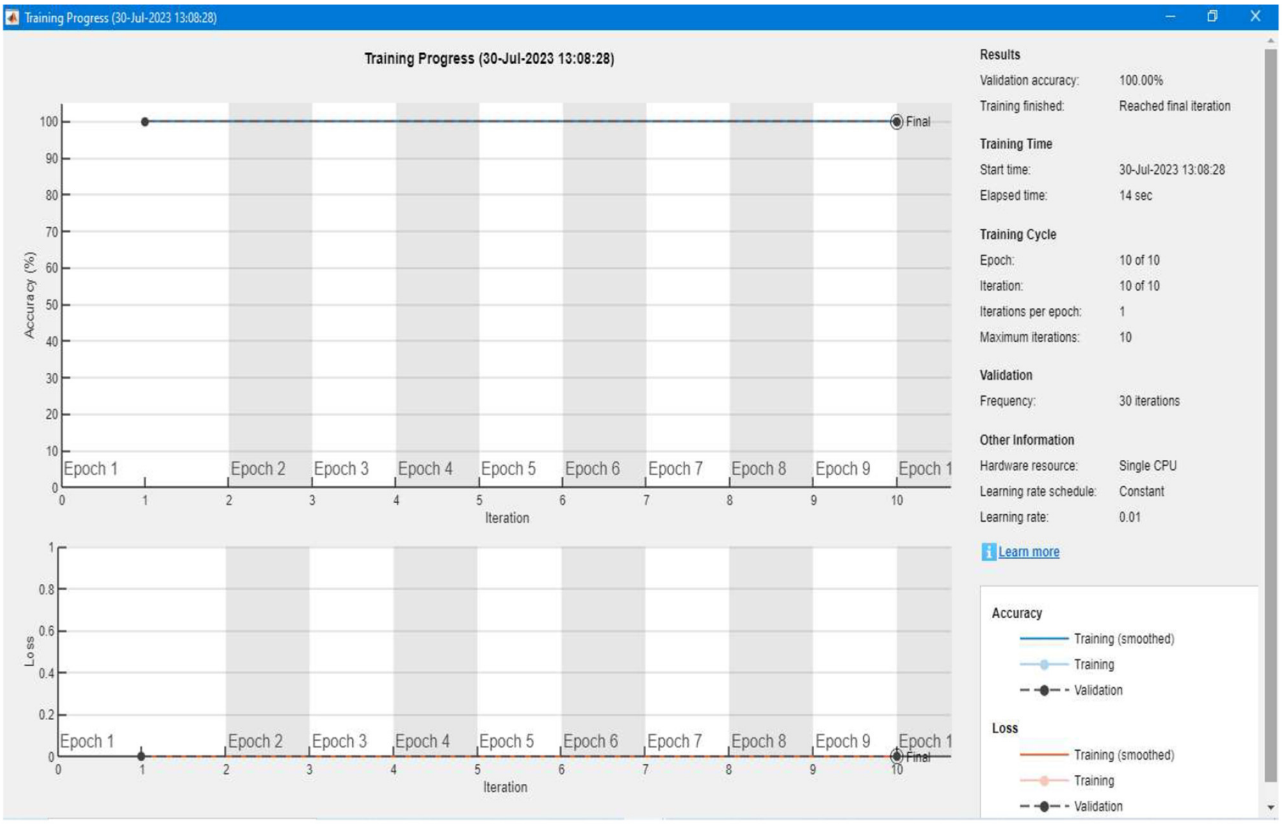


Analysis date: 30-Jul-2023 12:57:55

ANALYSIS RESULT

	Name	Type	Activations	Learnables	Total Learnab...
1	imageinput 600×600×1 images with 'zerocenter' normalization	Image Input	600×600×1	-	0
2	conv_1 8 3×3 convolutions with stride [1 1] and padding 'same'	Convolution	600×600×8	Weights 3×3×1×8 Bias 1×1×8	80
3	batchnorm_1 Batch normalization	Batch Normalization	600×600×8	Offset 1×1×8 Scale 1×1×8	16
4	relu_1 ReLU	ReLU	600×600×8	-	0
5	maxpool_1 2×2 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	300×300×8	-	0
6	conv_2 16 3×3 convolutions with stride [1 1] and padding 'same'	Convolution	300×300×16	Weights 3×3×8×16 Bias 1×1×16	1168
7	batchnorm_2 Batch normalization	Batch Normalization	300×300×16	Offset 1×1×16 Scale 1×1×16	32
8	relu_2 ReLU	ReLU	300×300×16	-	0
9	maxpool_2 2×2 max pooling with stride [2 2] and padding [0 0 0 0]	Max Pooling	150×150×16	-	0
10	conv_3 32 3×3 convolutions with stride [1 1] and padding 'same'	Convolution	150×150×32	Weights 3×3×16×32 Bias 1×1×32	4640
11	batchnorm_3 Batch normalization	Batch Normalization	150×150×32	Offset 1×1×32 Scale 1×1×32	64
12	relu_3 ReLU	ReLU	150×150×32	-	0
13	fc 1 fully connected layer	Fully Connected	1×1×1	Weights 1×720000 Bias 1×1	720001
14	softmax softmax	Softmax	1×1×1	-	0
15	classoutput crossentropyex	Classification Output	1×1×1	-	0

FIGURE 18
Showing different layers of the overall training model constructed to train the segmentation model.



Training Progress (30-Jul-2023 13:08:28)

Results
Validation accuracy: 100.00%
Training finished: Reached final iteration

Training Time
Start time: 30-Jul-2023 13:08:28
Elapsed time: 14 sec

Training Cycle
Epoch: 10 of 10
Iteration: 10 of 10
Iterations per epoch: 1
Maximum iterations: 10

Validation
Frequency: 30 iterations

Other Information
Hardware resource: Single CPU
Learning rate schedule: Constant
Learning rate: 0.01

FIGURE 19
Showing accuracy and validation of the training phase.

4.1 Implementation and results of phase one

4.1.1 Step one

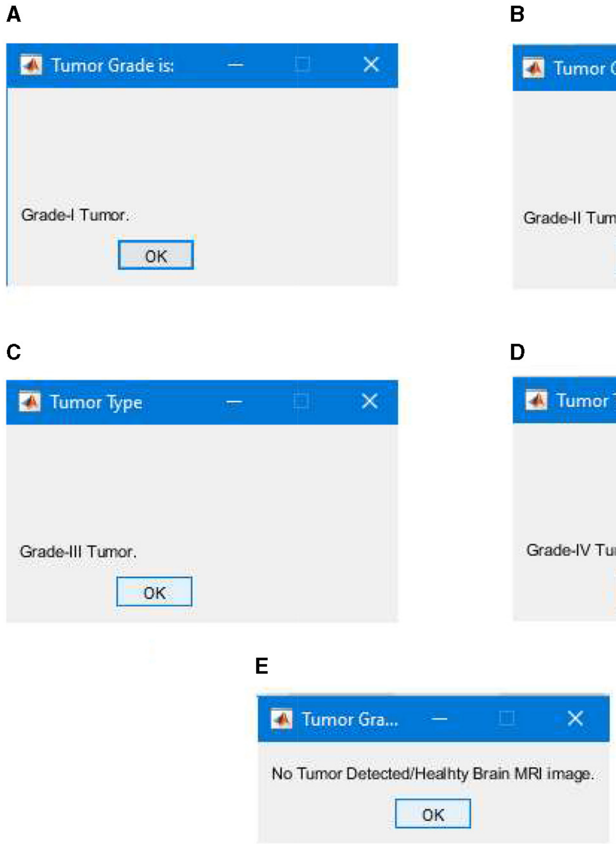
First, the brain tumor MRI image from the dataset is passed to the first step of our methodology, which is preprocessing. The preprocessing step is performed to remove any unwanted impurities, enhance the quality of the image, and convert all the images to a standard and equal size. Performing the preprocessing step provides better results at the end of the segmentation. The images we received from the two sources differed in size, intensity, and lighting. The preprocessing step is performed to make the images more useful. The difference in the original image (a) and the resultant image (b) after preprocessing is shown in Figure 14.

The intensity and size of both the images are different. The original image size is 812*812 pixels, and the resultant image size is 600*600 pixels. Preprocessing is performed to decrease or increase the size of the image to a standard size of 600*600 pixels.

4.1.2 Step two

In step two, the resultant image is further processed and converted to binary format, and morphological operations are performed; the conversion into binary format is undertaken to make the segmentation phase more efficient and faster. This step is performed to identify the tumor region and remove the boundaries or skull region to easily classify the tumor region. Once the image is converted to binary, morphological operations such as erosion and dilation are performed. Once the preprocessing resultant image is converted into binary, then there are only two pixel values, 0 and 1, left to deal with; 0 represents the background pixels or of the image which are not our targeted pixels, and 1 represents foreground pixels which are our target pixels. These pixels represent the tumor pixels and the boundaries and extra foreground pixels of the image that need to be removed to retain only the tumor foreground pixels.

Applying the erosion removes the extra or unwanted foreground pixels. This process also removes the boundaries of the skull in the image and some other parts. The erosion also affects the tumor foreground pixels, which affects the exact size of the tumor. To regain the actual size of the tumor size after erosion, we perform dilation; with the help of dilation, the tumor area that is removed or affected during erosion gets added to the tumor region again. This helps us to maintain the tumor region's original size and get accurate results in the segmentation. The results produced during step two of conversion into binary and morphological operations are shown in Figure 15. We can see that the tumor region in Figure 15A, which is the resultant image of step 1, can also be seen in Figure 15C after performing morphological operations.



Grade-I Tumor.

Grade-II Tumor.

Grade-III Tumor.

Grade-IV Tumour.

No Tumor Detected/Healhty Brain MRI image.

FIGURE 20
Segmentation results classifying tumor types into grades, **(A)** successful segmentation results of Grade I, **(B)** successful segmentation results of Grade II, **(C)** successful segmentation results of Grade III, **(D)** successful segmentation results of Grade IV, and **(E)** successful segmentation results of healthy brain MRI image with no brain tumor detected.

4.2 Implementation and results of phase two

In step two, the actual segmentation of the tumor types into grades is performed. To start the segmentation, we first labeled all the resultant images produced after the morphological operation of step two. All the images were stored in different folders according to the grades of the tumor; the folders were labeled Grade I Tumor, Grade II Tumor, Grade III Tumor, Grade IV Tumor, and Healthy Brain MRI image.

Labeling is performed to train our BCNN segmentation model. The model will train on the images stored in binary format in all the labeled folders. All the images stored in the labeled folders are in binary format. Approximately 90% of the resultant images processed from our actual brain MRI images are used to train our segmentation model. The segmentation model is trained to classify tumor types into their respective grades.

The actual performance of our proposed segmentation model includes the following steps:

4.2.1 Preparation of labeled data

All the resultant images of step two are divided into two categories. One category is labeled images, and the second is unlabeled images. As we can see in Figure 16, the labeled images are used to train our segmentation model. The labeled images are stored in separate folders (Grade I Tumor, Grade II Tumor, Grade III Tumor, Grade IV Tumor, and Healthy Brain MRI image). The remaining unlabeled images will be used in the testing phase of the segmentation model. Approximately 90% of the images were labeled for training purposes, and the remaining 10% was unlabeled and allocated for testing the segmentation model.

4.2.2 Loading all labeled images

All images stored in the labeled folders are loaded into the segmentation model for training purposes. They are loaded at once and stored as a variable described in the algorithm implementation; the labeled images stored as a variable are further used to train our segmentation model. Figure 17 displays a few images from all the labeled folders stored in a variable.

4.2.3 Development of the training phase

After labeling and loading the data, the segmentation model will train itself based on the parameter/labeled images. The model trains based on the size of tumor types already set in the labeled data and stored as a variable in the previous step. Figure 18 shows the different layers of the training phase of our segmentation model.

TABLE 2 Comparison of proposed and existing models.

Method	No. of tumor types	Segmentation of tumor into grades	Segmentation results			
			Accuracy	Precision	Recall	F-Measure
Convolutional neural networks (CNN) using grid search optimization (Irmak, 2021)	A total of five tumor types	Only Glioma brain tumors are classified into grades II, III, and IV	98.14%	98.31%	——	——
Multiclass support vector machine (M-SVM) (Maqsood et al., 2022)	Three tumor types	——	98.92%	——	——	——
Four models of deep convolutional neural networks: inceptionresnetv2, inceptionv3, transfer learning, brain-tumor-net model (Shoaib et al., 2022)	Three tumor types and normal cases	——	86.80%, 85.34%, 93.15%, 91.24%	86.85%, 85.12%, 93.14%, 91.20%	——	86.83%, 84.98%, 93.11%, 91.08%
Histogram differencing and KNN (Nida-Ur-Rehman et al., 2017)	Four tumor types and healthy brain MRI images	——	97.3%	——	——	——
Proposed Model: binary convolution neural network (BCNN)	Ten tumor types and healthy brain MRI images	Into four grades (Grade I to Grade IV) and healthy brain MRI images	99.40 %	99.32%	99.45%	99.28%

The training model consists of three convolution layers and one fully connected layer.

4.2.4 Training the segmentation model

After setting up the labeled data and developing the training model, the algorithm trains itself using the labeled data. Figure 19 shows a detailed representation of the training and validation of our proposed segmentation model. We can see that the accuracy of the training is 100%, and there is a 0% loss in the validation of the data. For each image, the model takes an average of 10 to 12 seconds of computation time to complete the overall training.

4.2.5 Testing and segmentation results

After completing the training, we tested our segmentation model; the testing was performed using the unlabeled images comprising 10% of the total images/data. The testing phase of our segmentation model achieved an overall true segmentation rate of 100%. All the tumor types were successfully classified into their respective grades (from Grade I to Grade IV). The model also accurately categorized the healthy brain MRI images, distinguishing them from those with tumors. The overall achievement of our proposed methodology showed significance and efficacy. Figure 20 illustrates a few results generated during the segmentation of the proposed segmentation model. We can see that the tumor regions in the brain MRI images are successfully classified into their respective grades, and the healthy brain MRI image is also successfully classified without detecting any tumor region.

4.3 Comparison of proposed and current research work

In this section, we compare our proposed model with existing research. The criteria for comparing are the overall accuracy of the segmentation model, precision, recall, and F-measure. In Table 2, the results of our proposed model are compared with existing research. Based on the comparison, it is evident that our proposed research model performs better than existing research models.

5 Conclusion and future work

In this study, we proposed a new model based on deep learning BCNN to classify the most common ten types of brain tumors into grades (from Grade I to Grade IV) based on the size and growth of the tumor. The tumor types that were considered in our research work are metastatic or secondary tumors, Meningioma, CNS Lymphoma, Glioblastoma, Astrocytoma, Pituitary Adenoma, Ependymomas, Medulloblastomas, Oligodendroglia's, and Hemangioblastomas. A dataset of 6,600 MRI images was used, including all types of tumor MRI images and healthy brain MRI images. The dataset was collected from the two main sources (Nida-Ur-Rehman et al., 2017; Radiopaedia's, 2023) and verified by an FCPS neurosurgeon for their validity.

The methodology that we proposed had two phases. The first phase consisted of preprocessing and conversion to binary and morphological operations. The image was loaded in step one, and its intensity and contrast were set to get more accurate results in the next step and phase two. We used the two-dimensional adaptive filter to filter and set up the image contrast for this. Next, during preprocessing, we standardized the size of all the MRI images to 600*600 using the nearest neighbor interpolation and converted all the images to a standard size. In step two of phase one, all the images were converted into binary. Morphological operations were performed to remove any noise or unwanted pixels that could result in errors or give false segmentation results.

In step two, we executed the main segmentation model based on the BCNN developed using three convolution layers and one fully connected layer. The images that were generated after step two of phase one were divided into two categories: labeled and unlabeled. The labeled images were stored in folders named according to tumor grades (Grade I to Grade IV) and healthy brain MRI images. The labeled images were used to train our algorithm; overall, 90% of the images were labeled and used to train our algorithm. In the next step, we tested our algorithm using the remaining 10% of images that were unlabeled.

The overall results of our segmentation model were very satisfactory, with 99.40% accuracy, 99.32% precision, 99.45% recall, and 99.28% F-Measure score. These results demonstrate the significance and efficacy of our proposed model, which successfully classified all the tumor types into their respective grades (from Grade I to Grade IV). Our study has also curated a new dataset for the research community with more than 6,000 brain MRI images that contain the 10 most common types of brain tumor and healthy brain MRI images. This robust framework enhances the accuracy of brain tumor segmentation and sets a new benchmark for early detection and grading of brain tumors, thereby contributing to the advancement of neuro-oncological diagnostics. The technique provides a more effective contribution to the clinical practitioner to easily, quickly, and accurately classify and segment brain tumors in the early stages, which will help them provide better treatment to patients suffering from this deadly disease. This model can effectively reduce the number of deaths caused by brain tumors by facilitating early and accurate detection of brain tumors and treatment.

5.1 Future work

Going forward, this study will aim to address the following challenges and limitations of the proposed research work and methodology:

- There are more than 100 types of brain tumors; these can be included for segmentation into grades.
- Other imaging technologies, such as CT scans, can be used along with MRI images to classify the brain using different deep learning techniques.
- Different deep-learning techniques can be used to classify other tumor types effectively using RGB images.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Author contributions

AZ: Conceptualization, Supervision, Writing – review & editing. LK: Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. QR: Formal analysis, Methodology, Writing – original draft. TR: Conceptualization, Data curation, Writing – review & editing. MS: Conceptualization, Validation, Writing – review & editing. SA: Funding acquisition, Visualization, Writing – review & editing. MI: Conceptualization, Formal analysis, Writing – review & editing. MN: Resources, Validation, Visualization, Writing – review & editing. SH: Conceptualization, Formal analysis, Writing – review & editing. HM: Data curation, Formal analysis, Funding acquisition, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research has received funding from King Saud University through Researchers Supporting Project number (RSP2024R387), King Saud University, Riyadh, Saudi Arabia.

Acknowledgments

The authors extend their appreciation to King Saud University for funding this work through Researchers Supporting Project number (RSP2024R387), King Saud University, Riyadh, Saudi Arabia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Ahmed, I., Nida-Ur-Rehman, Q., Masood, G., and Nawaz, M. (2016a). "Analysis of brain MRI for tumor detection & segmentation," in *Proceedings of the World Congress on Engineering* (London: World Congress on Engineering), 456–461,

Ahmed, I., Nida-Ur-Rehman, Q., Masood, G., Nawaz, M., Adnan, A., and Khan, M. (2016b). Segmentation of brain tumor from healthy tissues using multimodal MRI images. *Int. J. Comp. Sci. Inform. Secu.* 14, 676—689.

Akter, A., Nosheen, N., Ahmed, S., Hossain, M., Yousuf, M. A., Almoyad, M. A. A., et al. (2024). Robust clinical applicable CNN and U-Net based algorithm for MRI classification and segmentation for brain tumor. *Expert Syst. Appl.* 238, 122347. doi: 10.1016/j.eswa.2023.122347

Amin, J., Sharif, M., Haldorai, A., Yasmin, M., and Nayak, R. S. (2021). Brain tumor detection and segmentation using Machine Learning: a comprehensive survey. *Complex Intellig. Syst.* 8, 3161–3183. doi: 10.1007/s40747-021-00563-y

Badža, M. M., and Barjaktarović, M. C. (2020). Segmentation of brain tumors from MRI images using a convolutional neural network, *Appl. Sci.* 10, 1999. doi: 10.3390/app10061999

Bahadure, N. B., Ray, A. K., and Thethi, H. P. (2018). Comparative approach of MRI-based brain tumor segmentation and segmentation using genetic algorithm. *J. Digit. Imag.* 31, 477–489. doi: 10.1007/s10278-018-0050-6

Balachandran, G. (2010). *Cerebellar hemangioblastoma*. Case study. Radiopaedia.org (accessed July 22, 2023). doi: 10.53347/rID-10779

Brusic, A. (2021). *Brain metastases - breast cancer*. Case study. Radiopaedia.org (accessed July 22, 2023). doi: 10.53347/rID-88568

Di Muzio, B. (2023). *Meningioma*. Case study. Radiopaedia.org (accessed July 22, 2023). doi: 10.53347/rID-54380

Efford, N. (2000). Digital image processing: a practical introduction using Java, Inc. 75 Arlington Street, Suite 300 Boston, MA, United States: Addison-Wesley Longman Publishing Co,

Gaillard, F. (2010). *Glioblastoma NOS - hemorrhagic*. Case study. Radiopaedia.org (accessed July 22, 2023). doi: 10.53347/rID-9969

Gaillard, F. (2016). *Cystic pituitary adenoma and meningioma*. Case study. Radiopaedia.org (accessed July 22, 2023). doi: 10.53347/rID-49382

Gaillard, F. (2018). *Glioblastoma*. Case study. Radiopaedia.org (accessed July 22, 2023). doi: 10.53347/rID-63888

Gaillard, F. (2021), *Astrocytoma*. Case study. Radiopaedia.org (accessed July 22, 2023). doi: 10.53347/rID-85660

Garg, G., and Garg, R. (2021). Brain tumor detection and classification based on hybrid ensemble classifier. *CoRR*. abs/2101.00216. Available online at: https://arxiv.org/abs/2101.00216

Interpolation Methods (2024). *Hasty.ai Documentation*. Available online at: https://hasty.ai/docs/mp-wiki/augmentations/interpolation-methods (accessed October 22, 2023).

Irmak, E. (2021). Multi-segmentation of brain tumor MRI images using deep convolutional neural network with fully optimized framework. *Iranian J. Sci. Technol. Trans. Electr. Eng.* 45, 1015–1036. doi: 10.1007/s40998-021-00426-9

Jones, J. (2021). *CNS Embryonal tumor NOS*. Case study. Radiopaedia.org (accessed July 22, 2023). doi: 10.53347/rID-90346

Le, N., Yamazaki, K., Quach, K. G., Truong, D., and Savvides, M. (2021). "A multi-task contextual atrous residual network for Brain Tumor Detection & Segmentation," in *25th International Conference on Pattern Recognition (ICPR)* (Milan: ICPR), 5943–5950.

Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., et al. (2021). The 2021 WHO segmentation of tumors of the central nervous system: a summary, *Neurooncology* 23, 1231–1251. doi: 10.1093/neuonc/noab106

Maqsood, S., Damaševičius, R., and Maskeliunas, R. (2022). Multi-modal Brain tumor detection using deep neural network and multiclass SVM. *Medicina*. 58, 1090. doi: 10.3390/medicina58081090

Nida-Ur-Rehman, Q., Ahmed, I., Masood, G., Khan, M., and Adnan, A. (2017). Segmentation of brain tumor in multimodal MRI using histogram differencing & KNN. *Int. J. Adv. Comp. Sci. Appl.* 8, 249–256. doi: 10.14569/IJACSA.2017.080434

Oztek, M. A., Noda, S. M., Romberg, E. K., Cole, B. L., Wright, J. N., Ishak, G. E., and Perez, F. A. (2023). Changes to pediatric brain tumors in 2021 World Health Organization segmentation of tumors of the central nervous system. *Pediat. Radiol.* 53, 523–543 doi: 10.1007/s00247-022-05546-w

Phan, T. C., and ThanhHieu, L. (2024). Tumor segmentation and classification using machine learning approaches. *Int. J. Data Inform. Intellig. Comp.* 3, 1–11. doi: 10.59461/ijdiic.v3i1.89

Radiopaedia's (2023). Available online at: https://radiopaedia.org (accessed October 22, 2023).

Ramesh, S., Sasikala, S., and Paramanandham, N. (2021). Segmentation and segmentation of brain tumors using modified median noise filter and deep learning approaches. *Multimedia Tools Appl.* 80, 11789–11813. doi: 10.1007/s11042-020-10351-4

RMH Neuropathology (2013). *CNS lymphoma and demyelination*. Case study. Radiopaedia.org (accessed July 22, 2023). doi: 10.53347/rID-25592

Schubert, R. (2011). *Supratentorial ependymoma*. Case study. Radiopaedia.org (accessed July 22, 2023). doi: 10.53347/rID-14088

Shoaib, M. R., Elshamy, M. R., Taha, T. E., El-Fishawy, A. S., and Abd El-Samie, E. F. (2022). Efficient deep learning models for brain tumor detection with segmentation and data augmentation techniques. *Concurr. Comp.: Practice Exp.* 34, 21. doi: 10.1002/cpe.7031

Solanki, S., Singh, U. P., Chouhan, S. S., and Jain, S. (2023). Brain tumor detection and segmentation using intelligence techniques: an overview. *IEEE Access*. 11, 12870–12886. doi: 10.1109/ACCESS.2023.3242666

Soomro, T. A., Zheng, L., Afifi, A. J., Ali, A., Soomro, S., Yin, M., and Gao, J. (2023). Image segmentation for mr brain tumor detection using machine learning: a Review. *IEEE Rev. Biomed. Eng.* 16, 70–90. doi: 10.1109/RBME.2022.3185292

Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imag.* 9, 611–629. doi: 10.1007/s13244-018-0639-9

Frontiers in Computational Neuroscience

TYPE Original Research
PUBLISHED 28 June 2024
DOI 10.3389/fncom.2024.1425008

OPEN ACCESS

EDITED BY
Mohammad Shabaz,
Model Institute of Engineering and Technology, Jammu, India

REVIEWED BY
Zulfiqar Ali,
University of Essex, United Kingdom
Pooja Sapra,
Parul University, India

*CORRESPONDENCE
Ghadir Altuwaijri
✉ ga.altuwaijri@mu.edu.sa

†These authors have contributed equally to this work

RECEIVED 29 April 2024
ACCEPTED 03 June 2024
PUBLISHED 28 June 2024

CITATION
Arora J, Altuwaijri G, Nauman A, Tushir M, Sharma T, Gupta D and Kim SW (2024) Conditional spatial biased intuitionistic clustering technique for brain MRI image segmentation.
Front. Comput. Neurosci. 18:1425008.
doi: 10.3389/fncom.2024.1425008

Conditional spatial biased intuitionistic clustering technique for brain MRI image segmentation

Jyoti Arora[1†], Ghadir Altuwaijri[2*], Ali Nauman[3†], Meena Tushir[1], Tripti Sharma[1], Deepali Gupta[4] and Sung Won Kim[3]

[1]MSIT, New Delhi, India, [2]Department of Computer Engineering, College of Computer and Information Sciences, Majmaah University, Majmaah, Saudi Arabia, [3]School of Computer Science and Engineering, Yeungnam University, Gyeongsan, Republic of Korea, [4]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

In clinical research, it is crucial to segment the magnetic resonance (MR) brain image for studying the internal tissues of the brain. To address this challenge in a sustainable manner, a novel approach has been proposed leveraging the power of unsupervised clustering while integrating conditional spatial properties of the image into intuitionistic clustering technique for segmenting MRI images of brain scans. In the proposed technique, an Intuitionistic-based clustering approach incorporates a nuanced understanding of uncertainty inherent in the image data. The measure of uncertainty is achieved through calculation of hesitation degree. The approach introduces a conditional spatial function alongside the intuitionistic membership matrix, enabling the consideration of spatial relationships within the image. Furthermore, by calculating weighted intuitionistic membership matrix, the algorithm gains the ability to adapt its smoothing behavior based on the local context. The main advantages are enhanced robustness with homogenous segments, lower sensitivity to noise, intensity inhomogeneity and accommodation of degree of hesitation or uncertainty that may exist in the real-world datasets. A comparative analysis of synthetic and real datasets of MR brain images proves the efficiency of the suggested approach over different algorithms. The paper investigates how the suggested research methodology performs in medical industry under different circumstances including both qualitative and quantitative parameters such as segmentation accuracy, similarity index, true positive ratio, false positive ratio. The experimental outcomes demonstrate that the suggested algorithm outperforms in retaining image details and achieving segmentation accuracy.

KEYWORDS

fuzzy C-means, intuitionistic method, conditional spatial fuzzy C-means, MRI images, segmentation

1 Introduction

One of the core issues in clinical research methods is to segment the MRI image of human brain. The segmented image helps to detect different diseases related to the brain. Due to the intricate structure of MRI brain images and use of the inherent imaging mechanism includes the presence of noise, delineation of the image boundaries and many other challenges in the segmentation of these MRI images. In literature, several image segmentation methods can be categorized as thresholding (Suzuki and Toriwaki, 1991), region growing, level set methods (Li et al., 2011), model-based methods (Blahova et al., 2023) and unsupervised clustering.

Clustering, a primary technique in unsupervised learning, involves grouping a set of patterns into clusters, which can take the form of hard or soft clustering. Soft clustering is preferred over hard clustering due to its ability to assign each pixel varying degrees of membership across all clusters (Li et al., 2022). Fuzzy C-means (FCM), initially given by Dunn (1973) and refined by Bezdek (1981), is a prominent algorithm. The FCM method proves less effective in handling noisy images, primarily because of its vulnerability to noise. The absence of spatial membership matrix in FCM leads to unreliable out-comes in its results.

To address the noise issue, several improved versions of FCM have been proposed (Chen and Zhang, 2004; Wang et al., 2004; Krindis and Chatzis, 2010), which use the image's local spatial and grayscale information and prove to give better results in segmentation of MRI images. The FCM was altered by Ahmed et al. (2002) adding neighborhood information in the membership matrix. The refinement of similarity metrics incorporates information from all pixels closer to the cluster center, provided they lie within the spatial window and homogeneous region. However, this algorithm exhibits sensitivity to randomly defined initial cluster centers and an associated increase in computational complexity. Pedrycz (1996) proposed modified FCM algorithm by integrating an auxiliary variable and data attributes into the clustering process. By taking into account the domain in a feature space and the values deduced from a particular conditional variable, this method uncovers unique patterns within a set of patterns.

A conditional or auxiliary factor-guided conditional spatial fuzzy C-means (csFCM) methodology was given by Adhikari et al. (2015). This method introduces local spatial interactions among neighboring pixels through a fuzzy weighted membership function. Its advantages include defining more homogenous segments compared to other methods, robust to noise and the elimination of spurious blobs. But the Fuzzy clustering technique does not incorporate measure of the uncertainty degree that are inherent in the image datasets.

A rapid generalized FCM for image segmentation given by Cai et al. (2007), wherein the similarity metric integrates spatial and gray-level details to generate an image with a sum of weights that operates non-linearly. In a similar way, Yang and Zhang (2011) presented a novel penalized FCM, where the penalty term functions as a reconfigure within the algorithm, drawing inspiration from neighborhood maximization. In the literature, number of Fuzzy based segmentation algorithms are proposed for understanding of the anatomical and the functional aspects of the MRI brain images. The segmentation of these MRI images provides a crucial theoretical basis for the analysis and treatment of various brain ailments (Ren et al., 2019). Conventional clustering algorithms such as FCM failed to give the accurate results (Ahmed et al., 2002). To over-come these issues, Yang et al. introduced kernel-based clustering approach embedded with spatial information to violate the effect of noise for the task of segmentation of images (Yang and Tsai, 2008). Chaira (2011) introduced the Intuitionistic based fuzzy approach, which integrates entropy function along with intuitionistic theory for the segmentation of medical images. Notably, IFCM (Intuitionistic Fuzzy C-means) exhibits reduced sensitivity to outliers compared to fuzzy clustering methods. Integrated approaches of clustering were proposed to improve the accuracy of MRI image segmentation. Caldairou et al. (2011) have proposed integrated approach of FCM with the non-local information related to image with the aim of image restoration. Dubey and Mushrif (2016) have investigated different approaches of unsupervised clustering techniques for this purpose. Singh et al. (2024) proposed kernel based FCM clustering approach with bias correction for segmenting of MRI brain images. In this approach image is pre-processed using LZM based filtering and further segmented using kernelized approach of clustering.

The problem with different clustering approaches is the overhead of defining different parameters which is not an easy task. Since medical images are not linearly separable, these clustering techniques are not able to achieve high segmentation accuracy. The majority of these techniques performs severely due to the imperfection of the devices through which image is acquired, poor magnetic field, and other image artifacts.

Besides being susceptible to noise, another significant challenge in MRI image segmentation involves addressing the ambiguity inherent in pixel values. To handle this type of problem, intuitionistic-based clustering algorithms are used, which consist of intuitionistic membership degree characterized by a hesitation degree. Furthermore, by integrating conditional spatial functions into the segmentation framework, the algorithm gains the ability to adapt its smoothing behavior based on the local context of the image with the intuitionistic membership matrix giving more weight to the pixels with more similarity. The intuitionistic based clustering algorithms have been proved to lead novel perspectives in computer vision and therefore, various domains of image segmentation (Vlachos and Sergiadis, 2005; Arora and Tushir, 2019; Thao et al., 2019; Chen et al., 2021) have witnessed the application of clustering algorithms grounded in Intuitionistic metrics.

In this research paper, we introduce an intuitionistic based clustering method integrated with the spatial properties of the image which is guided by the weighted conditional factor termed as Conditional Spatial Intuitionistic Fuzzy C-means (csIFCM) for segmenting MRI brain images. Furthermore, the proposed approach makes noticeable advancements by presenting more adaptable solution toward the segmentation of MRI images. Accurate segmentation of medical image plays a pivotal role in different fields of medical applications, ranging from diagnosis of disease to planning of treatment. Accurate MRI brain image segmentation facilitates in-depth examination of anatomical features and pathological anomalies in clinical research. Therefore, highlighting the wide practical utility of the proposed approach is crucial for facilitating its adoption and integration into existing medical workflows. The primary contributions of this research are outlined below:

- The images are pre-processed using the process of normalization in order to bring the values of the pixels into more conventional form (Internet Brain Segmentation Repository, 2024).
- The proposed approach integrates spatial properties of the image guided by the weighted conditional factor into the membership matrix of the intuitionistic approach of the clustering.
- This algorithm demonstrates its effectiveness by producing robust results with better segmentation accuracy even when faced with challenges such as noise and intensity inhomogeneity.
- The proposed model's performance was evaluated using a range of standard metrics and the results obtained confirms the outstanding performance over the existing techniques.

The subsequent sections are structured as follows: Section 2 delves into the existing literature concerning the FCM algorithm, csFCM, and IFCM algorithms. Section 3 provides the detailed methodology about the proposed work. The experimental details and the application of the proposed algorithm to diverse synthetic and real MRI images are delineated in section 4. Lastly, Section 5 remarks the concluding summary of the research (Anand et al., 2023; Uppal et al., 2023).

2 Materials and methods

2.1 Fuzzy C-means

The FCM is a widely used clustering approach that aims to divide the data into groups such that each data point has a particular degree of membership μ_{ik}, that binds the datapoint with a particular cluster with certain percentage. The membership μ_{ik}, that each data point has with the cluster center is calculated by measuring the ratio of the distance between them and others.

The process of FCM is defined by the given equation of the objective function as in Eq. (1).

$$J_{FCM} = \sum_{i=1}^{C}\sum_{k=1}^{N}\mu_{ik}^{m} d\left(x_k, v_i\right)^2 \tag{1}$$

Here μ_{ik} represents membership function, $d\left(x_k, v_i\right)$ represents the distance metrics between every point of the data x_k, center of cluster v_i and the variable m∈[1,∞] determines the amount of fuzziness. In order to satisfy the imposed probability constraint and minimize the objective function of FCM, the degree of membership and center are determined. Eq. (2) defines the imposed constraint.

$$\sum_{i=1}^{c}\mu_{ik} = 1, \mu_{ik} \in [0,1], 0 \le \sum_{k=1}^{N}\mu_{ik} \le N \tag{2}$$

where number of clusters are defined manually by the variable c.

Here Eqs (3) and (4) defines the membership matrix and cluster center as:

$$\mu_{ik} = \frac{1}{\sum_{i=1}^{c}\left(\frac{d\left(x_k, v_i\right)}{d\left(x_k, v_j\right)}\right)^{\frac{2}{m-1}}} \tag{3}$$

$$v_i = \frac{\sum_{k=1}^{N}\mu_{ik}^{m} x_k}{\sum_{k=1}^{N}\mu_{ik}^{m}} \tag{4}$$

The primary problem with the FCM is that it is susceptible to noise and other visual artifacts, and its objective function lacks any spatial information.

Generally, the flow of the unsupervised clustering algorithm can be given in following steps.

2.2 Conditional spatial fuzzy C-means

In order to remove the drawback of FCM, improved version of spatial algorithm known as csFCM was given by Pedrycz (1996). csFCM includes the conditioning aspect of the clustering mechanism as the spatial properties of an image. This conditioning aspect allows smoothening of the pixel within its specified vicinity. In csFCM, firstly μ_{ik} and v_i are calculated as given in Eqs (3) and (4). Furthermore, spatial membership function μ_{ik} is calculated using conditional variable h_{ik} as in Eq. (5).

$$u_{ik} = \frac{h_{ik}}{d\left(x_k, v_i\right)^{\frac{2}{m-1}} / \sum_{i=1}^{C} d\left(x_k, v_i\right)^{\frac{2}{m-1}}} \tag{5}$$

Here h_{ik}, represents an auxiliary conditional variable that determines the extent to which a pixel is associated with the specific cluster by taking into account its spatial neighborhood and is calculated as Eq. (6).

In contrast to the FCM algorithm, the csFCM algorithm introduces a conditional element into the clustering process. The algorithm factors in conditioning variables, denoted as h1, h2, …, h_n for all pixels x1, x2, …, x_k, respectively.

$$h_{ik} = \frac{\sum_{j \in N\left(x_k\right)}\mu_{ik}}{R} \tag{6}$$

$N\left(x_k\right)$ is a fixed size square window having the pixel x_k as its center and R denotes the cardinality of pixels in the neighborhood. Further the weighted membership z_{ik} of csFCM and new cluster center t_i is calculated as in Eqs (7, 8).

$$z_{ik} = \frac{\left(\mu_{ik}\right)^p \left(u_{ik}\right)^q}{\sum_{i=1}^{C}\left(\mu_{ik}\right)^p \left(u_{ik}\right)^q} \tag{7}$$

$$t_i = \frac{\sum_{k=1}^{N} z_{ik}^{m} x_k}{\sum_{k=1}^{N} z_{ik}^{m}} \tag{8}$$

where in Eq. (7), local μ_{ik} and global u_{ik} membership values are weighted by constants p and q. These constants are used to regulate the respective importance of both the memberships for construction of the final weighted membership function and cluster center.

2.3 Intuitionistic fuzzy C-means

The theory of intuitionistic fuzzy sets (IFS) was introduced by Atanassov (1986). Unlike regular fuzzy sets (FS), IFS take into account a data point's membership and non-membership values while also taking into account the presence of a third parameter, i.e., hesitation degree. In IFS, the limitation imposed on the non-membership degree is that it is not the complement of the membership value (Kang et al., 2018).

The IFS (S_f) for dataset X can be represented as:

$$S_f = \left\{ x_{ik}, \mu_S\left(x_{ik}\right), \lambda_S\left(x_{ik}\right), \pi_S\left(x_{ik}\right), x_{ik} \in X \right\} \tag{9}$$

In the Eq. (9), the third parameter known as hesitation degree is introduced which differentiates the FS from IFS. Here, $\mu_S\left(x_{ik}\right) \to [0,1], \lambda_S\left(x_{ik}\right) \to [0,1]$ and $\pi_S\left(x_{ik}\right) \to [0,1]$ represents the membership matrix, non-membership matrix and hesitation degree matrix of the data point x_{ik} in an S_f with the following condition as in Eq. (10):

$$0 \le \mu_S\left(x_{ik}\right) + \lambda_S\left(x_{ik}\right) \le 1 \tag{10}$$

Chaira (2011) proposed novel Intuitionistic Fuzzy C-means (IFCM) approach for segmenting medical images. The IFCM introduced third parameter with the presence of membership and non-membership degree known as hesitation degree. It is calculated with the help of fuzzy membership complements using yagers complement and the sugeno complement. Dubey et al. (2016) in his research has introduced a new measure of fuzzy complement in the presence of uncertainty. The non-membership calculated using these set of fuzzy complement does not give non-membership as complement of membership. Thus, there is another factor known as hesitation degree which is $\pi_S\left(x_{ik}\right) = 1 - \mu_S\left(x_{ik}\right) - \lambda_S\left(x_{ik}\right)$ and $\pi_S\left(x_{ik}\right)$ is a measure of hesitation degree.

3 Proposed conditional spatial intuitionistic fuzzy C-means

The three drawbacks of FCM algorithm are:

- The FCM's objective function does not integrate spatial information; it treats each pixel as an individual intensity value. Image noise, which emerges during image acquisition, can lead to altered pixel intensity values, introducing both noise and intensity inhomogeneity (Zijdenbos and Dawant, 1994). Consequently, due to FCM's susceptibility to noise, noisy pixels tend to be misclassified in images.
- The relative distance between value of the pixel of the image and the cluster center determines the membership degree of the FCM. Pixels in close proximity to the centroid are attributed higher membership degrees, while those in distant clusters receive lower membership degrees (Hua et al., 2021). Consequently, the values of the membership becomes delicate in the presence of noise.
- Furthermore, FCM neglects to account for the presence of uncertainty or hesitation that might be inherent in real-world datasets (Ullah et al., 2023).

To some extent, csFCM proved to provide better clustering results in image segmentation. However, this algorithm does not consider the degree of hesitation or uncertainty that may exist in the real-world datasets, so the noisy pixels are not properly classified in its neighborhood. To overcome the problem of csFCM and other research work, we have proposed conditional spatial intuitionistic fuzzy C-means (csIFCM) for generating better segmentation results (Dhiman et al., 2022). In conventional FCM, the values of the non-membership degree typically complements the values of the membership degree. However, within the context of an intuitionistic approach, the non-membership degree is adjusted using fuzzy generators. This adaptation aims to effectively manage the inherent uncertainty (hesitation degree) in the dataset.

In the proposed algorithm, first we have normalized every pixel value of the image using the process of normalization in order to range the pixel value between 0 and 1. Further initial values of the number of the clusters (segments), random membership matrix, and random values of cluster centers are initialized. Then the value of the conditional spatial membership u_{ik}^* is calculated as mentioned in Eq. (11) using spatial membership function h_{ik} given by Eq. (12).

$$u_{ik}^* = \frac{h_{ik}}{d\left(x_k, v_i\right)^{\frac{2}{m-1}} / \sum_{i=1}^{C} d\left(x_k, v_i\right)^{\frac{2}{m-1}}} \tag{11}$$

$$h_{ik} = \frac{\sum_{j \in N\left(x_k\right)} \mu_{ik}}{R} \tag{12}$$

$N\left(x_k\right)$ is a fixed size square window having the pixel x_k as its center and R denotes the cardinality of pixels in the neighborhood. Then, a non-membership matrix is calculated using yager's fuzzy generator (Chaira, 2011) as given in Eq. (13).

$$\lambda_S = \left[1 - \left(u_{ik}^*\right)^{\alpha} \right]^{1/\alpha} \tag{13}$$

Here α is a hyperparameter that controls the degree of hesitation. Further, we calculate the hesitation degree as Eq. (14).

$$\pi_S\left(x_{ik}\right) = 1 - u_{ik}^*\left(x_{ik}\right) - \lambda_S\left(x_{ik}\right) \tag{14}$$

Subsequently, an conditional spatial intuitionistic membership matrix is computed by adding together the conditional membership and hesitation degree, as detailed in the Eq. (15).

$$\mu^*{}_S = \pi_S\left(x_{ik}\right) + u\left(x_{ik}\right) \tag{15}$$

Further the weighted intuitionistic membership matrix w_{ik} of csIFCM and joint cluster center g_{ik} is calculated as

$$w_{ik} = \frac{\left(\mu_{ik}^*\right)^p \left(u_{ik}^*\right)^q}{\sum_{i=1}^{C} \left(\mu_{ik}^*\right)^p \left(u_{ik}^*\right)^q} \tag{16}$$

$$g_{ik} = \frac{\sum_{k=1}^{N} w_{ik}^m x_k}{\sum_{k=1}^{N} w_{ik}^m} \tag{17}$$

The Eq. (16), represents the mathematical formulation of weighted membership function, where the parameters p and q are utilized to regulate the extent of intuitionistic fuzziness and the level of spatial membership. These parameters play a pivotal role in constructing the ultimate intuitionistic weighted membership function and updating the cluster centers. This w_{ik} matrix finally defines the weighted membership of the pixel with the segment of an image. In regions characterized by homogeneity, the spatial function reinforces the original membership function, leaving the clustering outcome unaltered.

However, in the case of noisy pixels, which typically do not belong to the cluster, classification becomes easier through considerations of the neighboring pixels. Consequently, inaccurately classified pixels stemming from spurious clusters or noisy regions can be effectively rectified. This attribute renders the proposed csIFCM algorithm notably robust against noise and other image artifacts, ultimately enhancing the accuracy of image segmentation.

The process is repeated till the termination criterion $(w^*_{i+1} - w^*_i \le \varepsilon)$ is met.

Proposed csIFCM Algorithm

```
Input: Image $X$ with $(x_1, x_2, x_3 \ldots\ldots\ldots x_k)$ normalized pixel points, number of
    clusters $c$.
Output: Weighted intuitionistic membership matrix $w_{ik}$, joint cluster centers $g_{ik}$
Initialize Partition Matrix $\mu_{ik}$ and cluster centers $v_i$, initial parameters
Calculate non-membership matrix using yagers compliments, hesitation degree
    and intuitionistic membership matrix.
for $i = 1 to n$
    for $k = 1 to c$
        Repeat for $j = 1,2,3\ldots$
            Update the conditional spatial membership matrix $u^*_{ik}$ and
            intuitionistic membership matrix $\mu^*_{ik}$.
            Update new weighted membership $w_{ik}$ matrix and joint cluster
            centers $g_{ik}$ using Eqs (16) and (17)
        Until $(w^*_{i+1} - w^*_i \le \varepsilon)$
    end
end
Return $w_{ik}$ and $g_{ik}$.
```

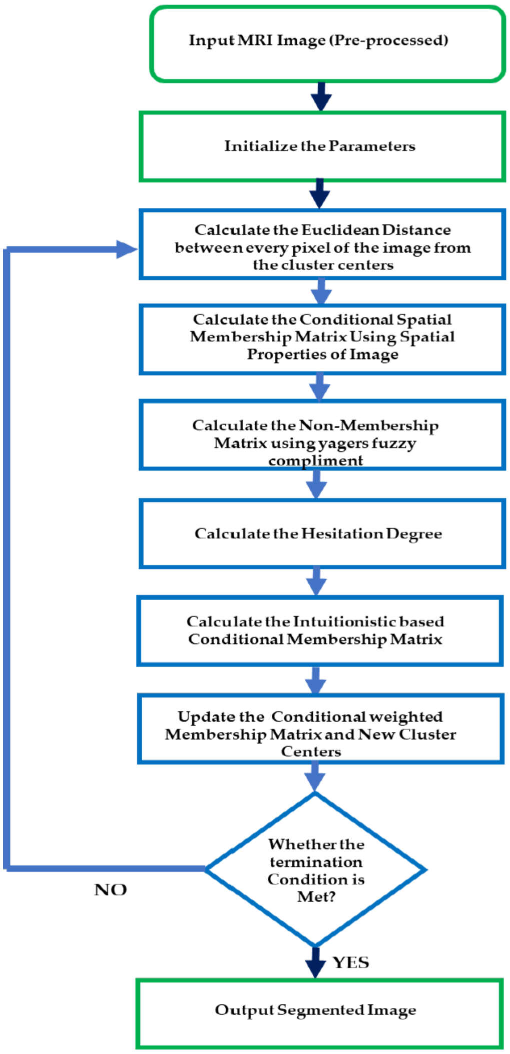


FIGURE 1
Proposed algorithm csIFCM for segmentation of MRI images.

In the proposed csIFCM algorithm, steps are executed as per the above given algorithm. Figure 1 shows the flow of the methodology.

4 Experiment results and discussion

Within this section, we undertake an evaluation of our proposed approach's performance. This evaluation encompasses both synthetic images derived from the phantom dataset and actual MRI images of the human brain sourced from The Brain Atlas of Harvard Medical School, Harvard University, and the Internet Brain Segmentation Repository (Internet Brain Atlas, 2024; Internet Brain Segmentation Repository, 2024). This evaluation entails a comprehensive assessment encompassing qualitative and quantitative analyses. We evaluate the efficiency of the proposed csIFCM in comparison to the other algorithms, including FCM, IFCM, and csFCM. Notably, all clustering algorithms were implemented utilizing MATLAB (R2015a). Additionally, trial and error are used to determine the ideal values for the weighted membership function's exponents (m, p, and q). Non-membership matrix and hesitation degree is calculated using the self-chosen alpha parameter which are used to calculate the intuitionistic membership matrix used in proposed csIFCM method.

4.1 Initialization of parameters

The parameters p and q have a significant influence on the weighted membership function w and joint cluster center, thereby effecting the accuracy of csIFCM. To fix the values of p and q for csIFCM, we have calculated the segmentation accuracy of the MRI images by segmenting the image at different values of p and q (Anand et al., 2023; Uppal et al., 2023). Figure 2 represents the segmentation accuracy attained on different slices (10–20) (Internet Brain Atlas, 2024) of MRI image with respect to different values of p and q. Table 1. represents the average value of segmentation accuracy on different slices of MRI images with respect to values of p and q. It is observed from Table 1 and Figure 2 that csIFCM is giving optimal results on $p = 1$ and $q = 2$. The outcomes underscore the necessity for balanced emphasis on both local and global membership values during the convolution process.

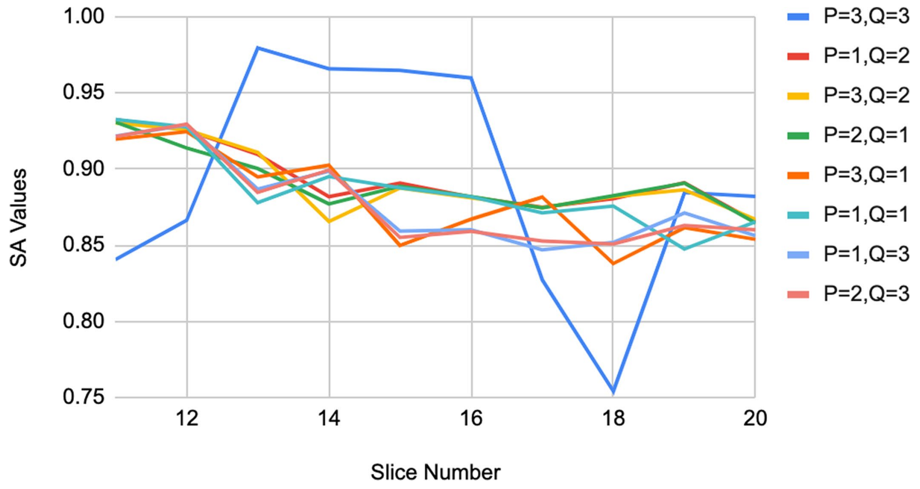


FIGURE 2
Results of segmentation accuracy on MRI images for different values of p and q.

This equilibrium is vital for generating the ultimate intuitionistic weighted membership values and joint cluster centers, wherein equal significance is attributed to both the conditional spatial membership matrix u_{ik}^{*} and the intuitionistic membership matrix μ_{ik}^{*}. Therefore, for all conducted experiments, the optimal parameter values are set as p = 1 and q = 2.

4.2 Qualitative and quantitative analysis

The qualitative results provide visual details about the different segments in the image. This evaluation demonstrates the algorithm's resilience in the face of noise. To quantitatively validate the quality of segmentation, we employed various metrics, including the false negative ratio $\left(fn_r\right)$, false positive ratio $\left(fp_r\right)$, similarity index (ρ), and overall segmentation ratio. The segment's fp_r indicates the error arising from surplus pixels, while the fn_r represents the inaccuracy occurs due to omitted pixels. The ρ refers to the pixels that align between the ground truth and the experimental outcomes. The segmentation accuracy (Zijdenbos and Dawant, 1994; Hua et al., 2021; Dhiman et al., 2022; Ullah et al., 2023) is characterized as the proportion of accurately classified pixels in relation to the entire pixel count within the ground truth image.

$$fp_r = \frac{|S_2| - |S_1 \cap S_2|}{|S_1|} \tag{18}$$

$$fn_r = \frac{|S_1| - |S_1 \cap S_2|}{|S_1|} \tag{19}$$

$$\rho = \frac{2|S_1 \cap S_2|}{|S_1| + |S_2|} \tag{20}$$

where S_1 and S_2 in Eqs (18)–(20) denote the pixels belonging to the ground truth value of the segment and experimental result obtained from the respective algorithm. Weighted membership function Parameters p and q have a significant influence on the weighted membership function w and joint cluster center g, thereby affecting the segmentation accuracy.

4.2.1 Synthetic image of phantom

A synthetic image was produced utilizing the built-in MATLAB function, phantom(). This function generates an image of a head phantom designed for assessing the numerical precision of various algorithms. The resulting grayscale intensity image comprises a prominent large ellipse symbolizing the brain, within which multiple smaller ellipses are embedded to symbolize distinct features within the brain. The quantitative outcomes of the segmentation attained by the various methods are displayed in Table 2. Table 2 indicates that the clustering method csIFCM, which is being proposed, is doing better in terms of overall segmentation accuracy.

The overall segmentation ratio should have a value of 1, with a greater value being better, for the better outcome. The suggested algorithm has a higher overall segmentation ratio (0.9631) than the other algorithms.

Figure 3 shows the overall segmentation accuracy achieved after segmentation and the proposed algorithm proved to give maximum accuracy. Figure 4 shows qualitative results of the phantom image

TABLE 1 **Segmentation accuracy of MRI images by different values of p and q for csIFCM.**

p/q	1	2	3
1	0.891	**0.8975**	0.882
2	0.894	0.8972	0.882
3	0.884	0.895	0.848

Bold values represent the results of the proposed algorithm.

TABLE 2 Quantitative measures of phantom image.

Algorithm	Segment	Similarity index	False positive ratio	True positive ratio	Overall segmentation accuracy
FCM	1	0.8305	0.0008	0.7106	0.906
	2	0.9687	0.0008	0.9324	
	3	0.9596	0.0097	0.9326	
	4	0.9598	0.0097	0.9329	
IFCM	1	0.4574	2.3729	1	0.8972
	2	0.9862	0.0131	1	
	3	0.9862	0.0131	1	
	4	0.9561	0.0101	0.9268	
csFCM	1	1	0	1	0.9187
	2	0.8601	0.0007	0.7544	
	3	0.9723	0.0018	0.9420	
	4	0.9651	0.0093	0.9423	
csIFCM	1	0.9398	0.0015	0.8877	**0.9631**
	2	0.9865	0.0013	0.9736	
	3	0.9866	0.0013	0.9737	
	4	0.9843	0.0047	0.9738	

Bold values represent the results of the proposed algorithm.

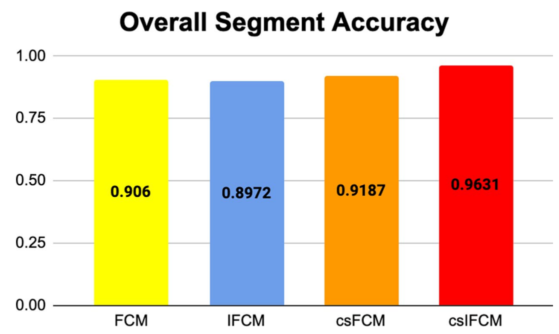


FIGURE 3
Results of overall segmentation accuracy of different algorithms for phantom image.

which has been segmented into four segments representing different regions of interest in the image.

To show the effectiveness of the algorithm against noise, we have embedded an image with 0.01% of Gaussian noise. Figures 4A,B depict the original and noisy images, respectively. The noisy image's segmented sections, produced by the FCM algorithm, are displayed in Figures 4C–F. Figures 4G–J displays the IFCM result. The results of csFCM are displayed in Figures 4K–N, and the suggested csIFCM algorithm produces results with more qualitative accuracy and greater noise resilience in Figures 4O–R.

The second experimentation involved the analysis of two real MRI brain images sourced from references (Internet Brain Atlas, 2024; Internet Brain Segmentation Repository, 2024). Specifically, the experiments were conducted on T-1 weighted axial slices ranging from slide numbers 85–115 (Internet Brain Atlas, 2024), considering noise. The primary objective was the segmentation of images into four distinct segments. It's noteworthy that these images are not provided with the ground truth value.

4.2.2 Real MR brain image

To quantify the results, a comparative analysis was performed by introducing 2% Gaussian noise to the original images and subsequently comparing the outcomes. This approach was employed as the ground truth values of the given data set was not provided. Specifically, the algorithm's performance was scrutinized by examining the variation in results when noise was added to the original images, thereby checking its effectiveness in handling noisy scenarios.

Table 3 shows the quantitative results produced by the implementation on T-1 weighted MRI axial image by FCM, IFCM, csFCM, and proposed csIFCM. Figure 5 shows the overall segmentation accuracy over all the segments and thus indicates that the proposed csIFCM outperforms all the other algorithms and validate the effectiveness of the proposed csIFCM over the other techniques under comparison.

Figure 6 shows the qualitative result of segmentation generated by FCM, IFCM, csFCM, and proposed csIFCM. A deep investigation reveals a better visualization of details in the MRI image in Figures 6O–R, where the proposed csIFCM is showing more robust results and demonstrates its superiority as compared to other algorithms. The csIFCM algorithm exhibits robustness in the presence of noise, retaining well-defined image edges and preserving a greater amount of image details.

Second, we employed actual MRI brain scans from the Internet Brain Segmentation repository (Internet Brain Atlas, 2024), which also provides manually segmented (ground truth values) data for verifying the outcomes of novel segmentation techniques. We have segmented real T1-weighted MRI brain images in 2D axial slices 10 to 20 using FCM, IFCM, csFCM, and the proposed csIFCM algorithms. Figure 7 displays the segments obtained for slice 10. Figure 7A shows

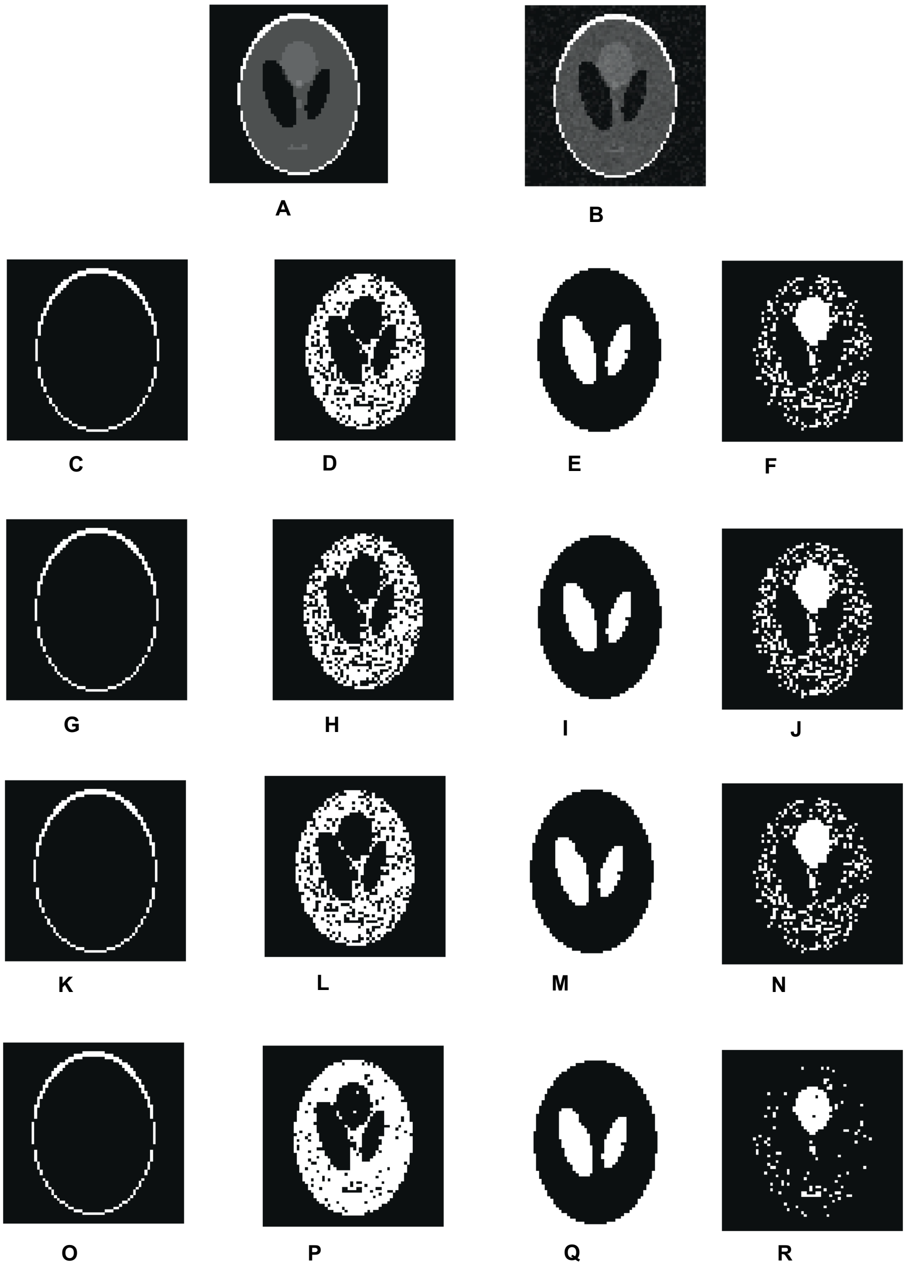


FIGURE 4
Visual results of segmentation on synthetic image of phantom. **(A)** Phantom image. **(B)** Noisy images. **(C–F)** Results of FCM. **(G–J)** Results of IFCM. **(K–N)** Results of csFCM. **(O–R)** Results of csIFCM.

the original image, which will be segmented into 4 clusters. Figures 7B–E shows the segmentation results obtained by FCM, Figures 7F–I showcases the segmentation results of IFCM, Figures 7J–M shows the segmentation results of csFCM and Figures 7N–Q shows the segmentation results of proposed csIFCM. The outcomes clearly confirm the superior performance of the proposed algorithm compared to the other algorithms under comparison. The csIFCM algorithm exhibits robustness in the presence of noise, retaining well-defined image edges and preserving a greater amount of image details.

In quantitative terms, we have computed both the true positive rate (TPR) and the false positive rate (FPR) (Adhikari et al., 2015; Arora and Tushir, 2019). This calculation facilitates the portrayal of a satisfactory balance between these two performance metrics

$$TPR = \frac{TP}{P} \tag{21}$$

$$FPR = \frac{FP}{P} \tag{22}$$

In this particular context, P symbolizes the count of positive instances, while N signifies the count of negative instances. When a

TABLE 3 Quantitative measures of T-1 weighted MRI axial image in the presence of noise.

Algorithm	Segment	Similarity index	False positive ratio	True positive ratio	Overall segmentation accuracy
FCM	1	0.9263	0.0045	0.9097	0.9747
	2	0.9825	0.0062	0.9919	
	3	0.9910	0.0078	0.9902	
	4	0.9902	0.0092	0.9900	
IFCM	1	0.9265	0.1389	0.9883	0.9753
	2	0.9837	0.0808	0.9488	
	3	0.9915	0.0643	0.9191	
	4	0.9907	0.0109	0.9897	
csFCM	1	0.9268	0.0829	0.9106	0.9757
	2	0.9845	0.1662	0.9320	
	3	0.9915	0.0682	0.9935	
	4	0.9907	0.1260	0.9819	
csIFCM	1	0.9300	0.0592	0.9206	**0.9767**
	2	0.9880	0.0627	0.9330	
	3	0.9917	0.0093	0.9927	
	4	0.9909	0.0108	0.9908	

Bold values represent the results of the proposed algorithm.

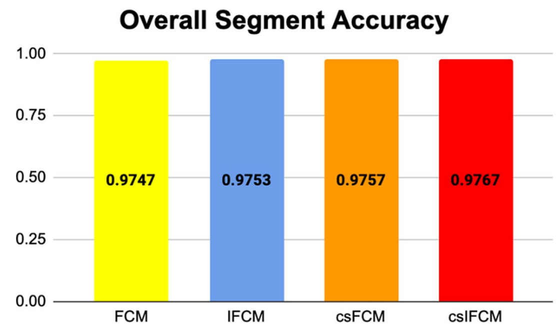


FIGURE 5
Plot showing overall segmentation accuracy for different algorithms.

prediction yields a positive result and the actual value is likewise positive, it falls under the category of true positive. Conversely, when a prediction yields a positive result but the actual value is negative, it is classified as a false positive. Table 4 shows the average value of the results obtained by the different segments for all the algorithms. The results indicate that the proposed csIFCM algorithm gives better results as compared to the FCM, IFCM, csFCM. The csIFCM gives maximum segmentation accuracy.

Figure 8 shows the plot of similarity index achieved from FCM, IFCM, csFCM and proposed csIFCM algorithm on image slices (10–20) of the T-1 weighted MR brain images. The overall segmentation accuracy and similarity index is represented for each slice. These results indicate that the proposed csIFCM technique is performing better in Figure 8A, with the highest average value of similarity index for each slice and in Figure 8B, the value of segmentation accuracy is better in each slice as compared to other algorithms and demonstrate its superiority over the FCM, IFCM, csFCM.

4.2.2.1 Time complexity analysis

To compute the similarity between each pixel requires $O(n*c*d)$ operations, where n is the number of data points, c is the number of clusters and d is the dimensionality of the data. To update the membership degrees for each data point, this typically requires $O(n*c)$ operations. In intutiotionistic approach, non-membership is calculated using membership matrix in constant time. The overall time complexity of the standard IFCM is $O(n*c*d*I)$ where I is the number of iterations required for convergence. Incorporating the conditional spatial information, includes calculation of spatial information in the time complexity of $O(n^2)$.Further, to calculate the weighted spatial membership matrix involves constant time. Therefore, the overall time complexity of incorporating conditional spatial information in csIFCM can be approximated as $O(n*c*d*I+N*K)$ where N is the number of datapoints in the spatial neighborhood of each point and K is the computational cost of processing the conditional spatial information.

4.2.3 Limitation of the conditional spatial intuitionistic fuzzy C-means

The limitation of csIFCM is its senstivity to choice of initial parameters. The initial parameters are selected randomly which may result in the increase of convergence time. The performance of csIFCM depends on how the spatial neighborhood is defined and selection of the size of neighborhood window along with the pixel under consideration which defines spatial relationship criterion. Defining a suitable spatial neighborhood enables the accurate capture of spatial structure, leading to optimal segmentation results.

Additionally, csIFCM experiences increased computational time complexity compared to IFCM due to the incorporation of spatial information and the calculation of the weighted membership matrix. The computation of spatial relationships or distances between data points adds an additional overhead to the clustering process, which can be significant, especially for large datasets or complex spatial structures.

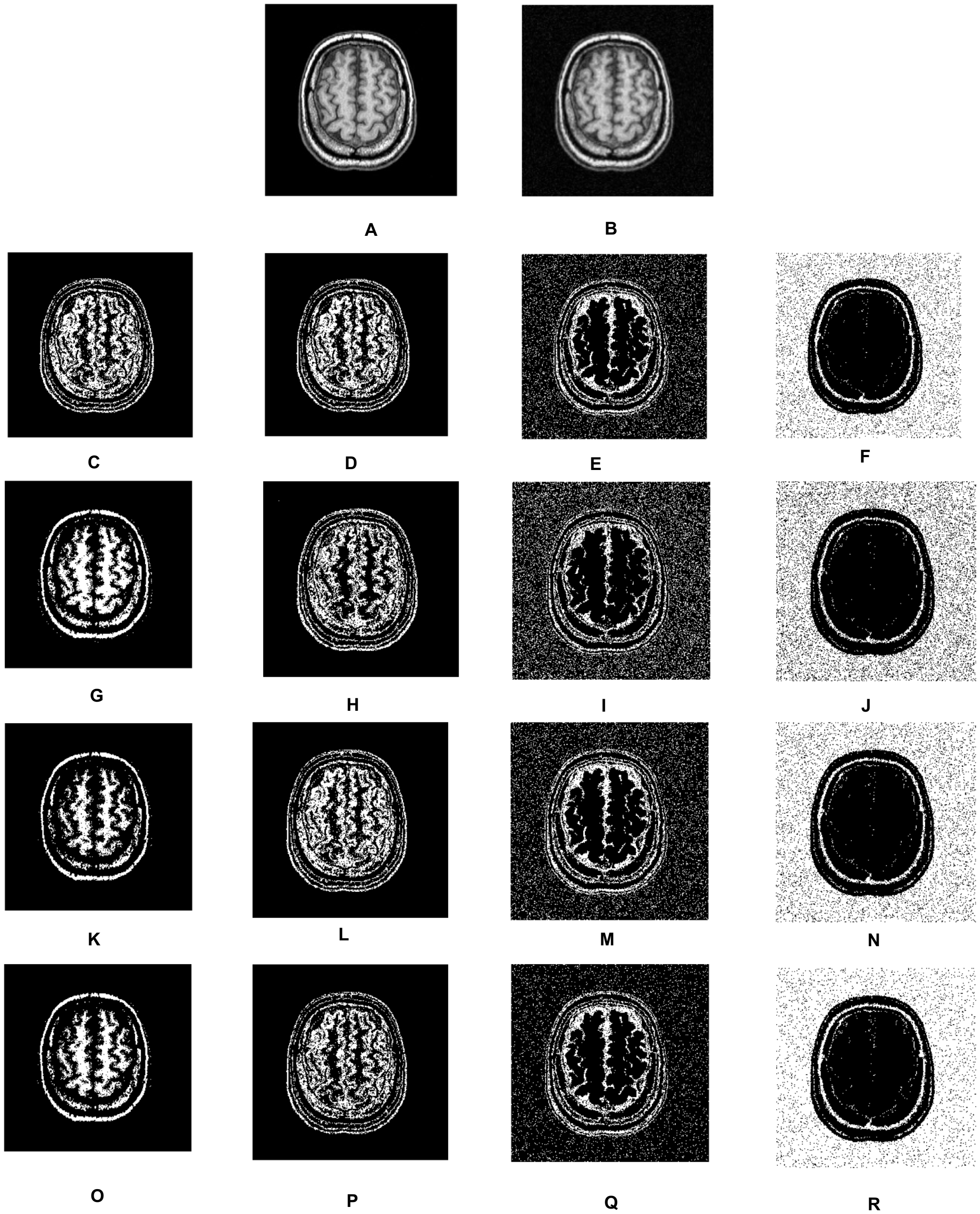


FIGURE 6
Visual results of real T-1 MR axial slice (10) scan. **(A)** Real image. **(B)** Noisy images. **(C–F)** Segmentation results of FCM. **(G–J)** Segmentation results of IFCM. **(K–N)** Segmentation results of csFCM. **(O–R)** Segmentation results of csIFCM.

5 Conclusion and future scope

This work introduces a novel algorithm named Conditional Spatial Intuitionistic Fuzzy C-means (csIFCM) for the segmentation of MRI images. By incoporating both local gray-level and spatial information through the introduction of a conditional spatial variable, csIFCM addresses the limitations of existing methods, particularly in scenarios involving noise and intensity inhomogeneity. Our experiments encompass synthetic phantom images, as well as real and simulated MRI brain images. We can explore kernel metrics that will help to segment non-linear data with higher accuracy.

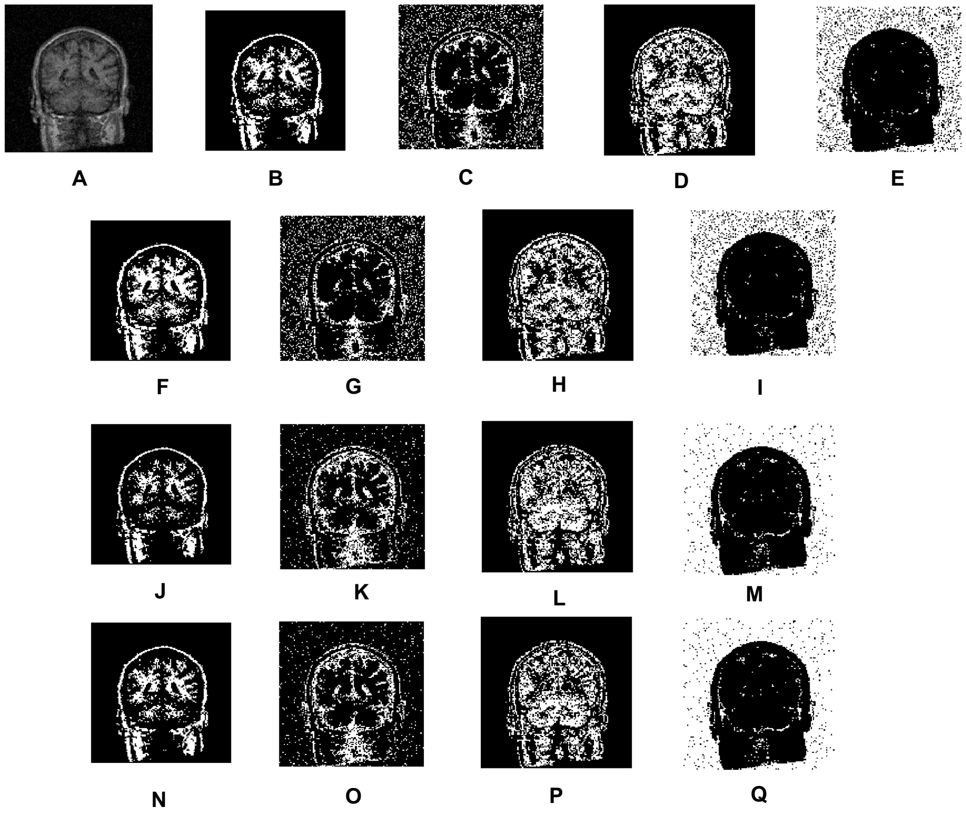


FIGURE 7
Visual results of real MRI scan. **(A)** Real images. **(B–E)** Segmentation results of FCM. **(F–I)** Segmentation results of IFCM. **(J–M)** Segmentation results of csFCM. **(N–Q)** Segmentation results of csIFCM.

TABLE 4 **Quantitative measures of T-1 weighted MR axial image of slices 10–20.**

Algorithm	Similarity index	True positive ratio	False positive ratio	Segmentation accuracy
FCM	0.8866	0.8655	0.120	0.8713
IFCM	0.8883	0.8778	0.114	0.8753
csFCM	0.8977	0.8919	0.093	0.8808
csIFCM	**0.9100**	**0.9002**	**0.083**	**0.8975**

Bold values represent the results of the proposed algorithm.

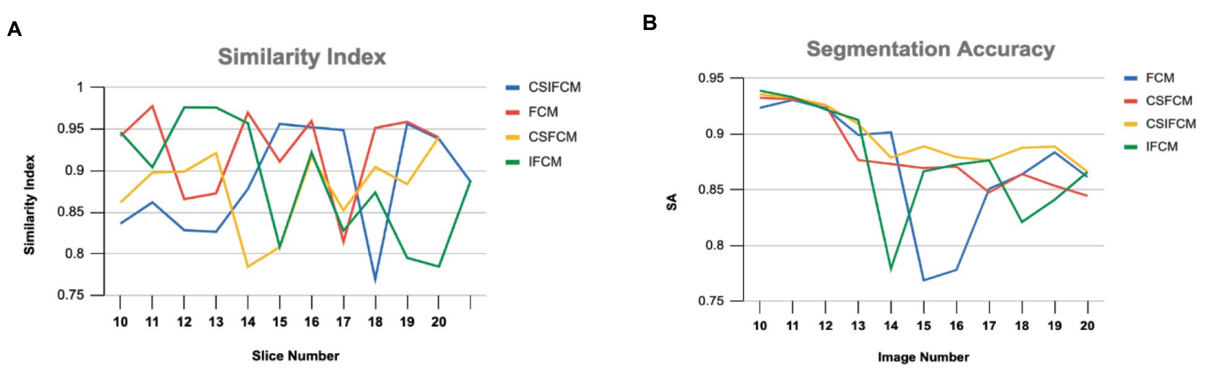


FIGURE 8
(A) Similarity Index of slices (10–20) of T-1 weighted MR axial image of FCM, IFCM, csFCM, and csIFCM. **(B)** Segmentation Accuracy of slices (10–20) of T-1 weighted MR axial image of FCM, IFCM, csFCM, and csIFCM.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: http://www.med.harvard.edu/aanlib/cases/caseNA/pb9.htm; http://www.cma.mgh.harvard.edu/ibsr/.

Ethics statement

Written informed consent was not required from the individual(s) for the publication of any potentially identifiable images or data included in this article as the dataset is taken from a public repository.

Author contributions

JA: Conceptualization, Writing – original draft. GA: Writing – review & editing, Software, Project administration, Funding acquisition, Supervision, Methodology. AN: Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization. MT: Writing – review & editing, Project administration, Formal analysis. TS: Writing – review & editing, Validation, Resources, Project administration. DG: Writing – review & editing, Supervision, Methodology, Investigation. SK: Writing – review & editing, Validation, Project administration, Formal analysis.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors would like to thank the Deanship of Scientific Research at Majmaah University for supporting this work under Project Number: R-2024-1149.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adhikari, S. K., Singh, J. K., Basu, D. K., and Nasipuri, M. (2015). Conditional spatial fuzzy C-means clustering algorithm for segmentation of MRI images. *App. Soft Compt.* 34, 758–769. doi: 10.1016/j.asoc.2015.05.038

Ahmed, M., Yamany, S., Mohamed, N., Farag, A., and Moriarty, T. A. (2002). Modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Trans. Med. Imag.* 21, 193–199. doi: 10.1109/42.996338

Anand, V., Gupta, S., Gupta, D., Gulzar, Y., Xin, Q., Juneja, S., et al. (2023). Weighted average ensemble deep learning model for stratification of brain tumor in MRI images. *Diagnostics* 13:1320. doi: 10.3390/diagnostics13071320

Arora, J., and Tushir, M. (2019). A new semi-supervised intuitionistic fuzzy C-means clustering. *EAI Endorsed Scal. Inf. Syst.* 7:e1. doi: 10.4108/eai.13-7-2018.159622

Atanassov, K. (1986). Intuitionistic fuzzy sets. *Fuzzy Sets Sys.* 20, 87–96. doi: 10.1016/S0165-0114(86)80034-3

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York, NY: Plenum.

Blahova, L., Horecny, J., and Kostolny, J., Segmentation of MRI images using clustering algorithms. 2023 International Conference on Information and Digital Technologies (IDT), Zilina, (2023), 169–178.

Cai, W., Chen, S., and Zhang, D. (2007). Fast and robust fuzzy C-means clustering algorithm incorporating local information for image segmentation. *Pattern Recogn.* 40, 825–838. doi: 10.1016/j.patcog.2006.07.011

Caldairou, B., Passat, N., Habas, P. A., Studholme, C., and Rousseau, F. (2011). A non-local fuzzy segmentation method: application to brain MRI. *Patt. Recgn.* 44, 1916–1927. doi: 10.1016/j.patcog.2010.06.006

Chaira, T. (2011). A novel intuitionistic fuzzy c means clustering algorithm and its application to medical images. *Appl. Soft Compt.* 11, 1711–1717. doi: 10.1016/j.asoc.2010.05.005

Chen, H., Xie, Z., Huang, Y., and Gai, D. (2021). Intuitionistic fuzzy C-means algorithm based on membership information transfer-ring and similarity measurement. *Sensors* 21:696. doi: 10.3390/s21030696

Chen, S., and Zhang, D. (2004). Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. *IEEE Trans. Syst. Man Cybern.* 34, 1907–1916. doi: 10.1109/TSMCB.2004.831165

Dhiman, G., Juneja, S., Viriyasitavat, W., Mohafez, H., Hadizadeh, M., Islam, M. A., et al. (2022). A novel machine-learning-based hybrid CNN model for tumor identification in medical image processing. *Sustain. For.* 14:1447. doi: 10.3390/su14031447

Dubey, Y. K., and Mushrif, M. M. (2016). FCM clustering algorithms for segmentation of brain MR images. *Adv. Fuzzy Syst.* 2016, 1–14. doi: 10.1155/2016/3406406

Dubey, Y. K., Mushrif, M. M., and Mitra, K. (2016). Segmentation of brain MR images using rough set based intuitionistic fuzzy clustering. *Biocybernet. Biomed. Eng.* 36, 413–426. doi: 10.1016/j.bbe.2016.01.001

Dunn, J. C. (1973). A fuzzy relative or the ISODATA process and its use in detecting compact and well-separated clusters. *J. Cyben.* 3, 32–57. doi: 10.1080/01969727308546046

Hua, L., Gu, Y., Gu, X., Xue, J., and Ni, T. (2021). A novel brain MRI image segmentation method using an improved multi-view fuzzy c-means clustering algorithm. *Front. Neur.* 15:662674. doi: 10.3389/fnins.2021.662674

Internet Brain Atlas. (2024). Available at: http://www.med.harvard.edu/aanlib/cases/caseNA/pb9.htm.

Internet Brain Segmentation Repository. (2024). Available at: http://www.cma.mgh.harvard.edu/ibsr/.

Kang, Y., Wu, S., Cao, D., and Weng, W. (2018). New hesitation-based distance and similarity measures on intuitionistic fuzzy sets and their applications. *Int. J. Syst. Sci.* 49, 783–799. doi: 10.1080/00207721.2018.1424965

Krindis, S., and Chatzis, V. (2010). A robust fuzzy local information C- means clustering algorithm. *IEEE Trans. Img. Proc.* 19, 1328–1337. doi: 10.1109/TIP.2010.2040763

Li, C., Huang, R., Ding, Z., Gatenby, J. C., Metaxas, D. N., and Gore, J. C. (2011). A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. *IEEE Trans. Image Process.* 20, 2007–2016. doi: 10.1109/TIP.2011.2146190

Li, M., Zhou, J., Wang, D., Peng, P., and Yu, Y. (2022). Application of clustering-based analysis in MRI brain tissue segmentation. *Compt. Math Methods Med.* 2022, 1–16. doi: 10.1155/2022/7401184

Pedrycz, W. (1996). Conditional fuzzy C-means. *Pattern Recogn. Lett.* 17, 625–631. doi: 10.1016/0167-8655(96)00027-X

Ren, T., Wang, H., Feng, H., Xu, C., Liu, G., and Ding, P. (2019). Study on the improved fuzzy clustering algorithm and its application in brain image segmentation. *Appl. Soft Compt.* 81:105503. doi: 10.1016/j.asoc.2019.105503

Singh, C., Ranade, S. K., Kaur, D., and Bala, A. (2024). A kernelized-bias-corrected fuzzy C-means approach with moment domain filtering for segmenting brain magnetic resonance images. *Soft Compt.* 28, 1909–1933. doi: 10.1007/s00500-023-09379-z

Suzuki, H., and Toriwaki, J. I. (1991). Automatic segmentation of head MRI images by knowledge guided thresholding. *Comput. Med. Imaging Graph.* 15, 233–240. doi: 10.1016/0895-6111(91)90081-6

Thao, N. X., Ali, M., and Smarandache, F. (2019). An intuitionistic fuzzy clustering algorithm based on a new correlation coefficient with application in medical diagnosis. *J. Int. Fuzzy Syst.* 36, 189–198. doi: 10.3233/JIFS-181084

Ullah, F., Nadeem, M., Abrar, M., Al-Razgan, M., Alfakih, T., Amin, F., et al. (2023). Brain tumor segmentation from MRI images using handcrafted convolutional neural network. *Diagnostics* 13:2650. doi: 10.3390/diagnostics13162650

Uppal, M., Gupta, D., Juneja, S., Gadekallu, T. R., el Bayoumy, I., Hussain, J., et al. (2023). Enhancing accuracy in brain stroke detection: multi-layer perceptron with Adadelta, RMSProp and AdaMax optimizers. *Front. Bioeng. Biotechnol.* 11:1257591. doi: 10.3389/fbioe.2023.1257591

Vlachos, K., and Sergiadis, G. D., Towards intuitionistic fuzzy image processing. International Conference on Computational Intelligence for Modelling: Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06), Vienna. (2005), 2–7.

Wang, X., Wang, Y., and Wang, L. (2004). Improving fuzzy c-means clustering based on feature-weight learning. *Patt. Recog. Lett.* 25, 1123–1132. doi: 10.1016/j.patrec.2004.03.008

Yang, M.-S., and Tsai, H.-S. (2008). A Gaussian kernel-based fuzzy c-means algorithm with a spatial bias correction. *Patt. Recgn. Lett.* 29, 1713–1725. doi: 10.1016/j.patrec.2008.04.016

Yang, X., and Zhang, G. (2011). A kernel fuzzy C-means clustering-based fuzzy support vector machine algorithm for classification problems with outliers or noises. *IEEE Trans. Fuzzy Syst.* 19, 105–115. doi: 10.1109/TFUZZ.2010.2087382

Zijdenbos, P., and Dawant, B. M. (1994). Brain segmentation and white matter lesion detection in MR images. *Crit. Rev. Biomed. Eng.* 22, 401–465.

frontiers | Frontiers in Neuroinformatics

TYPE Original Research
PUBLISHED 23 August 2024
DOI 10.3389/fninf.2024.1451529

OPEN ACCESS

EDITED BY
Najib Ben Aoun,
Al Baha University, Saudi Arabia

REVIEWED BY
Mutasem Hamdan,
Samsung Research and Development Institute, United Kingdom
Mehdi Hellou,
The University of Manchester,
United Kingdom

*CORRESPONDENCE
Abdullah S. Alharthi
✉ asalharthi@kku.edu.sa

RECEIVED 19 June 2024
ACCEPTED 29 July 2024
PUBLISHED 23 August 2024

CITATION
Alharthi AS (2024) Interpretable machine learning comprehensive human gait deterioration analysis.
Front. Neuroinform. 18:1451529.
doi: 10.3389/fninf.2024.1451529

Interpretable machine learning comprehensive human gait deterioration analysis

Abdullah S. Alharthi*

Department of Electrical Engineering, College of Engineering King Khalid University, Abha, Saudi Arabia

Introduction: Gait analysis, an expanding research area, employs non-invasive sensors and machine learning techniques for a range of applications. In this study, we investigate the impact of cognitive decline conditions on gait performance, drawing connections between gait deterioration in Parkinson's Disease (PD) and healthy individuals dual tasking.

Methods: We employ Explainable Artificial Intelligence (XAI) specifically Layer-Wise Relevance Propagation (LRP), in conjunction with Convolutional Neural Networks (CNN) to interpret the intricate patterns in gait dynamics influenced by cognitive loads.

Results: We achieved classification accuracies of 98% F1 scores for PD dataset and 95.5% F1 scores for the combined PD dataset. Furthermore, we explore the significance of cognitive load in healthy gait analysis, resulting in robust classification accuracies of 90% ± 10% F1 scores for subject cognitive load verification. Our findings reveal significant alterations in gait parameters under cognitive decline conditions, highlighting the distinctive patterns associated with PD-related gait impairment and those induced by multitasking in healthy subjects. Through advanced XAI techniques (LRP), we decipher the underlying features contributing to gait changes, providing insights into specific aspects affected by cognitive decline.

Discussion: Our study establishes a novel perspective on gait analysis, demonstrating the applicability of XAI in elucidating the shared characteristics of gait disturbances in PD and dual-task scenarios in healthy individuals. The interpretability offered by XAI enhances our ability to discern subtle variations in gait patterns, contributing to a more nuanced comprehension of the factors influencing gait dynamics in PD and dual-task conditions, emphasizing the role of XAI in unraveling the intricacies of gait control.

KEYWORDS

deep convolutional neural networks (CNN), deep learning, ground reaction forces (GRF), gait, interpretable neural networks, Parkinson's disease, perturbation

1 Introduction

Gait refers to the distinctive walking pattern unique to each individual (Saleh and Hamoud, 2021). It involves a cyclic sequence of movements in both lower limbs (Jing et al., 2019), providing valuable information about individuals' physical and physiological attributes, including weight, gender, health, and age (Wang and Zhang, 2020; Sadeghzadehyadi et al., 2021).

Gait analysis holds immense importance across various domains, such as healthcare, sport, biometrics, and human–robot interaction. It serves as a rich source of information, adding to the understanding and assessment of various conditions, including neurodegenerative disorders like Parkinson's disease (PD) (Alotaibi and Mahmood, 2015; Yuqi et al., 2019; Chaabane et al., 2023).

Previous studies (Castro et al., 2017; Huang et al., 2021; Wang and Yan, 2021; Erdaş et al., 2022; Vidya and Sasikumar, 2022) have explored gait analysis in the context of PD, aiming to diagnose the condition and track disease progression (Yuan and Zhang, 2018; Zhang S. et al., 2019; Mogan et al., 2023). However, these analyses often rely on clinical evaluation and subjective surveys, resulting in semi-subjective assessments (Wu et al., 2016; Arshad et al., 2021; Khan et al., 2023). Additionally, gait alterations under cognitive load known as "dual tasks" have been investigated, revealing variations influenced by factors such as environmental conditions and emotional states (Delgado-Escaño et al., 2018; Alharthi et al., 2019; Castro et al., 2020; Slijepcvic et al., 2021).

The existing gait analysis in literature faces limitations, particularly in accurately representing the non-linearity and non-stationary of gait cycle (Whittle, 2023). Traditional methods, such as visual observation and harmonic analysis, may fall short of capturing the intricate dynamics of gait (Goodfellow et al., 2016). To address these limitations, this study incorporates explainable artificial intelligence (XAI) techniques. XAI, including layerwise relevance propagation (LRP), enhances the transparency of deep learning models, adding to the interpretation of predictions. We selected LRP over other XAI methods, such as SHAP (SHapley Additive ExPlanations) (Ribeiro et al., 2016), Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju et al., 2017), and Local Interpretable Model-agnostic Explanations (LIME) (Lundberg and Lee, 2017). As noted by Adebayo et al. (2018), not all proposed XAI methods are robust, and the validity of their explanations should be critically assessed.

In this paper, we contribute a comprehensive approach to gait analysis by leveraging sensor fusion, deep convolutional neural networks (CNN), and XAI techniques, specifically LRP. The utilization of CNNs facilitates automatic feature extraction from raw sensor data, while the incorporation of LRP enhances interpretability. This novel combination adds significant value to the fields by providing insights that inform not only gait analysis but also sensor design and data processing for improved healthcare applications.

2 Background

2.1 Related studies

Gait, the intricate walking pattern unique to each individual, has captivated humans (Wang and Zhang, 2020). Figuratively, the gait cycle, as depicted in Figure 1, encapsulates the rhythmic sequence of movement in the lower limb during walking. Early civilizations recognized the distinctiveness of gait as a personal identifier, and over time, methodologies for studying gait have evolved from rudimentary visual observation to sophisticated techniques (Yuqi et al., 2019).

In ancient times, the recognition of individuals based on their gait laid the foundation of contemporary studies (Saleh and Hamoud, 2021). Recent advancements, such as the integration of CNN, have enabled person recognition through intricate gait models (Jing et al., 2019). These efforts underscore the enduring importance of gait analysis, with applications ranging from healthcare to biometrics (Alotaibi and Mahmood, 2015).

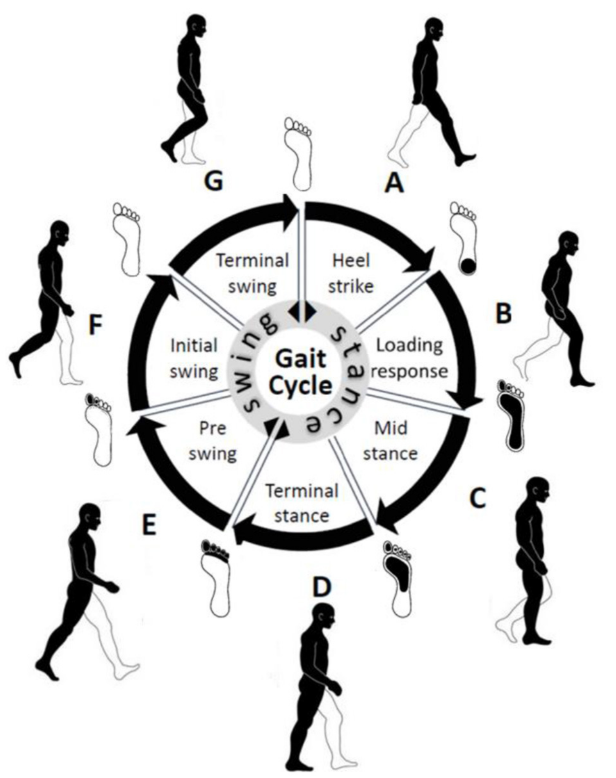


FIGURE 1
Important gait events and intervals in a normal gait cycle. In the center, the stance phase represents 60% of the gait cycle and the swing phase represents 40% of the gait cycle.

The landscape of gait analysis has witnessed a notable surge in recent literature, with cutting-edge technologies at the forefront. For instance, a fusion network incorporating long short-term memory (LSTM) and CNNs demonstrated heightened accuracy in abnormal gait recognition (Sadeghzadehyadi et al., 2021). Another study applied a CNN-LSTM network to decipher spatiotemporal patterns of gait anomalies (Wang and Zhang, 2020), highlighting a continuous evolution of gait analysis methodologies.

Gait biometrics has emerged as a focal point, with studies exploring joint CNN-based methods (Chaabane et al., 2023). Moreover, predicting the severity of neurodegenerative diseases using CNNs showcased promising outcomes (Yuqi et al., 2019). Lightweight attention-based CNN models efficiently recognized gait patterns using wearable sensors, pushing the boundaries of gait analysis capabilities (Alotaibi and Mahmood, 2015). These contemporary studies collectively underscore the growing importance of leveraging advanced technologies for accurate and nuanced gait analysis.

Recent gait recognition literature has focused on solving view- and clothing-invariant problems using advanced machine learning methods like generative adversarial networks (GANs). Zhang P. et al. (2019) designed a view transformation GAN (VT-GAN) with a generator, discriminator, and similarity preserver, achieving competitive results on the CASIA-B dataset. Babaee et al. (2019) used GANs to reconstruct complete gait energy images (GEIs) from incomplete ones, showing effectiveness on the OU-ISIR dataset. Chen et al. (2021) proposed Multi-View Gait GAN (MvGGAN) for cross-view gait recognition, demonstrating improved performance on CASIA-B and OUMVLP datasets. Recent study on wearable and floor sensors has focused on medical applications, such as analyzing the impact of muscle fatigue on gait (Balakrishnan et al., 2020), health monitoring (Muheidat and Tawalbeh, 2020), and age-related differences (Costilla-Reyes et al., 2021). Turner and Hayes (2019) proposed using an LSTM network to classify pressure sensor signals from shoes, aiming to diagnose gait abnormalities. Tran et al. (2021) developed multi-model LSTM and CNN to classify IMU spatiotemporal signals, outperforming previous results on the whuGAIT (Zou et al., 2020) and OU-ISIR (Ngo et al., 2014) datasets.

In the field of gait analysis, the integration of explainable artificial intelligence (XAI) represents a pioneering approach. XAI techniques, exemplified by methods such as layerwise relevance propagation (LRP), address the opacity challenge inherent in deep learning models (Erdaş et al., 2022). LRP has shown success in image classification (Samek et al., 2017a; Jolly et al., 2018) and gait-based subject identification (Horst et al., 2019) when combined with CNNs. Our study stands as a beacon of innovation, presenting a comprehensive approach that seamlessly integrates sensor fusion, CNN, and XAI techniques for gait analysis (Khan et al., 2023).

While existing studies have explored gait analysis through the lens of deep learning models, our distinctive contribution lies in the transparent interpretation facilitated by XAI. Building on recent advancements, we propose using LRP to enhance the interpretability of CNN predictions (Castro et al., 2020). This not only adds intrinsic value to gait analysis but also provides profound insights that extend beyond, influencing advancements in sensor design and data processing for refined healthcare applications (Alharthi et al., 2019). Our study represents a departure from conventional convolutional gait analysis approaches, introducing a paradigm shift in the synergy between gait analysis, deep learning, and explainability.

2.2 Gait parameters

Gait refers to the coordinated sequence of muscle contractions that result in walking. The brain generates commands that travel through the spinal cord to activate the lower neural center, leading to muscle contractions aided by feedback from joints and muscles. This allows for coordinated movements of the trunk and lower limbs, resulting in periodic cycles for each foot. These cycles consist of two phases: the stance phase (when the foot is in contact with the ground) and the swing phase (when the foot is not in contact with the ground). The stance phase is further divided into four intervals (A, B, C, and D), while the swing phase is divided into three intervals (E, F, and G) (Whittle, 2023) as shown in Table 1 and Figure 1.

3 Materials and methods

The categorization of gait ground reaction force (GRF) signals poses a formidable challenge, necessitating the application of sophisticated machine learning methodologies. Illustrated in Figure 2, this study delineates the framework for data acquisition and analysis. Gait data presented in Sections 3.6.1 and 3.6.2 serve as the training set for a neural network tasked with classifying these signals, and the resulting output is iteratively refined through backpropagation to pinpoint the key foot profiles crucial for classification. Detailed in subsequent sections are the experiments conducted utilizing various deep convolutional neural network (CNN) models to process and categorize spatiotemporal 3D matrices derived from raw sensor signals.

3.1 Convolutional neural networks

CNNs excel in classification tasks by abstracting high-level features from extensive datasets through convolutional operations. Mathematical representation in one-dimensional convolution operations is expressed as $C(i)$, with i denoting the index of an element in the new feature map (Goodfellow et al., 2016, ch. 9):

$$C(i) = (\omega \circ x)[i] = \sum_{d} x\left(i-d\right)\,\omega(d) \tag{1}$$

Gait is captured as a two-dimensional signal as spatial and temporal; therefore, the convolution operation in Equation 1 can be extended to two dimensions, such that the spatiotemporal input is a large set of data points, and the kernel is a set of data smaller in size than the input. Then the convolution operation slides the kernel over the input computes elementwise multiplication and adds the values in a smaller future map. With a 2-D input x and a 2-D kernel ω with (i, j), (d, k) are iterators, the mathematical representation of convolution in two dimensions can expressed as $C(i, j)$ with (i, j) is the index of an element in the new feature map (Goodfellow et al., 2016):

$$C(i,j) = (\omega \circ x)[i,j] = \sum_{d}\sum_{k} x\left(i-d,\, j-k\right)\omega(d,k) \tag{2}$$

In this study, we implement three CNN architectures for analyzing gait deterioration. The first model (Figure 3A) is a CNN designed for PD severity classification, comprising four convolutional layers, each followed by average pooling and two fully connected layers, totaling 10 stacked layers. The second CNN architecture (Figure 3B), tailored for processing GRF signals, draws inspiration from inception neural network architectures. It features two stages with parallel streams fused via concatenation layers, resulting in 18 stacked layers. The third CNN (Figure 3C) is a quadruplet network, amalgamating elements from Siamese and triplet networks. It includes convolutional layers, max-pooling, and average pooling, with separate activations, weights, and biases for each stream. This architecture aims to capture spatial and temporal gait signals simultaneously, enhancing generalization capabilities on unseen data.

3.2 Backpropagation

It is short for "backward propagation of errors"; it is an algorithm based on gradient descent. As explained by Andrew Ng (Ma et al., 2024), the method moves in reverse order from the output layer to the input layer while calculating the gradient

TABLE 1 Gait intervals.

Sequence	Gait interval	Description
A	Heel strike	Initial contact uses this term to describe the contact of the extended limb's heel with the walking surface
B	Loading response	Foot flat is a single-support interval that follows the initial double-support interval. During this phase, the body weight is transferred onto the supporting limb. The trunk is at its lowest position, the knee is flexed, and the ankle undergoes plantar flexion
C	Mid-stance	Single-support interval that occurs between opposite toe-off and heel-off. It commences from the elevation of the opposite limb until both ankles align in the coronal plane
D	Terminal stance	Heel-off begins when the supporting heel rises from the ground in preparation for the opposite swing. During this phase, the trunk is sinking from its highest point, and the knee has an extant peak near the time of heel rise, while the ankle undergoes dorsiflexion after heel rise. The swing phase consists of three intervals: pre-swing, initial swing and mid-swing, and terminal swing
E	Pre-swing	The second double-limb support interval. During this phase, opposite initial contact occurs, and the hip begins to flex, the knee flexes, and the ankle undergoes plantar flexion. The toe is in the last contact before the swing phase, completing the push-off initiated in interval D
F	Initial swing	Mid-swing interval commences with the toe-off into single-support and starting to swing. The body weight shifts to the opposite forefoot, and the knee joint undergoes maximum flexion. The hip flexes, and the limb advances in preparation for a stride
G	Terminal swing	The last interval of the gait cycle and the end of the swing phase. The interval starts at maximum knee flexion and ends with maximum extension of the swinging limb forward. The hip continues to flex while the knee extends with regard to gravity, and the ankle continues dorsiflexion to end neutral, ready for the next heel strike

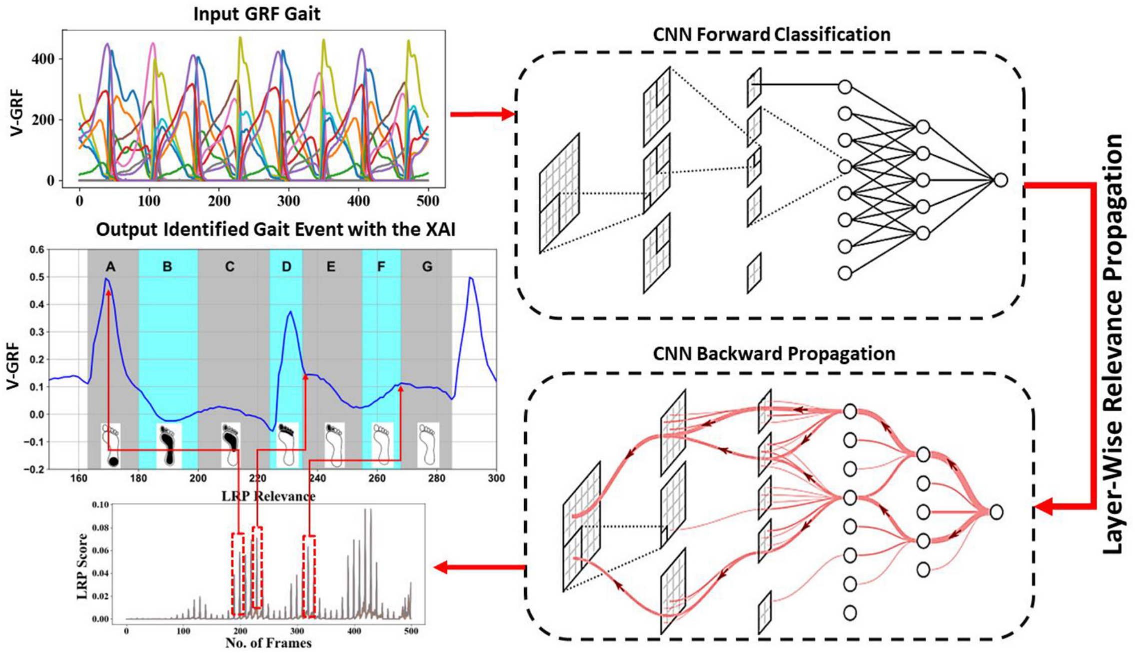


FIGURE 2
Overview of data acquisition and analysis of CNN. Gait data as input to CNN for classification; interpreting the CNN model by LRP, a deeper red color represents a higher contribution to the classification process. Relevance linked to the foot profile in the input single.

of the error function based on the network weights, the aim is to minimize $J(\theta)$ using an optimal set of parameters in θ. It is based on performing the partial derivative to minimize the cost function. The partial derivative is expressed as $\frac{\partial}{\partial\theta^{l}_{i,j}} J(\theta)$. The output layer calculates the error of the network layers L with: $\eth^{(L)} = \alpha^{(l)} - y$, such that the error of node j in layer l is denoted as $\eth^{(l)}_{j}$ and the activation of node j of layer l is denoted as $\alpha^{(l)}_{j}$ and y is the output of the output layer, then the backpropagation can be expressed for neural networks as (Ma et al., 2024):

$$\eth^{(L)} = ((\theta^{(l)})^{(T)} \eth^{(l+1)}) \circ \alpha^{(l)} \circ (1 - \alpha^{(l)}) \tag{3}$$

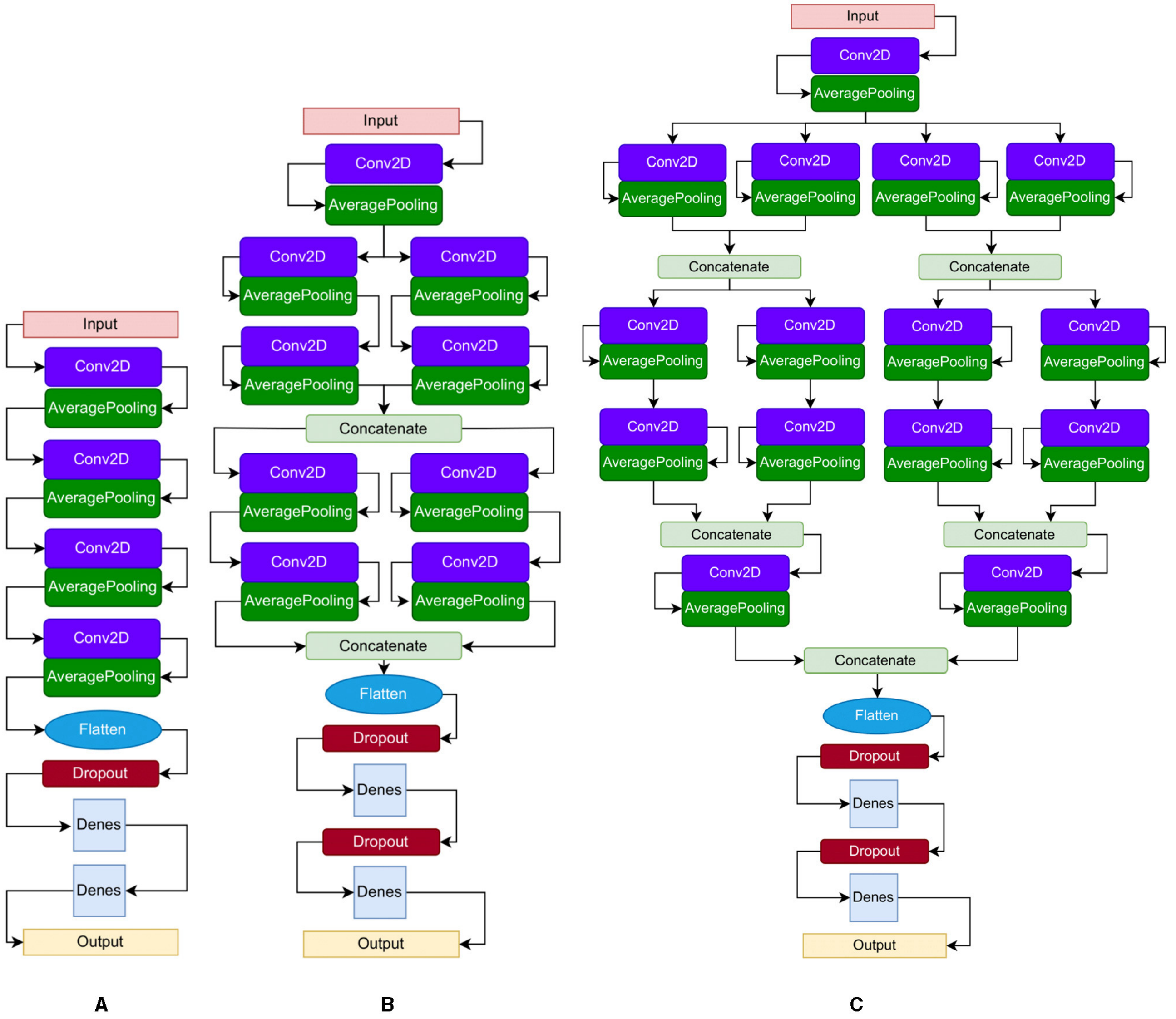


FIGURE 3
Proposed CNN architectures: **(A)** single CNN, **(B)** parallel CNN, **(C)** quadruplets CNN. The boxes: convolution layers and fully connected layers; pooling layers; concatenation layers and flattening layers; dropout layers.

Here, the $\eth^*$ values of the output layer L are calculated by multiplying the $\eth^*$ values in the next layer (in the reverse direction) with the θ matrix of layer l; hence, T denotes matrix. We then perform elementwise multiply ($^\circ$) with the g', which is the derivative of the activation function, which is evaluated with the input values given by $z^{(l)}$, where $g^{'}\left(z^{(l)}\right) = \alpha^{(l)} \circ (1 - \alpha^{(l)})$.

The partial derivatives needed for backpropagation are performed by multiplying the activation values and the error values for each training example t and m is the number of training data as Ma et al. (2024):

$$\frac{\partial}{\partial \theta_{i,j}^{l}} J(\theta) = \frac{1}{m}\left[\sum_{t=1}^{m} \alpha_j^{(t)(l)} \eth_j^{(t)(l+1)}\right]. \tag{4}$$

3.3 Evaluation measure

The confusion matrix is a common accuracy measure in gait analysis (Ruuska et al., 2018). It is a table showing correct and incorrect predictions for each class, including true positive (TP), true negative (TN), false positive (FP), and false negative (F).

In this paper, we use the confusion matrix because a number of TP, TN, FP, and FN samples are values of interest to understand the confusion in gait classes for further analysis using LRP.

From this confusion matrix table, performance measures are obtained, such as accuracy, recall, precision, and F1 using the following equations.

- Accuracy: an indicator of the ratio between the correctly predicted data to the total number of samples in the dataset, defined as follows: $\frac{TP+TN}{TP+TN+FP+FN}$.

- Recall: the proportion of positive classes identified correctly, defined as follows: $\frac{TP}{TP+FN}$.
- Precision: the fraction of positive cases correctly identified over all the positive cases predicted, defined as $\frac{TP}{TP+FP}$.
- F1-Score: the harmonic mean of Precision and Recall, defined as follows: $\frac{2{*}Precision{*}Recall}{Precision+Recall}$.

3.4 Layerwise relevance propagation

Layerwise relevance propagation (LRP) (Bach et al., 2015; Montavon et al., 2017, 2018) is a backward propagation method used to identify the most influential parts of the input vector in the model prediction of an artificial neural network (ANN). In this thesis, we measure the contribution of individual components of the input x_i (e.g., sensor signals at specific time frames) to the prediction $f_c(x)$ of a gait class c made by the CNN classifier f. The prediction is redistributed backward through the network via backpropagation until reaching the input layer. LRP generates a "heat map" over the original signal, highlighting sections with the highest contributions to the model's prediction, such as areas with the greatest variability among classes. It is important to note that a neural network comprises multiple layers of neurons, where neurons are activated as described in Montavon et al. (2018).

$$a_k=\sigma\left(\sum_j a_j\omega_{jk}+b_k\right) \tag{5}$$

Here, a_k is the neuron activation and a_j is the activation of the neuron in the previous layer in a forward direction; ω_{jk} denotes the weight received in the forward direction by neuron k from neuron j in the previous layer, and b_k is the bias. The sum is computed over all the j^{th} neurons that are connected to the k^{th} neuron. σ is a non-linear monotonically increasing activation function. These activations, weights, and biases are learned by CNN during supervisory training. During training, the output $f_c(x)$ is evaluated in a forward pass and the parameters $(\omega_{jk}+b_k)$ are updated by back-propagating using model error. For the latter, we base our computations on categorical cross-entropy (Zhang and Sabuncu, 2018).

The LRP approach decomposes the CNN output for a given prediction function of gait class c as f_c for input x_i and generates a "relevance score" R for the ith neuron received from R_j for the jth neuron in the previous layer, which is received from R_k, for the kth neuron in the lower layer, where the relevance conservation principle is satisfied as:

$$\sum_i R_{i\longleftarrow j}=\sum_j R_{j\longleftarrow k}=\sum_k R_k=f_c(x) \tag{6}$$

The LRP starts at the CNN output layer after removing the *Softmax* layer. In this process, a gait class c is selected as an input to LRP, and the other classes are eliminated. The backpropagation for unspooling for the pooling layer is computed by redirecting the signal to the neuron for which the activation was computed in the forward pass. As a generalization, consider a single output neuron i in one of the model layers, which receives a relevance score R_j from a lower-layer neuron j, or the output of the model (class c). The scores are redistributed between the connected neurons throughout the network layers, based on the contribution of the input signals x_i using the activation function (computed in the forward pass and updated by back-propagating during training) of neuron j as shown in Figure 2. The latter will hold a certain relevance score based on its activation function and pass its value to consecutive neurons in the reverse direction. Finally, the method outputs relevance scores for each sensor signal at a specific time frame. These scores represent a heat map, where the high relevance scores at specific time frames highlight the areas that contributed the most to the model classifications.

3.5 Perturbation analysis

Human gait, characterized by its inherent variability among individuals and even within a single individual, poses a significant challenge for developing reliable and robust models capable of accommodating such diversity in input data. Within the realm of gait analysis, layerwise relevance propagation (LRP) emerges as a promising methodology for interpreting the significance of input data points. However, the effectiveness of LRP in the context of gait analysis hinges on its resilience to noise and fluctuations in the input data stream.

To address this concern, a systematic exploration of the impact of random perturbation noise on LRP relevance scores is undertaken. This analysis serves a dual purpose: first, to inform the selection of the most appropriate LRP method, and second, to guide the design of a deep convolutional neural network (CNN) model capable of withstanding the inherent variability of gait patterns. The intricacies of this perturbation analysis methodology are elucidated in subsequent sections.

The iterative procedure proposed by Samek et al. (2017b), commonly referred to as the "greedy" approach, serves as the cornerstone for selecting the optimal LRP method and evaluating the relevance scores generated for gait classification. This iterative process involves progressively removing information from the spatiotemporal input signal, prioritizing regions with the highest relevance scores for perturbation using a "most relevant first" (MoRF) approach (Samek et al., 2017b). At each iteration, the model's performance is rigorously assessed by re-predicting test data with the accumulated perturbations. The selection of the preferred LRP method is informed by observing the most significant decline in accuracy during the initial iterations, indicating the criticality of the perturbed regions for accurate classification performance. Subsequent iterations demonstrate a slower decline in accuracy as less crucial regions are perturbed, thus providing insight into the relative importance of different input features.

Moreover, the evaluation of the significance of CNN model architecture entails a comprehensive analysis of the impact of perturbations on model performance. This process involves systematically removing the highest relevance scores obtained from the selected LRP method and evaluating the model's performance by re-predicting the test data for each perturbed model. Models exhibiting substantial performance deterioration after only a few perturbation steps are deemed most amenable to leveraging LRP.

This decline in performance signifies the critical role of the removed regions in facilitating accurate classification, thereby highlighting meaningful relationships between input patterns and learned classes. Conversely, regions with minimal impact on classification performance upon removal suggest lesser relevance in discerning such relationships, thus informing subsequent model refinement efforts.

3.6 Gait data

In this paper, we investigate gait deterioration due to Parkinson's disease (PD) and under dual-task conditions (walking while performing cognitive tasks as detailed in Section 3.6.2). Specifically, we compare the effects of dual-tasking and PD on gait events. The data for this study are detailed in the following section.

3.6.1 Parkinson's disease data

In this study, we utilized the open access benchmark available on PhysioNet.org (Goldberger et al., 2003) to analyze ground reaction force (GRF) data in Parkinson's disease (PD) patients. The dataset included 93 PD patients (mean age: 66.3 years; 63% men) with varying degrees of PD progression based on Hoehn and Yahr Scale staging criteria (Frenkel-Toledo et al., 2005; Yogev et al., 2005; Hausdorff et al., 2007), as outlined in Table 2, and described in detail in Table 3. Additionally, the dataset also included GRF measurements from 73 healthy controls (mean age: 66.3 years; 55% men). During the data collection process, participants were instructed to walk for ∼2 min while wearing eight sensors placed underneath each foot to measure the force [N] as a function of time. The output of the 16 sensors was recorded at a frequency of 100 frames per second. Moreover, the sum of the eight sensors of each foot was added to each subject sample along with the timestamp, resulting in a total of 19 columns. The dataset was collected by three research groups, namely the Ga group (Yogev et al., 2005), the Ju group (Hausdorff et al., 2007), and the Si group (Frenkel-Toledo et al., 2005). The sub-parts of the dataset were named after these research groups. The Ju and Si groups recorded usual healthy walking at a self-selected speed, while the Ga group included additional samples for each subject, where they performed a dual task while walking (Yogev et al., 2005). Overall, this dataset provides valuable insights into the gait patterns of PD patients and healthy individuals, which could be used to develop effective interventions for gait-related impairments in PD.

Each sample recorded in the dataset contains 19 columns of data with varying column lengths, as for some subjects' gait was recorded for a longer time (12,119 frames) than for others (<1,000 frames). In order to make the input data length consistent, the datasets were split into equal-size parts of 500 frames such that single long recordings are divided into several chunks of 500 frames. The timestamp columns were deleted as it doesn't report information about gait. The final sample size is 18 columns and 500 rows or frames as shown in Figure 4A. This choice is justified as the gait cycle is ∼1 s, and the sample captures heel strike and toe-off for both feet over five gait cycles. The input dataset is a tensor with dimensions $m \times 500 \times 18$ where $m = 2{,}698$ for the Ga group (Yogev et al., 2005), 2,198 for the Ju group (Hausdorff et al., 2007), and 1,509 Si group (Frenkel-Toledo et al., 2005). Data standardization is performed as a pre-processing step to reduce the redundancy and dependency among the data, such that the estimated activations, weights, and biases will update similarly, rather than at different rates, during the training process. The standardization involves rescaling the distribution of values with mean at zero and rescaling the standard deviation to unity.

TABLE 2 Number of subjects with the severity rating.

Severity (0) healthy	Severity (2)	Severity (2.5)	Severity (3)	Group
18	15	15	6	Ga (Balakrishnan et al., 2020)
26	12	12	4	Ju (Muheidat and Tawalbeh, 2020)
29	29	29	0	Si (Costilla-Reyes et al., 2021)

TABLE 3 Discerption of datasets subject.

Subjects	Number	Male	Female	Group
PD patients	29	20	9	Ga (Balakrishnan et al., 2020)
Healthy subjects	18	10	8	Ga (Balakrishnan et al., 2020)
PD patients	29	16	13	Ju (Muheidat and Tawalbeh, 2020)
Healthy subjects	26	12	14	Ju (Muheidat and Tawalbeh, 2020)
PD patients	35	22	13	Si(Costilla-Reyes et al., 2021)
Healthy subjects	29	18	18	Si(Costilla-Reyes et al., 2021)

$$\widehat{x_{n,s}} = \frac{x_{n,s} - \mu(x_{n,s})}{\vartheta(x_{n,s})} \tag{7}$$

Here, $\widehat{x_{n,s}}$ is PD data rescaled such that μ is the mean values and ϑ is the standard deviation. Then, the dataset is randomly split into training 60%, hold-out validation 20%, and testing 20% with a *random state* parameter with a different seed.

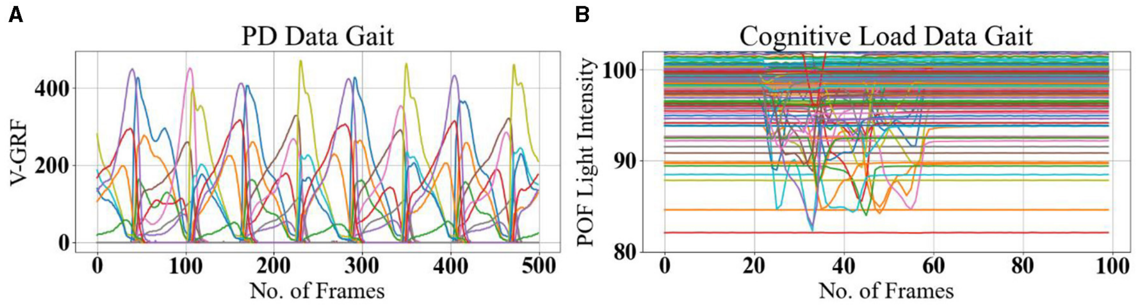


FIGURE 4
Example gait data. **(A)** PD Gait was recorded at 100 frames per second, with a sample length of 500 timeframes. The signals represent pressure sensor signals under each foot (different colors for each of the eight sensors). **(B)** Cognitive Load Gait, recorded at 20 frames per second, with a sample length of 100 timeframes. The signals represent POF sensors transmitted light intensity is affected by surface bending due to pressure under each foot (different colors for each of the 116 POF sensors).

3.6.2 Cognitive load data

The iMagiMat footstep imaging system is an innovative floor sensor head that utilizes photonic guided-path tomography technology (Ozanyan et al., 2005; Cantoral et al., 2011; Cantoral-Ceballos et al., 2015; Ozanyan, 2015). The system can capture temporal samples from strategically placed distributed POF sensors on top of a deformable underlay of a commercial retail floor carpet in an unobtrusive manner. Each sensor is made up of low-cost POF (step-index PMMA core with fluorinated polymer cladding and polyethylene jacket, total diameter 1 mm, NA = 0.46) terminated with an LED (Multicomp OVL-3328 625 nm) at one end and a photodiode (Vishay TEFD4300) at the other. The sensors are designed to allow collaborative sensor fusion and deliver spatiotemporal sampling that is adequate for discerning gait events.

The iMagiMat system covers a 1 m × 2 m area managed by 116 POF sensors arranged in three parallel plies, sandwiched between the carpet top pile and the carpet underlay. The system includes a lengthwise ply with 22 POF sensors at 0° angle to the walking direction and two independent plies, each consisting of 47 POF sensors, arranged diagonally at 60 and -60°, respectively (see Cantoral-Ceballos et al., 2015, for the iMAGiMAT system). The system is managed by electronics contained in a closed hard-shell periphery at carpet surface level and is organized into eight-channel modules, including LED Driver boards and input trans-impedance amplifier boards to receive the data and send it to a CPLD (complex programmable logic device) to reformat the data for processing by a Raspberry pi single-board computer for export via Ethernet/Wi-Fi. The operational principle of the system is based on recording the deformation caused by the variations of ground reaction force (GRF). As bending affects the POF sensors, transmitted light intensity is affected by surface bending. This captures the specifics of foot contact and generates robust data without constraints of speed or positioning anywhere on the active surface.

For this experiment, 21 physically active subjects aged 20–40 years, 17 men and four women, without gait pathology or cognitive impairment, participated. The study was carried out under the University of Manchester Research Ethics Committee (MUREC) with ethical approval number 2018-4881-6782. All participants were informed about the data recording protocol according to the ethics board's general guidelines, and written consent was obtained from each subject prior to the experiments. Each participant was asked to walk normally or while performing cognitively demanding tasks along the 2 m length direction of the iMagiMat sensor head. The captured gait data was unaffected by start and stop, as it was padded on both ends with several unrecorded gait cycles before the first footfall on the sensor. With a capture rate of 20 timeframes/s (each timeframe comprising the readings of all 116 sensors), experiments yielded 5 s long adjacent time sequences, each containing 100 frames. The recorded gait spatiotemporal signals were able to capture ∼4–5 uninterrupted footsteps at each pass.

A dual-task gait test detects mild cognitive impairment (Wang et al., 2023); therefore, five manners of walking were defined as normal gait plus four different dual tasks, and experiments were recorded for each subject, with 10 gait trials for each manner of walking in a single assessment session. Thus, the total number of samples is $10 \times 5 = 50$ per-subject. The five manners of walking are defined in Table 4. A set of measured data as $x_{n,s} = [x_{n,1} \& \ldots \& x_{n,116}] \in \mathbb{R}^{n \times 116}$ is harvested from the iMagiMat system, where n is the number of the data block (100 frames) and s enumerates the POF sensors, as shown in Figure 4B. A total number of 1,050 samples are recorded for 21 subjects and placed in a 3D matrix of dimensions $1,050 \times 100 \times 116$. The recorded amplitude of data varies due to the weight of each subject; therefore, data standardization is implemented as a pre-processing step, to ensure that the data are internally consistent, such that the estimated activations, weights, and biases update similarly, rather than at different rates, during the training process and testing stage. The standardization involves rescaling the distribution of values with a zero mean unity standard deviation, using Equation 7, where $\widehat{x_{n,s}}$ is gait data rescaled so that μ is the mean and ϑ is the standard deviation. Then, the dataset is randomly split into training 60%, hold-out validation 20%, and testing 20% with a *random state* parameter with a different seed.

4 Experiment and results

All algorithms for LRP computation are implemented in *Python* 3.7.3 programming language using

TABLE 4 Cognitive load experiment data.

Manner of walking	Description
M1	Normal Gait: walking at a normal self-selected speed
M2	Gait while listening to a story: audio input through headphones, followed by answering questions after gait recording is completed
M3	Gait with serial 7 subtractions: normal walking speed attempted while simultaneously performing serial 7 subtractions (counting backward in sevens from a given random 3-digit number)
M4	Gait while texting: normal walking speed attempted while simultaneously typing text on a mobile device keyboard
M5	Gait while talking walking at a normal self-selected speed while talking or answering questions

TABLE 5 PD data models F1-score for each dataset and F1-score with datasets combined.

CNN model	Ga	Ju	Si	GaUJuUSi
Single	**98%**	**98%**	**98%**	**96%**
Parallel	96%	97%	96%	**96%**
Quadruplet	97%	97%	98%	95%
LSTM	91%	93%	80%	94%
SGD	88%	84%	80%	83%
KNN	81%	90%	78%	79%
GPC	82%	85%	89%	81%

The best performance is in bold.

Keras 2.2.4, *TensorFlow* 1.14.0, *and iNNvestigate GitHub repository* (Alber et al., 2018). The codes are executed on a desktop with *Intel Core i7 6700 CPU @3.4 GHz*. The deep CNN model is applied to the datasets to test the validity of the algorithms for identifying gait signatures. The implementation and the perturbation analysis are detailed in the following section. We compare the CNN predictions to manually labeled ground truth in several experiments, including PD severity staging, individuals' identity, and the effects of cognitive load on normal gait. The models' classification performance is evaluated using confusion matrices. The performance of the LRP methods is examined in detail in the discussion subsection.

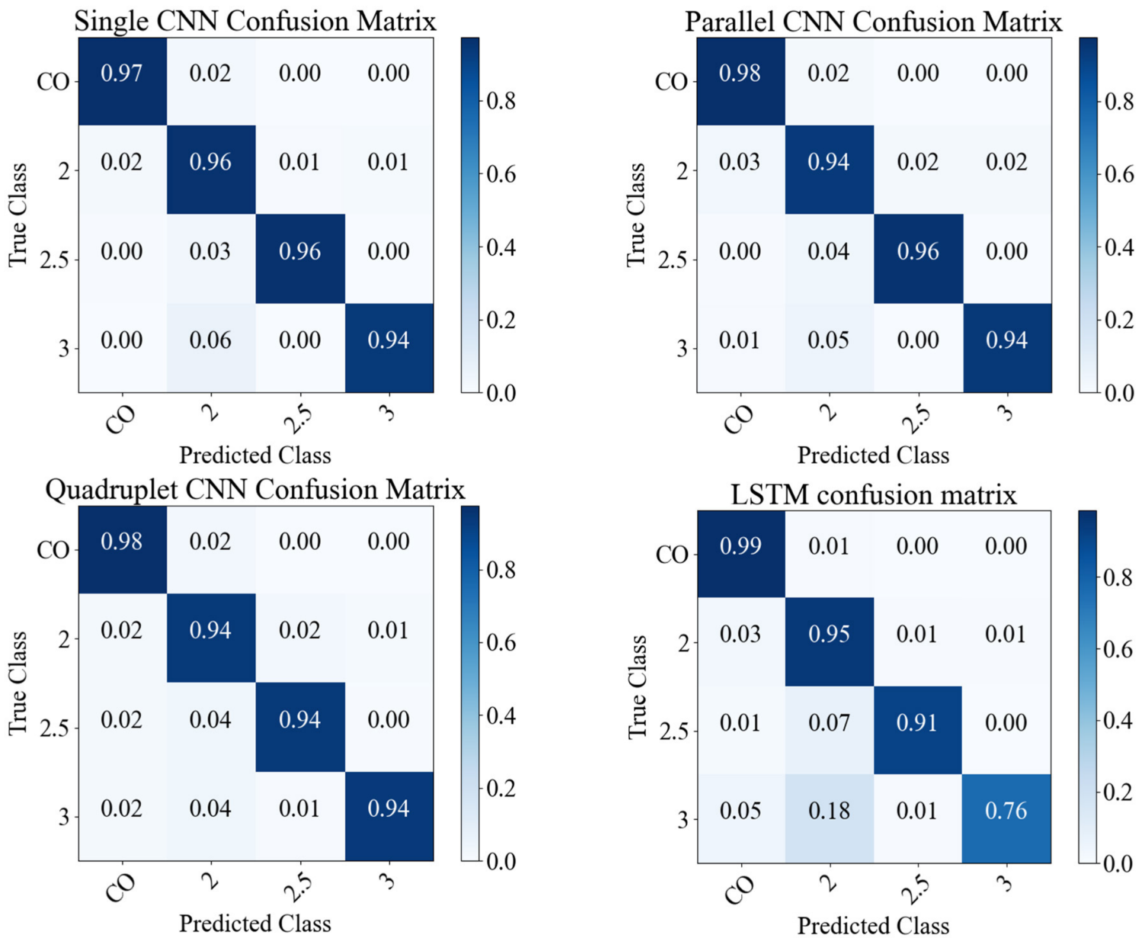
4.1 Classification experiments

We introduce a variety of algorithms and architectures, including a CNN model, LSTM, Stochastic Gradient Descent (SGD), K-Nearest Neighbors (KNN), and Gaussian Process Classifier (GPC). The SGD updates model parameters iteratively using a single or a few randomly selected data points to compute the gradient, optimizing the objective function efficiently for large datasets (Zhang, 2004). The KNN algorithm employs Euclidean distance techniques to determine the distance between data samples (Altman, 1992). The GPC leverages Gaussian processes to define a distribution over functions, making predictions by averaging over all possible functions, thus providing probabilistic classification outputs and well-calibrated uncertainties (Xiao et al., 2019). Through experimentation detailed in the following sections, we utilize CNN and LSTM methods as automatic feature extractors and classifiers. The CNN models shown in Figure 3 map the gait spatiotemporal signal $\widehat{x_{n,s}}$ to an output label y by learning an approximation function $y = f\left(\widehat{x_{n,s}}\right)$. The networks consist of an input layer, convolution layers (see Equation 2), pooling layers, fully connected layers, batch normalization layers, and an output layer with a softmax classifier. The set of stacked layers in Figure 3 utilizes Conv2D kernels (*filter size* × *number of feature maps* × *number of filters*), MaxPooling and AveragePooling layers.

To improve the model performance, a regularization method is utilized as follows: (1) Batch normalization [to normalize the activations of the previous layer at each batch, by maintaining the mean activation close to 0 and the activation standard deviation close to 1 (Ioffe and Szegedy, 2015)]. (2) The Batch normalization followed by dropout (Srivastava et al., 2014), after the last pooling layers were flattened, by transforming a matrix to one single-column vector. An Adam (adaptive moment estimation) (Kingma and Ba, 2015) is utilized to train the model. The used optimizer parameters are $\alpha = 0.002$, $\beta 1 = 0.9$, $\beta\ 2 = 0.999$, $\varepsilon = 1e - 08$. Here, α is the learning rate or the fraction of weights updated where larger values (e.g., 0.3) result in faster initial learning before the rate is updated. Smaller values (e.g., 1.0E-5) slow learning right down during training; $\beta\ 1$ and $\beta\ 2$ are the exponential decay rates for the first- and second-moment estimates, respectively; ε is a small number to avoid division by zero. The loss is computed using categorical cross-entropy in every iteration to minimize the network error (Zhang and Sabuncu, 2018). The convolutional layers weight parameters are initiated with a *Glorot* uniform (Glorot and Bengio, 2010) with zero bias. The model is trained and validated (for several experiments) using a batch size of 100 samples for each iteration; 200 epochs are found optimal to train the model based on backpropagation Equations 3 and 4. The training and validation sizes are set to be 70 and 10%, respectively, where 20% is reserved for testing the model accuracy.

4.1.1 Experiment (1) on PD severity staging

In this experiment, CNNS, LSTM, SGD, KNN, and GPC models are trained and tested on the PD dataset to classify the severity of PD into five stages: normal (*CO*), mild (2), moderate (2.5), and severe (3). Table 5 presents the models' *F*1-score for each dataset and the *F*1-score with datasets combined. Figure 5 presents the confusion matrix for the CNNs and LSTM with datasets combined. The best performance is achieved by the CNN single and parallel with F1-score 96% for the data set combined and for each data, where the LSTM performance was 79% for PD stage 3. In the statistical



Single CNN Confusion Matrix

True Class \ Predicted Class	CO	2	2.5	3
CO	0.97	0.02	0.00	0.00
2	0.02	0.96	0.01	0.01
2.5	0.00	0.03	0.96	0.00
3	0.00	0.06	0.00	0.94

Parallel CNN Confusion Matrix

True Class \ Predicted Class	CO	2	2.5	3
CO	0.98	0.02	0.00	0.00
2	0.03	0.94	0.02	0.02
2.5	0.00	0.04	0.96	0.00
3	0.01	0.05	0.00	0.94

Quadruplet CNN Confusion Matrix

True Class \ Predicted Class	CO	2	2.5	3
CO	0.98	0.02	0.00	0.00
2	0.02	0.94	0.02	0.01
2.5	0.02	0.04	0.94	0.00
3	0.02	0.04	0.01	0.94

LSTM confusion matrix

True Class \ Predicted Class	CO	2	2.5	3
CO	0.99	0.01	0.00	0.00
2	0.03	0.95	0.01	0.01
2.5	0.01	0.07	0.91	0.00
3	0.05	0.18	0.01	0.76

FIGURE 5
The predictions of models on the 1,281 sample shown as confusion matrices: single CNN, parallel CNN, quadruplet CNN, and LSTM.

analysis, the performance of SGD, KNN, and GPC models was below 90%.

4.1.2 Experiment (2) on cognitive load impact on gait

The aim of this experiment is to show that in healthy subjects the influence of cognitive load on gait varies from subject to subject and the normal gait can be predicted with higher true positive rates than predictions under cognitive load. Five types of gait signatures, normal and four cognitively demanding task patterns, are learned for 21 subjects. The performance observed for the five classes is shown in Figure 6, as the median confusion matrix based on several runs with the CNNs in Figure 3 resulted in a $F1$-score of 50%, mean performance, and standard error of 48.25 ± 1.03%. The results show that normal gait is predicted by a true positive incidence of 92% ± 1.7%, while there is notable confusion between the dual tasks performed by the 21 subjects. The different random state parameters return the same result, where the normal gait true positive prediction is higher than 90% and substantial confusion between the dual-task cases.

4.1.3 Experiment (3) cognitive load impact on gait for each subject

In this experiment, gait patterns are investigated within each subject, to show that each subject gait under cognitive load can be learned and predicted. This is achieved by training, validating, and testing the CNNs in Figure 3 to classify each subject gait pattern using the normal gait and cognitive load. Each subject data are split using a random state to cover all five classes for testing with $m = 50$ samples. The model evaluation using the $F1$-score is detailed for each subject in Table 6. Gait data are predicted with more than 85% $F1$-score for 16 subjects, and for six subjects, $F1$-scores are between 65 and 77%.

4.1.4 Experiment (4) cognitive load impact on gait for each subject

To study patterns for each of the four dual tasks ($M2 - M5$) representing variants of cognitive load, we organize the data into four groups so that binary classification performance to distinguish between gait under normal (class 0) and cognitive load (one of the classes 1, 2, 3, or 4, depending on the particular data group) conditions can be studied separately for each dual task. The CNNs in Figure 3 are trained 16 times, implementing four runs with each of the four data groups. The $F1$-scores for each run are shown in Table 7. The first run in each data group is based on training and validating the CNNs on 20 subjects and testing the model on 1 subject, to see whether we can predict the gait of one person from 20 people. In the second run, the numbers are 19 and 2, respectively; in the third—17 and 4, respectively. The last run is based on splitting the data into 70% for training, 10% for validation, and 20% for testing, using $m = 420$ samples with a random state of 200 seed parameters (as the accuracy does not change with the random state seed). As shown in Table 7, the highest classification performance is achieved in the first runs (except for the group containing class 3). This is used essentially in the implementation of LRP to analyze the gait classes for that subject in the first run as reported in further comparison with statistical classifiers.

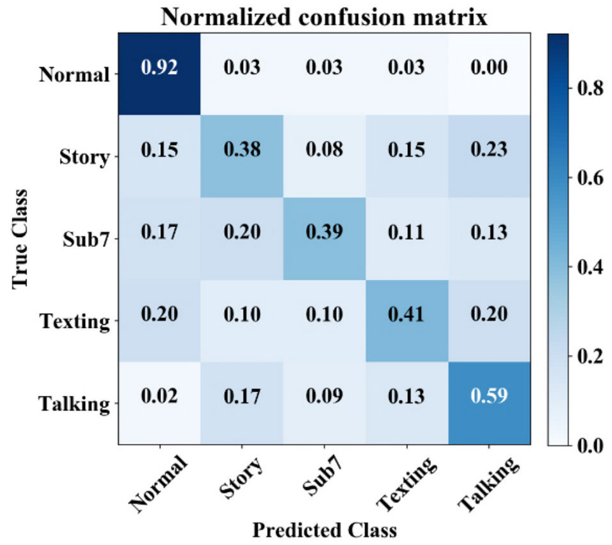


FIGURE 6
Confusion matrix for classification under cognitive load: 21 subjects, five classes. Experiment (4) Cognitive Load Impact on Gait Variations (Binary Classification).

4.2 LRP analysis and interpretation for explainability

In the following sections, we present the LRP analysis (see Equations 5 and 6) and interpretation for the best-performing model using perturbation presented in Section 3.5. Then, we present the explainability results of the investigated classification models for PD and cognitive load.

4.2.1 Model selection and XAI selection

4.2.1.1 Model selection

In this study, we conducted an in-depth analysis of the performance of various CNN models for the task of gait classification. We employed explainable AI (XAI) techniques to select the most suitable CNN model for this application. To identify the CNN model that best captures the relevant gait features, we utilized a perturbation-based approach presented in Section 3.5. Specifically, we systematically perturbed each of the three candidate CNN models by gradually replacing 7 × 7 regions within the input gait sequence with Gaussian noise and observed the impact on the classification accuracy over 100 steps. Rather than comparing the models to a baseline, we focused on the rate of decline in accuracy (with the means removed to isolate the rate of change) as a metric to identify the model with the steepest drop in performance. This approach is based on the premise that models that rely on more compact regions within the gait cycle sequence will exhibit a faster decline in accuracy when those regions are perturbed. The results, as depicted in Figure 7, show that the parallel CNN model (see Figure 7) experiences the most pronounced decrease in accuracy with perturbation, indicating that it captures the gait events that are most vulnerable to deterioration in individuals with PD. As depicted in Figure 7, after step 13, the quadruplet CNN begins capturing less relevant features, similar to the decline observed in the parallel CNN. This finding suggests that the parallel CNN model is the preferred candidate for accurate feature identification of the gait cycle events most sensitive to the effects of either PD or cognitive load. Figure 8 shows the assessment of the validity of the LRP heatmaps for subjects' identification of cognitive load.

Here, we apply the removal of the region based on both LRP sequential preset a flat (LRP-SPF) MoRF and random region removal and re-predicting gait class. As shown in Figure 8A, the model prediction strongly decays using the LRP for the

TABLE 6 Models classification accuracy for cognitive load impact on gait for each subject.

Subject number	F1-score	Subject number	F1-score	Subject number	F1-score
0	95%	7	87%	14	100%
1	65%	8	90%	15	75%
2	93%	9	90%	16	80%
3	90%	10	77%	17	100%
4	87%	11	91%	18	100%
5	91%	12	90%	19	80%
6	73%	13	100%	20	69%

TABLE 7 F1-score predictions for binary classification, normal vs. cognitive load.

Data group for classification	1 testing subject	2 testing subjects	4 testing subjects	Test with all subjects
Class 0 vs. class 1	100%	85%	81%	79%
Class 0 vs. class 2	95%	87%	58%	69%
Class 0 vs. class 3	60%	68%	63%	79%
Class 0 vs. class 4	100%	85%	74%	81%

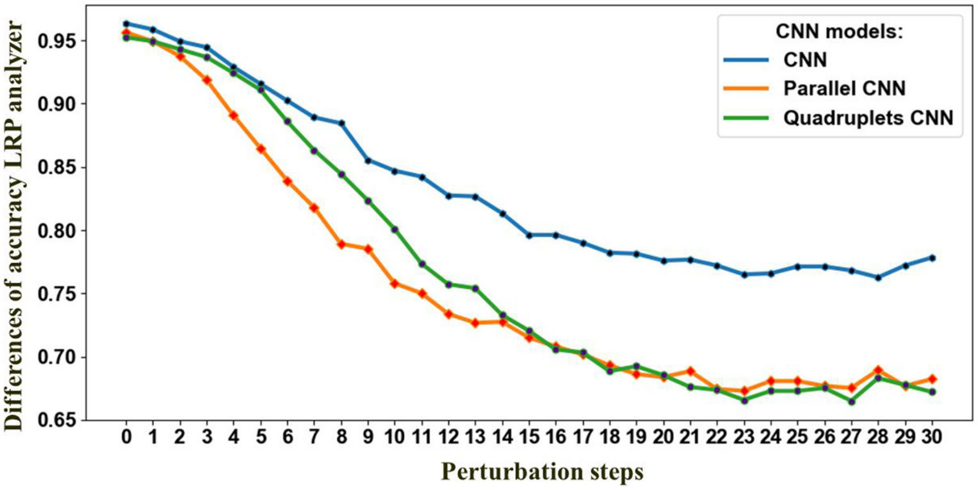


FIGURE 7
Perturbation effect on the proposed CNNs architectures. The decline in accuracy results from progressively removing information from the input data based on LRP-SPF and re-predicting, at each step, 30 steps total.

removal of information compared to the removal of random information. Figure 8B shows the model performance over 300 steps. The model reaches the lowest performance accuracy where the gait classes have to take a random prediction. Furthermore, it can be inferred from Figure 8 that the model is effective in finding the most relevant region to identify cognitive load of subjects and the LRP is consistent over the test samples.

4.2.1.2 XAI selection

To identify the most suitable backpropagation method for the three CNN models, we conducted a comprehensive evaluation of various LRP (layerwise relevance propagation) techniques. These included deep Taylor (Montavon et al., 2017), deep Taylor bounded (Kohlbrenner et al., 2019), deconvnet (deconvolution) (Zeiler and Fergus, 2014), guided backprop (guided backpropagation) (Springenberg et al., 2015), and LRP sequential preset a flat (LRP-SPF) (Kohlbrenner et al., 2019), all of which were implemented using the iNNvestigate GitHub repository. For each of the LRP methods, we assessed the CNN classification accuracy by performing a sequence of perturbation steps as described in Section 3.5 as described in **Model Selection**. To establish a baseline for comparison, we replaced regions of the input data with random Gaussian noise with one level at 0.1%, rather than using the LRP-based methods. We then subtracted the accuracy of LRP maps from the accuracy of randomly replaced regions to isolate the impact of the LRP techniques. As shown in Figure 9, the LRP curves recovered after around the 15th perturbation step as the remaining spatiotemporal regions became less relevant for the classification task. The baseline accuracy was reached around the 30th perturbation step, indicating that the remaining regions were unimportant for the classification. Importantly, the observed rate of change in accuracy was proportional to the importance of the information perturbed at each step as expected. This analysis allowed us to understand the relative significance of different regions within the input gait sequence for the classification performance of CNN models.

4.2.2 PD gait event assignment using LRP

Gait GRF data take the form of periodic sequences, which are characterized as repetitive cycles for each foot. We note that the normal gait cycle is initiated by the heel strike of one foot, followed by other gait events described in Figure 1 and Table 1, in strict order. Therefore, the LRP-generated heat map of the temporal variations in the GRF signal can reveal which events in the gait cycle are most relevant for the classifications. Consequently, gait event assignment is best performed on the data sequences in Figure 10 after spatial averaging and standardization. A representative spatially averaged sensor signal sequence is shown in Figure 10A for a healthy subject. The highlighted gray area corresponds to one gait cycle, while the plotted signal is given by

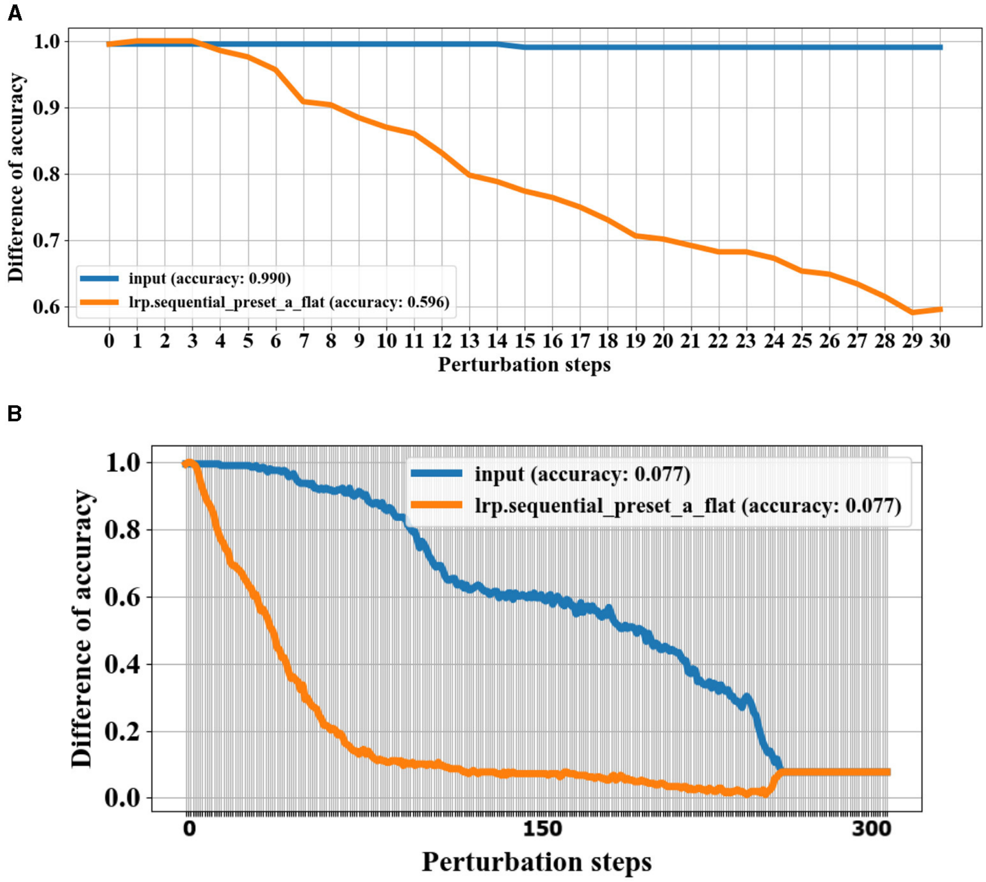


FIGURE 8
Validation of LRP heatmaps by perturbation technique for experiment 3 subject 13. Information with the highest relevance scores is progressively removed, and the test samples are re-predicted. A steeper initial decrease indicates better identification of gait events with the most weight in the classifications. **(A)** Shows the model predictions in 30 steps based on removing relevance scores using LRP sequential preset a flat (LRP-SPF) and random removal of information. **(B)** Shows the model performance after 300 steps of information removal.

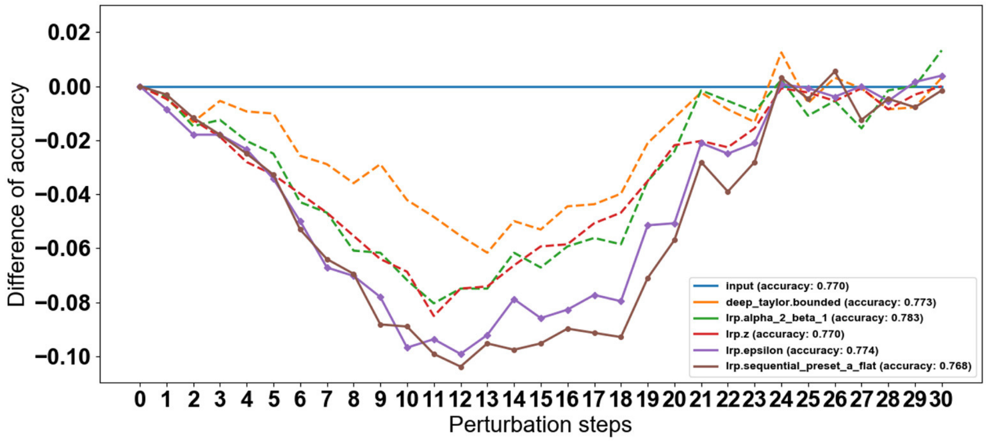


FIGURE 9
LRP method selection by perturbation steps progressively removes information with the highest relevance scores. A steeper initial decrease indicates better identification of gait events with the most weight in the classifications.

the spatial average (SA) metric, computed as follows:

$$SA[n] = \frac{1}{18}\sum_{i=1}^{18}(x_i[n]) \tag{8}$$

Here, x_i are the readings from individual sensors, and n enumerates the frames in each sample. Recall that each foot has eight sensors attached (16 total) and the two sums one for each eight sensors for each foot are available giving 18 signals in total.

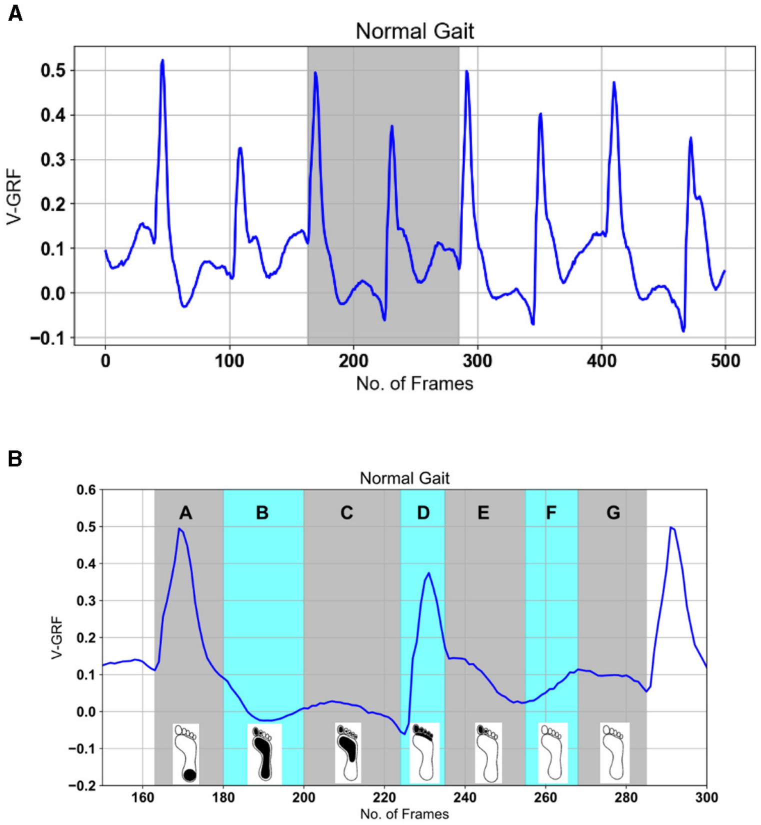


FIGURE 10
Gait events processed SA (see Equation 7) signal top. The highlighted gray area in **(A)** is explained in **(B)** based on gait events for one foot from Figure 4 as: A—heel strike, B—loading response or flat foot, C—mid-stance or single support, D—terminal stance or heel rising, E—pre-swing or double-limb support, F—initial swing and mid-swing or toe-off, G—terminal swing.

Figure 10B shows the expanded gait cycle from Figure 10A with the gait events color-coded and labeled as per Figure 1 and Table 1.

4.2.2.1 Interpretation

The LRP scores highlight the regions of the input data that contribute significantly to the model's classification of PD severity stages. The plot of LRP scores in Figure 11 displays calculated *SA* (top panels) aligned against the relevant "LRP scores" *SA*, which consists of sharp peaks, well defined in the temporal domain, thus attributable to time-stamped gait events. Figure 11 displays the spatially averaged data signals for the four classes with their respective LRP score maps. The most prominent peaks are attributed to observable gait events, labeled in consistency with the gait cycle in Figure 1 and Table 1. It is observed that the model focuses on specific gait features related to severity, such as changes in stride length, gait speed, and variability, to make accurate predictions. These are further discussed in Section 5.

4.2.3 Cognitive load impact on gait event assignment using LRP

The focus of this section is to identify the features picked up by the model to classify gait under cognitive load. To obtain accurate LRP relevance scores *Ri*, the true positive prediction of the model should be high. Therefore, the gait class with a high positive rate is considered for LPR analysis. The learned CNN model parameters in experiments 2 and 4 were frozen for LRP analysis. Experiment 3 is to check whether there is a variation in gait within a subject; therefore, it is not considered for LRP analysis. LRP sequential preset a flat (LRP-SPF) based on the **XAI Selection** criteria was utilized for this part as it has shown sensitivity to gait inconsistency. The iMagiMat system captures a sequence of periodic events as distinct, but similar cycles for each foot. This spatiotemporal sequence is generated by the change of light transmission intensity in the POF sensors: $x_i = [x_1 \& \ldots \& x_{116}] \in \mathbb{R}^{n \times 116}$. However, a typical interpretation of the gait cycle, based on visual observation, is derived much less from the spatial component than the temporal one. Thus, to progress toward interpreting the CNN classifications in terms of observable gait events, we average over the spatial domain using Equation 8.

Figure 12 displays randomly selected samples of normal gait classified with 100% true positives in experiment 2; Figure 13 shows predicted gait samples in experiment 3 for a single. Figure 14 displays randomly selected subjects for comparison of dual tasking with a normal gait. The top panels in Figures 12–14 display calculated *SA* aligned against the relevant "LRP scores" *SA*, generated from the calculated LRP scores and displayed in the bottom panels (to be discussed further in Section 5). The

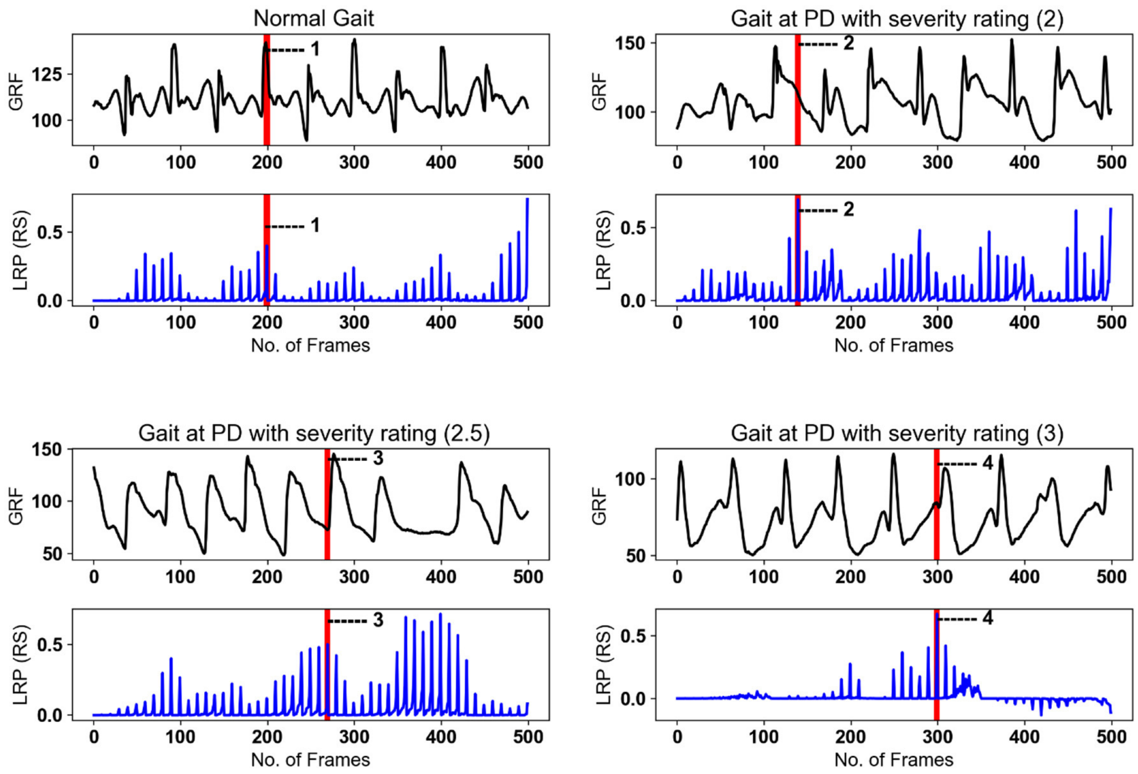


FIGURE 11
LRP method applied on randomly selected samples for healthy gait and three PD severity ratings. SA of gait spatiotemporal signals: black; SA for LRP relevance scores (RS) over the same temporal period: blue. Vertical red bars with number labels display consistency with gait events listed below with capital letters as per Figure 1 (Table 1) and Figure 10: 1—heal strike and foot flattening (A); 2—mid-stance and single support (C); 3—loading response after the double support interval (B), 4—terminal swing and ready for the heel strike (G).

SA temporal sequences have different values on the *y-axis* due to the nature of the captured gait signal, which is influenced by the individual anthropometry of subjects.

4.2.3.1 Interpretation of results

The LRP heatmaps demonstrate the regions in the input data that contribute significantly to the model's decision regarding cognitive load impact on gait. The model appears to focus on variations in gait features influenced by cognitive load, providing insights into the relationship between cognitive demand and gait characteristics.

The presented experiments demonstrate the effectiveness of the CNN model in various gait-related tasks, including PD identification, severity staging, subject identification, and assessing cognitive load impact on gait. The high F1-scores obtained in each experiment indicate the capability of model to make accurate predictions. The LRP analysis provides interpretability by highlighting important regions in the input data for decision-making. In the PD gait identification experiment, the model seems to focus on specific patterns in ground reaction forces related to PD-associated abnormalities. In PD severity staging, the model relies on gait features indicative of severity, such as stride length and variability. For subject cognitive load identification, the model captures unique gait patterns for each individual, and in assessing cognitive load impact, it considers variations influenced by cognitive demand. The ensemble approach consistently shows comparable or improved performance over the single model, indicating its effectiveness in enhancing predictive accuracy. The mean F1-scores across experiments suggest the model's robustness in handling diverse gait-related tasks. Overall, the presented CNN model, accompanied by LRP analysis, provides a powerful tool for gait analysis in the context of Parkinson's disease and related tasks. Further research and validation on larger datasets and diverse populations would contribute to the generalizability and applicability of the proposed model. Additionally, real-world deployment considerations, such as model interpretability in clinical settings, should be explored for practical implementation.

5 Discussion

The study presented delves into the promising realm of explainable artificial intelligence (AI) and deep learning methods for predicting gait deterioration. The focus is on identifying the impact of cognitive load and Parkinson's disease (PD) on gait patterns, and this is achieved by analyzing spatiotemporal data obtained from sensors placed under the feet. To carry out this investigation, convolutional neural networks (CNNs) were utilized. These powerful neural networks can effectively learn from complex spatiotemporal data and produce highly accurate predictions. In addition, the CNNs were perturbed to provide insights into the features within the spatiotemporal gait ground reaction force

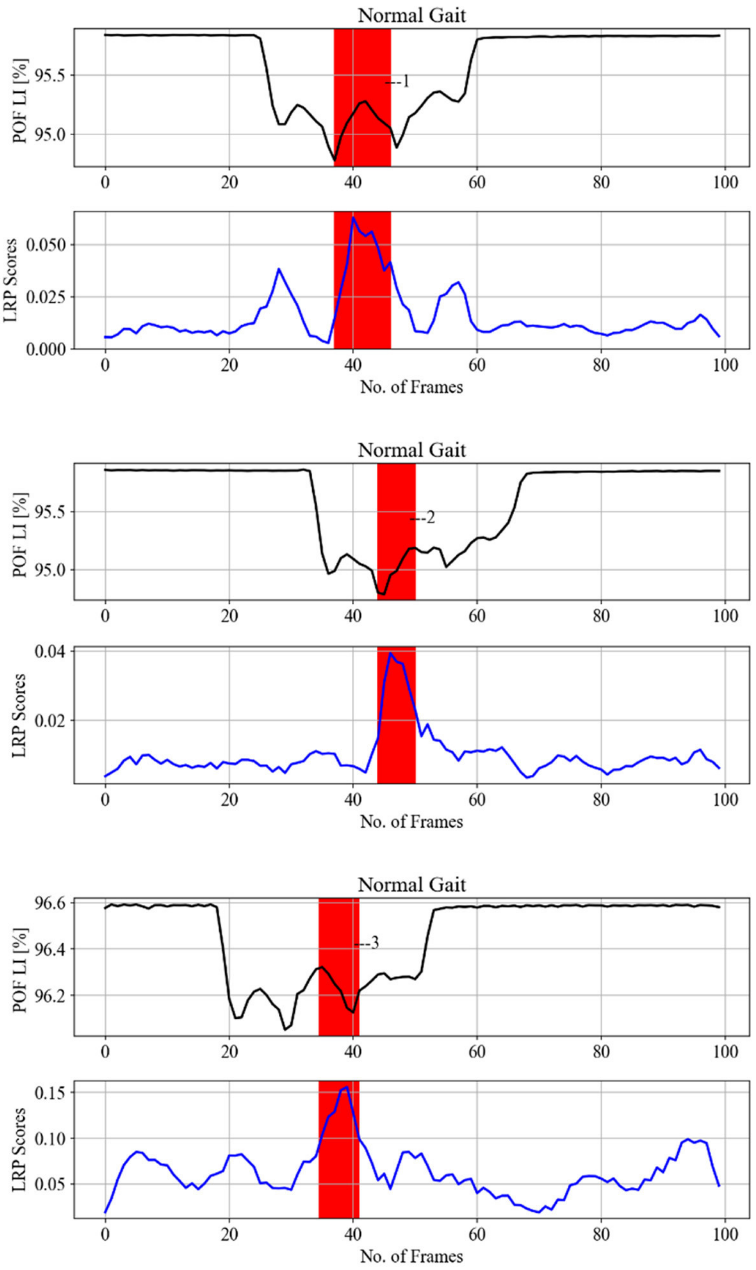


FIGURE 12
LRP methods applied on normal gait samples (from different subjects) from experiment 2 testing data, to identify gait events relevant to the CNN prediction to classify the cognitive load impact on gait. Gait events are 1,2,3—loading response or foot flat and double support.

(GRF) signals that are most relevant to the predictions of the models. The results of this study are presented in detail in the following sections, with each data classification and perturbation analyzed and discussed in depth.

5.1 PD data

The spatiotemporal signal in Figures 4A, 10 implies that gait has normal events. Abnormal gait, otherwise difficult to detect visually, can be detected by machine learning, in alignment with the knowledge of the ground truth labels. However, the magnitude of GRF in Newton shows a decrease attributable to the severity of PD. The main objective of this study was to find the best deep learning model for PD severity rating and relate the model predictions to the gait cycle events shown in Figure 3.

Research toward machine learning classifications from PD data, specifically PhysioNet data, is based on the use of manual feature extraction methods with classical machine learning methods as shown in Table 8. The best classification results from manual

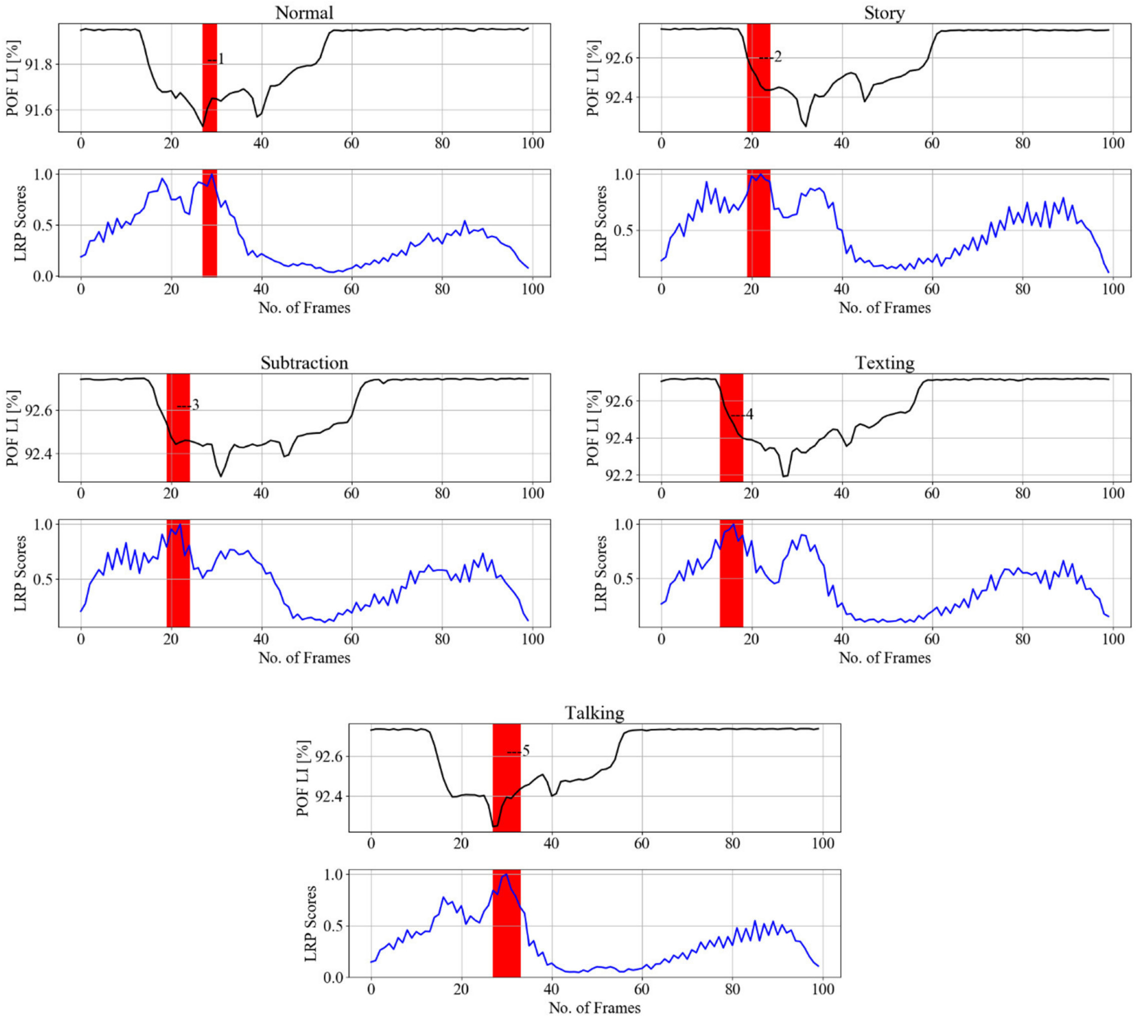


FIGURE 13
LRP methods applied on a single subject from experiment 3 testing data (each column is one pair), to identify gait events relevant for the CNN prediction to classify the cognitive load impact on gait. SA of gait spatiotemporal signals: black; SA for LRP relevance signals over gait temporal period: blue; POF LI (plastic optical fiber light intensity). Vertical red bars with numbers display correspondence to gait events as per Figure 15: 1, 5—loading response or foot flat and double support, 2, 3, 4—loading response or foot flat and single support.

extraction are reported in Abdulhay et al. (2018) using SVM classifier (92.7%). Our study on PD severity classification reported in Table 5 displays that the CNN outperformed the SGD, KNN, GPC algorithms, and LSTM. In this study, we explore three CNN architectures for automatic extraction and LRP analysis. The proposed CNNs identified PD, as well as rated the severity of the deviation from healthy gait, achieving better classification performance with an F1-score of 98% for each dataset and for the datasets combined with different random states (see Table 5). The best classification accuracy is achieved with the parallel CNN, with mean performance and standard errors of 95.5 and 0.28%, respectively. Additionally, the parallel CNN exhibit robustness at perturbation with Gaussian noise as shown in Figure 7. This suggests that the model is adequate for detecting gait deterioration from the spatiotemporal GRF signal. As an additional substantial enhancement, our LRP approach allows classification results to be related to visual observations similar to those established in medical practice to diagnose PD. In this section, we present key findings from our analysis, supported by visual representations. Figure 10 illustrates the spatiotemporal signal extracted from PD data, providing insights into the gait patterns of individuals with Parkinson's disease.

Moving on to Figure 10B, we delve into the gait cycle events identified in PD data. These events play a crucial role in understanding the dynamics of gait abnormalities associated with Parkinson's disease.

To further refine our analysis, Figure 11 presents gait cycle events specifically categorized for PD severity staging. This categorization allows for a nuanced exploration of how gait characteristics vary across different stages of Parkinson's disease. These figures serve as visual aids to enhance the comprehension of our findings and contribute to the broader understanding of gait abnormalities in the context of Parkinson's disease.

(1) PD Severity Level 0 (Healthy Gait): The CNN classifies the raw spatiotemporal signals as healthy or within three severity ratings as shown in the confusion matrix (Figure 5). The best LRP

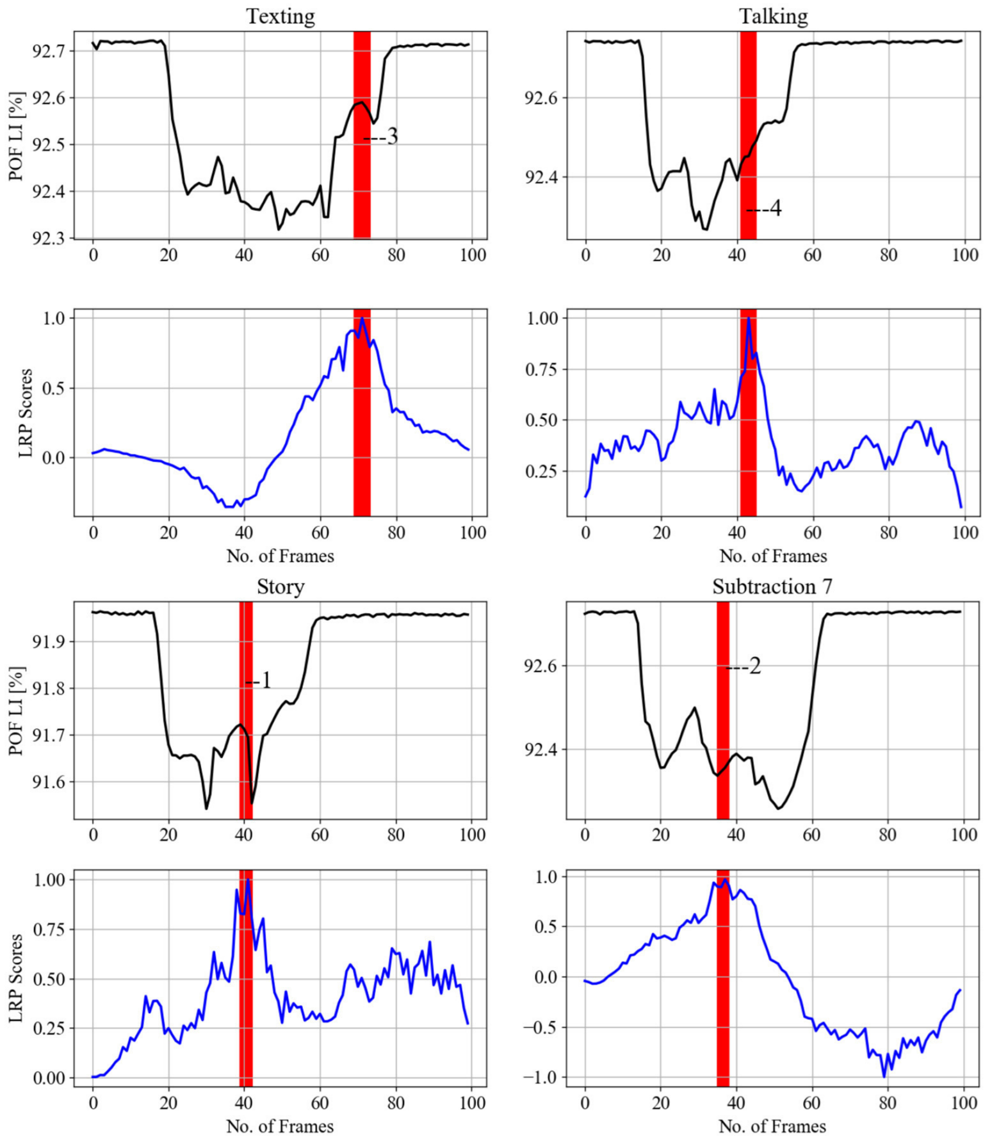


FIGURE 14
LRP methods applied on a single subject from experiment 4 testing data (each column is one pair), to identify gait events relevant for the CNN prediction to classify the cognitive load impact on gait. Gait events are as follows: 1—heel strike, 2—toe-off, 3—between foot swing and opposite heel strike, 4—between double support and toe-off.

method is selected by applying a perturbation technique, which detects the highest sensitivity to removal of information from the input data sequence (Figure 9). The selected LRP-SPF was found to be superior to well-known methods such as deconvolution and guided backpropagation.

Among the CNN architectures (Figure 3), the parallel CNN model shows the steepest decrease in the perturbation procedure. Therefore, that model is learned and used to generate the heatmap or relevance for randomly selected samples (Figure 11). The gait cycle events identified as key at each level of PD severity are listed below:

PD Severity Level 0 (Healthy Gait): (1) Heel strike and foot flattening (A).

This indicates that the healthy person's ability to maintain balance is stronger than the PD patients', with strong balance suggesting that the forces are applied rhythmically to achieve the lower limbs' synchronized movement with stable posture.

PD Severity Level 2: (1) Mid-stance and single support (C).

The heatmap shows that the subjects affected with PD level 2 have a weaker balance in single support, where this feature is marked by the model by 96% F1-score.

PD Severity Level 2.5: Loading response after the double-support interval (B).

This shows that the subject has weaker foot landing or flat foot landing after the balance is compromised by the single support.

PD Severity Level 3: (4) Terminal swing and ready for the heel strike (G).

Here the balance is compromised by weak GRF resulting from unstable body posture and implies a high risk of falling. This

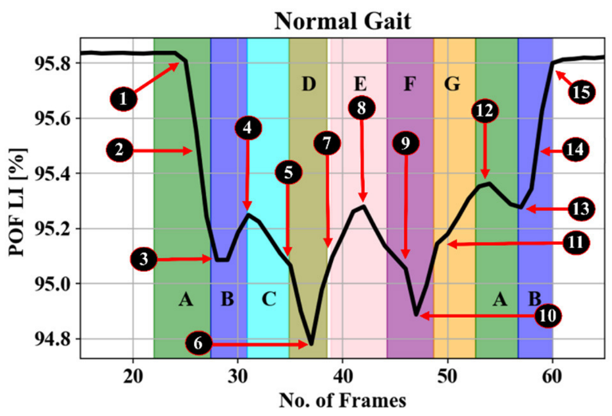


FIGURE 15
Representative gait cycle spatial average of spatiotemporal signals (see Equation 8). Gait events recorded by iMAGiMAT sensors in a typical full gait cycle of two steps (Figure 4: A, B, C, D, E, F, G): 1—heel strike, 2—foot-flattening, 3—single support, 4—opposite heel strike, 5—opposite foot-flattening, 6—double support, 7—toe-off, 8—foot swing, 9—heel strike, 10—double support, 11—toe-off, 12—foot swing, 13—opposite heel strike, 14—single support, 15—toe-off.

TABLE 8 PD classification results on PhysioNet three datasets.

References	Methods	Accuracy (%)
Abdulhay et al. (2018)	SVM	92.7
Jane et al. (2016)	Q-BTDNN	91.5
Ertugrul et al. (2016)	1D-LBP + MLP	88.89
Medeiros et al. (2016)	PCA	81.00
Wu et al. (2017)	SVM	84.48
This study	Parallel CNN	95.5 ± 0.28

SVM, support vector machine; 1D-LBP+MLP, shifted 1D-local binary patterns + multi-layer perceptron; PCA, principal component analysis; Q-BTDNN, Q-back-propagated time delay ANN.

TABLE 9 F1-score predictions for comparison of CNN with classical classifiers.

Classifier	Experiment 3	Experiment 2
SGD	77%	42%, $N = 47\%$
KNN	87%	51%, $N = 81\%$
GPC	5%	22%, $N = 0\%$
CNN	100%	50%, $N = 92\%$

N, True positive prediction of normal gait.

conclusion is based on linking the stages of PD in Wang et al. (2023) (description of how the stage of PD affects the body posture during gait using visual observation) to the events that are highlighted by the model for a certain PD severity.

(2) Interpretation of Classifications: The above markers for classification align with the observations in the literature that PD-induced gait GRF deterioration affects body balance and posture. The latter is with the closest relevance to gait events identified by the heat maps in Figure 11 as the highest LRP scores, while the other gait events are less significant to the classifications. It is worth mentioning that these markers are identical by 95.5% in 1,281 samples, such that the removal of these regions in the 95.5% of samples resulted in a strong decay in the model prediction. The interpretation given above is in very good agreement with the description of the Hoehn and Yahr Scale staging criteria as follows: "Stage 0—No signs of disease, Stage 2—Symptoms on both sides but no impairment of balance, Stage 2.5—Mild symptoms on both sides, with recovery when the 'pull' test is given (the doctor stands behind the person and asks them to maintain their balance when physically pulled backward), Stage 3—Balance impairment, mild-to-moderate disease, physically independent" (International Parkinson and movement disorder society, 2004). However, the staging criteria do not refer to the gait events adversely influencing the body's postural balance, due to the advancement of the disease.

5.2 iMagiMat data

5.2.1 Classification of gait signatures under cognitive load

The present study investigates the importance of cognitive load influence on gait inconsistency. We present a comparison of classification performance between five types of gait: normal and under cognitive load in four different tasks. CNNs not only outperform, unsurprisingly, the classical classifier methods but also achieve an F1-score of 92% for normal gait (Figure 6 and Table 9) for cognitive load impact on gait in experiment 2 with 21 healthy adult data. Understandably the variation in the other cognitive demanding tasks gait is varying among subjects as each subject has a different way of dual taking.

Experiment 3 is, in essence, an extra validation of the adequacy of the spatiotemporal sampling of GRF by the 116 sensors and their fusion as well as the classification performance of the trained models. An F1-score of 100% is achieved for most of the test data. Although Experiment 3 has the character of a sanity check, the results support the value of floor sensor gait data as a biometric. Experiment 2 is conducted to study the possibility of classifying cognitive load on healthy subjects. It has shown that normal gait is classified with a higher true positive rate compared to any of the classes of gait under cognitive load. This experiment also indicates that the achieved true positive rates in predicting normal gait are higher for the CNN model compared to the classical classifiers (see Figure 6 and Table 9). Samples obtained under cognitive load are hard to fit due to the inconsistency of gait pattern changes among the subjects.

The results from the first two experiments suggest that while the dual-task data obviously contributes to the high F1-scores in experiments 2 and 3, it results in substantially degraded true positive rates in experiment 2. However, experiment 3 shows that when classifications are within a single subject the performance is notably better: for 16 subjects (out of 21) the gait under cognitive load the F1-score ranges between 80 and 100%, with the remaining five subjects the range being between 69 and 77%.

These observations can be discussed in the light of humans having a natural gait pattern evolved over millions of years; however, changes in gait when experiencing cognitive load at any particular instance are specific to the individual, expressing their response to the impaired ability to process cognitive information (Chopra et al., 2018). In experiment 4, we use binary classifications (see Table 6) to distinguish normal gait from gait under the 4 variants of cognitive load. The best classification results are obtained when the model learns normal or dual-task gait features for a single subject. This implies that although learned gait features under cognitive load may not be readily portable across subjects, they are consistent for each individual and can contribute substantially for correct subject classifications; however, the accuracy drops if more subjects are involved.

Figures 12–14 provide the link between the LRP relevance scores ("heat map") and the time sequence of the calculated *SA* signal in a single gait cycle window. The LRP score maxima are suitable pointers to the parts of the gait cycle which are most relevant for the classifications. For accurate heat maps of a specific gait class the model's true positive prediction in the confusion matrix must be close to 100% for most of the testing samples, which points to the results from experiment 2 (Figure 12), for normal gait heat maps—in Figure 13 and experiment 3 for a single subject predicted gait under the 4 variants of cognitive load. Focusing just on one complete gait period (two steps) is justified by the fact that on multiple repetitive occasions each subject will initiate a gait cycle (see full description of the gait cycle Figure 1 and Table 1) by performing a heel strike, strictly followed by other gait events described in Figure 15 and ending in a toe off.

The indication of events numbered 1, 2, 3 on Figure 12 implies that normal gait identified by loading response or Foot flat and double support for 21 subjects. This gait event is marked by the model by 92% true positive (see Figure 12) to distinguish normal gait from 4 cognitive load classes. Figure 13 indicates that loading response has high relevance for assigning a gait signature to one out of the 21 subjects gait samples, notably even under cognitive load, as indicated by with gait events numbered from 1 to 5. The indicated gait events are 1,5—loading response or foot flat and double support, 2,3,4—loading response or foot flat and single support. Figure 12 displays the binary classification of randomly selected subject gait events as: 1—heel strike, 2—toe-off, 3—between foot swing and opposite Heel strike, 4—between double support and toe-off. Figure 12 shows cognitive load gait samples for one subject as per experiment 4 summarized as follows:

1. Gait while listening to story: Heel strike is significant for distinguishing listing to story from normal walking.
2. Gait while performing serial 7 subtraction: Toe-off is significant for distinguishing 7 subtraction from normal walking.
3. Gait while texting in smart phone: the transition from foot swing to opposite Heel strike is significant for distinguishing texting from normal walking.
4. Gait while talking: the transition from double support to Toe-off is important to distinguishing talking from normal walking.

Figure 13 indicates that loading response has high relevance for assigning a gait signature to one out of the 21 subjects gait samples, notably even under cognitive load, as indicated by with gait events numbered from 1 to 5. The indicated gait events are 1,5—loading response or foot flat and double support, 2,3,4—Loading response or foot flat and single support. Figure 12 displays the binary classification of randomly selected subject gait events as: 1—heel strike, 2—toe-off, 3—between foot swing and opposite heel strike, 4—between double support and toe-off.

Overall, the LRP analysis indicates that subjects' normal gait is characterized by loading response, while the other cognitive load gait classes are classified by landing or lifting the feet on/from the surface of the iMagiMat system. For subject dual tasking, there are many second relevant scores used to predict the cognitive load of the subject based on gait signature.

6 Conclusion

In conclusion, this study demonstrates the potential of explainable artificial intelligence (XAI) and deep learning methods in predicting gait deterioration. The use of convolutional neural networks (CNNs) on spatiotemporal data obtained from sensors under the feet proves to be effective in identifying the impact of cognitive load and Parkinson's disease (PD) on gait patterns. The proposed CNN architectures show robustness and achieve high classification accuracy for PD severity and cognitive load classification. The local relevance propagation (LRP) analysis provides valuable insights into the features of the spatiotemporal gait ground reaction force (GRF) signals that are most relevant to the model's predictions. The identified gait events and their relevance scores align with existing literature on PD-induced gait deterioration and cognitive load effects on gait. Additionally, the perturbation analysis validates the robustness of the model predictions, and the comparison of LRP methods highlights the effectiveness of the selected LRP-SPF method. The study contributes to the understanding of the relationship between gait events, PD severity, and cognitive load providing a foundation for further research in the field of gait analysis and neurodegenerative diseases. The findings suggest that the proposed model can not only classify gait patterns accurately but also reveal the specific features contributing to these classifications. The experiments conducted in this study shed light on the challenges associated with gait classification under cognitive load. Overfitting observed in the learning curve underscores the importance of addressing the variability in gait patterns induced by cognitive tasks across different subjects. Despite the challenges, the model exhibits promising performance, particularly in distinguishing normal gait from cognitive-loaded gait patterns. The binary classifications in Experiment 5 further emphasize the potential of the model for subject-specific gait analysis. The consistency of learned gait features within individuals suggests the applicability of the model for personalized gait assessments, although caution is warranted when generalizing across a larger population. The interpretation of classifications through LRP heatmaps reveals the relevance of specific gait events in distinguishing between normal and cognitive-loaded gaits. Loading response emerges as a critical gait event for identifying normal gait, while other events such as heel strike and toe-off play distinct roles in classifying cognitive-loaded gaits. The perturbation analysis validates the robustness of the model against the removal of relevant information. The ability of the model to maintain high performance in the presence of random perturbations suggests that it focuses on genuine gait features rather than noise. In conclusion, this comprehensive study not only demonstrates the effectiveness of deep learning models in gait analysis by achieving 98% classification results but also provides interpretability through LRP analysis using perturbation analysis to result in a robust model. The combination of accurate classification, subject-specific insights, and robustness to perturbations positions the proposed model as a valuable tool in clinical settings for assessing gait abnormalities associated with cognitive load and neurodegenerative diseases.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The study was carried out under the University of Manchester Research Ethics Committee (MUREC) with ethical approval number 2018-4881-6782. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

AA: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University for funding this study through Large Research Project under grant number RGP2/254/45.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abdulhay, E., Arunkumar, N., Narasimhan, K., and Vellaiappan, E. (2018). Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease. *Future Gener. Comput. Syst.* 83, 366–373. doi: 10.1016/j.future.2018.02.009

Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., Kim, B., et al. (2018). "Sanity checks for saliency maps," in *Proc. Adv Neural Information Processing Systems (NeurIPS)* (San Diego, CA), 9505–9515.

Alber, M., Lapuschkin, S., Seegerer, P., Hägele, M., Schütt, K. T., Montavon, G., et al. (2018). *iNNvestigate Neural Networks*. Available at: https://github.com/albermax/innvestigate (accessed Jun 19, 2024).

Alharthi, A. S., Yunas, S. U., and Ozanyan, K. B. (2019). Deep learning for monitoring of human gait: a review. *IEEE Sens. J.* 19, 9575–9591. doi: 10.1109/JSEN.2019.2928777

Alotaibi, M., and Mahmood, A. (2015). "Improved gait recognition based on specialized deep convolutional neural network," in *2015 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)* (Washington, DC: IEEE). doi: 10.1109/AIPR.2015.7444550

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* 46:175. doi: 10.1080/00031305.1992.10475879

Arshad, M. Z., Jung, D., Park, M., Shin, H., Kim, J., and Mun, K.-R. (2021). "Gait-based frailty assessment using image representation of IMU signals and deep CNN," in *43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (IEEE). doi: 10.1109/EMBC46164.2021.9630976

Babaee, M., Zhu, Y., Köpüklü, O., Hörmann, S., and Rigoll, G. (2019). "Gait energy image restoration using generative adversarial networks," in *2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan* (New York, NY: IEEE), 2596–2600. doi: 10.1109/ICIP.2019.8803236

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., Samek, W., et al. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE* 10:e0130140. doi: 10.1371/journal.pone.0130140

Balakrishnan, A., Medikonda, J., and Namboothiri, P. K. (2020). "Analysis of the effect of muscle fatigue on gait characteristics using data acquired by wearable sensors," in *2020 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)* (Udupi), 137–140. doi: 10.1109/DISCOVER50404.2020.9278096

Cantoral, C. J., Nurgiyatna, N., Scully, P., and Ozanyan, K. B. (2011). "Smart carpet for imaging of objects' footprint by photonic guided-path tomography," in *IEEE Africon '11* (Victoria Falls: IEEE), 1–6.

Cantoral-Ceballos, J. A., Nurgiyatna, N., Wright, P., Vaughan, J., Brown-Wilson, C., Scully, P. J., et al. (2015). Intelligent carpet system, based on photonic guided-path tomography, for gait and balance monitoring in home environments. *IEEE Sens. J.* 15, 279–289. doi: 10.1109/JSEN.2014.2341455

Castro, F. M., Marín-Jiménez, M. J., Guil, N., and de la Blanca, N. P. (2020). Multimodal feature fusion for CNN-based gait recognition: an empirical comparison. *Neural Comput. Appl.* 32, 14173–14193. doi: 10.1007/s00521-020-04811-z

Castro, F. M., Marín-Jiménez, M. J., Guil, N., López-Tapia, S., and Pérez de la Blanca, N. (2017). "Evaluation of CNN architectures for gait recognition based on optical flow maps," in *2017 International Conference of the Biometrics Special Interest Group* (Darmstadt: IEEE). doi: 10.23919/BIOSIG.2017.8053503

Chaabane, N. B., Conze, P.-H., Lempereur, M., Quellec, G., Rémy-Néris, O., Brochard, S., et al. (2023). Quantitative gait analysis and prediction using artificial intelligence for patients with gait disorders. *Sci. Rep.* 13:23099. doi: 10.1038/s41598-023-49883-8

Chen, X., Luo, X., Weng, J., Luo, W., Li, H., Tian, Q., et al. (2021). Multi-view gait image generation for cross-view gait recognition. *IEEE Trans. Image Process.* 30, 3041–3055. doi: 10.1109/TIP.2021.3055936

Chopra, P., Castelli, D. M., and Dingwell, J. B. (2018). Cognitively demanding object negotiation while walking and texting. *Sci. Rep.* 8:17880. doi: 10.1038/s41598-018-36230-5

Costilla-Reyes, O., Scully, P., Leroi, I., and Ozanyan, K. B. (2021). Age-related differences in healthy adults walking patterns under a cognitive task with deep neural networks. *IEEE Sens. J.* 21, 2353–2363. doi: 10.1109/JSEN.2020.3021349

Delgado-Escaño, R., Castro, F. M., Ramos Cózar, J., Marín-Jiménez, M. J., and Guil, N. (2018). An end-to-end multi-task and fusion CNN for inertial-based gait recognition. *IEEE Access* 7, 1897–1908. doi: 10.1109/ACCESS.2018.2886899

Erdaş, Ç. B., Sumer, E., and Kibaroglu, S. (2022). CNN-based severity prediction of neurodegenerative diseases using gait data. *Digit. Health* 8:20552076221075147. doi: 10.1177/20552076221075147

Ertugrul, Ö. F., Kaya, Y., Tekin, R., and Almali, M. N. (2016). Detection of Parkinson's disease by shifted one dimensional local binary patterns from gait. *Expert Syst. Appl.* 56, 156–163. doi: 10.1016/j.eswa.2016.03.018

Frenkel-Toledo, S., Giladi, N., Peretz, C., Herman, T., Gruendlinger, L., Hausdorff, J. M., et al. (2005). Treadmill walking as a pacemaker to improve gait rhythm and stability in Parkinson's disease. *Mov. Disord.* 20, 1109–1114. doi: 10.1002/mds.20507

Glorot, X., and Bengio, Y. (2010). "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, ser. Proceedings of Machine Learning Research*, Vol. 9, eds. Y.W. The, and M. Titterington (Sardinia: PMLR), 249–256. Available at: https://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf (accessed August 5, 2024).

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2003). PhysioBank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* 101, e215–e220. doi: 10.1161/01.CIR.101.23.e215

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge: MIT Press, 326–415.

Hausdorff, M. J., Lowenthal, J., Herman, T., Gruendlinger, L., Peretz, C., Giladi, N., et al. (2007). Rhythmic auditory stimulation modulates gait variability in Parkinson's disease. *Eur. J. Neurosci.* 26, 2369–2375. doi: 10.1111/j.1460-9568.2007.05810.x

Horst, F., Lapuschkin, S., Samek, W., Müller, K.-R., and Schöllhorn, W. I. (2019). Explaining the unique nature of individual gait patterns with deep learning. *Sci. Rep.* 9:2391. doi: 10.1038/s41598-019-38748-8

Huang, H., Pan, Z., Ye, L., and Fangmin, S. (2021). A lightweight attention-based CNN model for efficient gait recognition with wearable IMU sensors. *Sensors* 21:2866. doi: 10.3390/s21082866

International Parkinson and movement disorder society (2004). *Rating scales*. Available at: https://www.movementdisorders.org/MDS/MDS-Rating-Scales/Short-Parkinsons-Evaluation-Scale-SPESSCales-for-Outcomes-in-PArkinsons-disease--Motor-Func.htm (accessed June 19, 2024).

Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv* [Preprint]. arXiv:1502.03167. doi: 10.48550/arXiv.1502.03167

Jane, Y. N., Nehemiah, H. K., and Arputharaj, K. (2016). A Q-backpropagated time delay neural network for diagnosing severity of gait disturbances in Parkinson's disease. *J. Biomed. Inform.* 60, 169–176. doi: 10.1016/j.jbi.2016.01.014

Jing, G., Gu, P., Ren, Q. and Zhang, J., and Song, X. (2019). Abnormal gait recognition algorithm based on LSTM-CNN fusion network. *IEEE Access* 7, 163180–163190. doi: 10.1109/ACCESS.2019.2950254

Jolly, S., Iwana, B. K., Kuroki, R., and Uchida, S. (2018). "How do convolutional neural networks learn design?" in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, pp. 1085–1090. doi: 10.1109/ICPR.2018.8545624

Khan, M. A., Kadry, S., Parwekar, P., Damaševičius, R., Mehmood, A., Ali Khan, J., et al. (2023). Human gait analysis for osteoarthritis prediction: a framework of deep learning and kernel extreme learning machine. *Complex Intell. Syst.* 9, 2665–2683. doi: 10.1007/s40747-020-00244-2

Kingma, D. P., and Ba, J. L. (2015) "Adam: a method for stochastic optimization," in *3rd International Conference on Learning Representations (ICLR)* (San Diego, CA). Available at: https://arxiv.org/abs/1412.6980 (accessed August 5, 2024).

Kohlbrenner, M., Bauer, A., Nakajima, S., Binder, A., Samek, W., Lapuschkin, S., et al. (2019). "Towards best practice in explaining neural network decisions with LRP," in *Computer Vision and Pattern Recognition (cs.CV); Machine Learning (stat.ML)* (New York, NY: Cornell University). arXiv:1910.09840. doi: 10.48550/arXiv.1910.09840

Lundberg, S. M., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proc. Adv. Neural Inf. Process. Syst.* 30, 1–10.

Ma, T., Avati, A., Katanforoosh, K., and Ng, A. (2024). *CS229 Lecture Notes, Deep Learning. Stanford CS229 Machine Learning*. Available at: https://cs229.stanford.edu/notes2020spring/cs229-notes-deep_learning.pdf (accessed July 16, 2024).

Medeiros, L., Almeida, H., Dias, L., Perkusich, M., and Fischer, R. (2016). "A gait analysis approach to track Parkinson's disease evolution using principal component analysis," in *IEEE 29th International Symposium on Computer-Based Medical Systems (CBMS) Belfast and Dublin, Ireland* (New York, NY: IEEE), 48–53. doi: 10.1109/CBMS.2016.14

Mogan, J. N., Chin, P. L., Kian, M. L., Mohammed, A., and Ali, A. (2023). Gait-CNN-ViT: multi-model gait recognition with convolutional neural networks and vision transformer. *Sensors* 23:3809. doi: 10.3390/s23083809

Montavon, G., Lapuschkin, S., Binder, A., Samek, W., and Müller, K.-R. (2017). Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognit.* 65, 211–222. doi: 10.1016/j.patcog.2016.11.008

Montavon, G., Samek, W., and Müller, K.-R. (2018). Methods for interpreting and understanding deep neural networks. *Digit. Signal Process* 73, 1–15. doi: 10.1016/j.dsp.2017.10.011

Muheidat, F., and Tawalbeh, L. A. (2020). In-home floor based sensor system-smart carpet- to facilitate healthy aging in place (AIP). *IEEE Access* 8, 178627–178638. doi: 10.1109/ACCESS.2020.3027535

Ngo, T. T., Makihara, Y., Nagahara, H., Mukaigawa, Y., and Yagi, Y. (2014). The largest inertial sensor-based gait database and performance evaluation of gait-based personal authentication. *Pattern Recognit.* 47, 228–237. doi: 10.1016/j.patcog.2013.06.028

Ozanyan, K. B. (2015). "Tomography defined as sensor fusion," in *2015 IEEE SENSORS* (Busan: IEEE), 1–4. doi: 10.1109/ICSENS.2015.7370554

Ozanyan, K. B., Castillo, S. G., and Ortiz, F. J. P. (2005). Guided-path tomography sensors for nonplanar mapping. *IEEE Sens. J.* 5, 167–174. doi: 10.1109/JSEN.2005.843895

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "'Why should I trust you?': explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining* (New York, NY: Association for Computing Machinery), 1135–1144. doi: 10.1145/2939672.2939778

Ruuska, S., Hämäläinen, W., Kajava, S., Mughal, M., Matilainen, P., Mononen, J., et al. (2018). Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle. *Behav. Processes* 148, 56–62. doi: 10.1016/j.beproc.2018.01.004

Sadeghzadehyadi, N., Batabyal, T., and Acton, S. T. (2021). Modeling spatiotemporal patterns of gait anomaly with a CNN-LSTM deep neural network. *Expert Syst. Appl.* 185:115582. doi: 10.1016/j.eswa.2021.115582

Saleh, A., and Hamoud, M. T. (2021). Analysis and best parameters selection for person recognition based on gait model using CNN algorithm and image augmentation. *J. Big data* 8:1. doi: 10.1186/s40537-020-00387-6

Samek, W., Binder, A., Lapuschkin, S., and Müller, K.-R. (2017a). "Understanding and comparing deep neural networks for age and gender classification," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy* (New York, NY: IEEE), 1629—1638, doi: 10.1109/ICCVW.2017.191

Samek, W., Binder, A., Montavon, G., Lapuschkin, S., and Müller, K. (2017b). Evaluating the visualization of what a deep neural network has learned. *IEEE Trans. Neural Netw. Learn. Syst.* 28, 2660–2673. doi: 10.1109/TNNLS.2016.2599820

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., et al. (2017). "Grad-CAM: visual explanations from deep networks via gradient-based localization," in *International Conference on Computer Vision (ICCV), Venice, Italy* (New York, NY: IEEE), 618–626. doi: 10.1109/ICCV.2017.74

Slijepcvic, D., Horst, F., Lapuschkin, S., Horsak, B., Raberger, A.-M., Kranzl, A., et al. (2021). Explaining machine learning models for clinical gait analysis. *ACM Trans. Comput. Healthc.* 3, 1–17. doi: 10.1145/3474121

Springenberg, J., Dosovitskiy, A., Brox, T., and Riedmiller, M. (2015). Striving for simplicity: the all convolutional net. *arXiv* [Preprint]. arXiv:1412.6806. doi: 10.48550/arXiv.1412.6806

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958. Available at: https://jmlr.org/papers/v15/srivastava14a.html (accessed August 2024).

Tran, L., Hoang, T., Nguyen, T., Kim, H., and Choi, D. (2021). Multi-model long short-term memory network for gait recognition using window-based data segment. *IEEE Access* 9, 23826–23839. doi: 10.1109/ACCESS.2021.3056880

Turner, A., and Hayes, S. (2019). The classification of minor gait alterations using wearable sensors and deep learning. *iEEE Trans. Biomed. Eng.* 66, 3136–3145. doi: 10.1109/TBME.2019.2900863

Vidya, B., and Sasikumar, P. (2022). Parkinson's disease diagnosis and stage prediction based on gait single analysis using EMD and CNN-LSTM network. *Eng. Appl. Artif. Intell.* 114:105099. doi: 10.1016/j.engappai.2022.105099

Wang, X., and Yan, K. (2021). Gait classification through CNN based ensemble learning. *Multimed. Tools Appl.* 80, 1565–1581. doi: 10.1007/s11042-020-09777-7

Wang, X., and Zhang, J. (2020). Gait feature extraction and gait classification using two-branch CNN. *Multimed. Data Tools Appl.* 79:29172930. doi: 10.1007/s11042-019-08509-w

Wang, Y., Yang, Q., Tian, C., Zeng, J., Yang, M., Li, J., et al. (2023). A dual-task gait test detects mild cognitive impairment with a specificity of 91.2%. *Front. Neurosci.* 16:1100642. doi: 10.3389/fnins.2022.1100642

Whittle, M. W. (2023). *Whittle's Gait Analysis*, 6th Edn. Amsterdam: Elsevier, 30.

Wu, Y., Chen, P., Luo, X., Wu, M., Liao, L., Yang, S., et al. (2017). Measuring signal fluctuations in gait rhythm time series of patients with Parkinson's disease using entropy parameters. *Biomed. Signal Process. Control* 31, 265–271. doi: 10.1016/j.bspc.2016.08.022

Wu, Z., Huang, Y., Wang, L., Wang, X., and Tan, T. (2016). A comprehensive study on cross-view gait based human identification with deep CNNs. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 209–226. doi: 10.1109/TPAMI.2016.2545669

Xiao, G., Cheng, Q., and Zhang, C. (2019). Detecting travel modes using rule-based classification system and Gaussian process classifier. *IEEE Access* 7, 116741–116752. doi: 10.1109/ACCESS.2019.2936443

Yogev, G., Giladi, N., Peretz, N., Springer, S., Simon, E. S., Hausdorff, M. J., et al. (2005). Dual tasking, gait rhythmicity, and Parkinson's disease: which aspects of gait are attention demanding? *Eur. J. Neurosci.* 22, 1248–1256. doi: 10.1111/j.1460-9568.2005.04298.x

Yuan, W., and Zhang, L. (2018). "Gait classification and identity authentication using CNN," in *Methods and Applications for Modeling and Simulation of Complex Systems. AsiaSim 2018. Communications in Computer and Information Science, vol 946*, eds in L. Li, K. Hasegawa, and S. Tanaka (Singapore: Springer). doi: 10.1007/978-981-13-2853-4_10

Yuqi, Z., Huang, Y., Wang, L., and Yu, S. (2019). A comprehensive study on gait biometrics using a joint CNN based method. *Pattern Recognit.* 93, 228–236. doi: 10.1016/j.patcog.2019.04.023

Zeiler, M. D., and Fergus, R. (2014). "Visualizing and understanding convolutional networks," in *Computer Vision - ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, Vol. 8689*, eds. D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Cham: Springer). doi: 10.1007/978-3-319-10590-1_53

Zhang, P., Wu, Q., and Xu, J. (2019). "VT-GAN: view transformation GAN for gait recognition across views," in *2019 International Joint Conference on Neural Networks (IJCNN)* (Budapest), 1–8. doi: 10.1109/IJCNN.2019.8852258

Zhang, S., Poon, S. K., Vuong, K., Sneddon, A., and Loy, C. T. (2019). "A deep learning-based approach for gait analysis in huntington disease," in *MEDINFO 2019: Health and Wellbeing e-Networks for All* (Amsterdam: IOS Press), 477–481.

Zhang, T. (2004). "Solving large scale linear prediction problems using stochastic gradient descent algorithms," in *Machine Learning, Proceedings of the Twenty-first International Conference (ICML 2004)* (Banff, AB: ICML), 116. doi: 10.1145/1015330.1015332

Zhang, Z., and Sabuncu, M. R. (2018). "Generalized cross entropy loss for training deep neural networks with noisy labels," in *2nd Conference on Neural Information Processing Systems (NeurIPS 2018)* (Montréal, QC). arXiv:1805.07836. doi: 10.48550/arXiv.1805.07836

Zou, Q., Wang, Y., Wang, Q., Zhao, Y., and Li, Q. (2020). Deep learning-based gait recognition using smartphones in the wild. *IEEE Trans. Inf. Forensics Secur.* 15, 3197–3212. doi: 10.1109/TIFS.2020.2985628

frontiers | Frontiers in Neuroinformatics

TYPE Original Research
PUBLISHED 30 October 2024
DOI 10.3389/fninf.2024.1495571

Spectral graph convolutional neural network for Alzheimer's disease diagnosis and multi-disease categorization from functional brain changes in magnetic resonance images

Hadeel Alharbi[1]*, Roben A. Juanatas[2], Abdullah Al Hejaili[3] and Se-jung Lim[4]*

[1]College of Computer Science and Engineering, University of Hail, Ha'il, Saudi Arabia, [2]College of Computing and Information Technologies, National University, Manila, Philippines, [3]Faculty of Computers and Information Technology, Computer Science Department, University of Tabuk, Tabuk, Saudi Arabia, [4]School of Electrical and Computer Engineering, Yeosu Campus, Chonnam National University, Gwangju, Republic of Korea

OPEN ACCESS

EDITED BY
Sadique Ahmad,
Prince Sultan University, Saudi Arabia

REVIEWED BY
Muhammad Rizwan,
University of Derby, United Kingdom
Wilfried Yves Hamilton Adoni,
Helmholtz Association of German Research Centres (HZ), Germany

*CORRESPONDENCE
Hadeel Alharbi
✉ h.alharbe@uoh.edu.sa
Se-jung Lim
✉ limsejung@chonnam.ac.kr

RECEIVED 12 September 2024
ACCEPTED 14 October 2024
PUBLISHED 30 October 2024

CITATION
Alharbi H, Juanatas RA, Al Hejaili A and Lim S-j (2024) Spectral graph convolutional neural network for Alzheimer's disease diagnosis and multi-disease categorization from functional brain changes in magnetic resonance images. *Front. Neuroinform.* 18:1495571.
doi: 10.3389/fninf.2024.1495571

Alzheimer's disease (AD) is a progressive neurological disorder characterized by the gradual deterioration of cognitive functions, leading to dementia and significantly impacting the quality of life for millions of people worldwide. Early and accurate diagnosis is crucial for the effective management and treatment of this debilitating condition. This study introduces a novel framework based on Spectral Graph Convolutional Neural Networks (SGCNN) for diagnosing AD and categorizing multiple diseases through the analysis of functional changes in brain structures captured via magnetic resonance imaging (MRI). To assess the effectiveness of our approach, we systematically analyze structural modifications to the SGCNN model through comprehensive ablation studies. The performance of various Convolutional Neural Networks (CNNs) is also evaluated, including SGCNN variants, Base CNN, Lean CNN, and Deep CNN. We begin with the original SGCNN model, which serves as our baseline and achieves a commendable classification accuracy of 93%. In our investigation, we perform two distinct ablation studies on the SGCNN model to examine how specific structural changes impact its performance. The results reveal that Ablation Model 1 significantly enhances accuracy, achieving an impressive 95%, while Ablation Model 2 maintains the baseline accuracy of 93%. Additionally, the Base CNN model demonstrates strong performance with a classification accuracy of 93%, whereas both the Lean CNN and Deep CNN models achieve 94% accuracy, indicating their competitive capabilities. To validate the models' effectiveness, we utilize multiple evaluation metrics, including accuracy, precision, recall, and F1-score, ensuring a thorough assessment of their performance. Our findings underscore that Ablation Model 1 (SGCNN Model 1) delivers the highest predictive accuracy among the tested models, highlighting its potential as a robust approach for Alzheimer's image classification. Ultimately, this research aims to facilitate early diagnosis and treatment of AD, contributing to improved patient outcomes and advancing the field of neurodegenerative disease diagnosis.

KEYWORDS

Alzheimer's disease (AD), image classification, Convolutional Neural Networks (CNN), SGCNN model, deep learning, ablation study

1 Introduction

Neurodegenerative disease, such as Alzheimer's disease (AD), is the most prevalent type of dementia that affects 60% to 80% of patients in the world (Turer and Sanlier, 2024; Vejandla et al., 2024). It is characterized by a decline in cognitive processes, including language, reasoning, and memory, ultimately leading to the inability to perform daily activities. The elderly population has a higher prevalence of the disease and is the fourth leading cause of death (Self and Holtzman, 2023). There is currently no cure for AD, despite extensive research and presently available medications for AD solely working on managing symptoms of the disease and putting into effect a vast financial burden for the health care system, patients and their families (Vejandla et al., 2024).

AD is progressively emerging as the most prevalent neurological disorder, with its numbers likely to rise by 2050 globally from 50 to 100 million (Zhao et al., 2024). There is an urgent need for reliable and efficient methods to detect AD at its initial stages. Early diagnosis leads to timely therapeutic interventions that potentially slow the disease progression and relieve the great burden on healthcare systems (Garg et al., 2023). Cognitive decline in AD starts many years before it is manifested clinically; the first stage may be Mild Cognitive Impairment (MCI), which may lead to AD. About 15%–20% of individuals over 60 years suffer from MCI, with 30%–35% progressing to AD within four years (Karran and De Strooper, 2022). The accumulation of coagulated tau proteins and amyloid-beta ($A\beta$) plaques causes neuronal death and brain shrinkage, which is the cause of the disease. This tissue loss occurs starting with the Gray Matter (GM), then going into the White Matter (WM), Corpus Callosum (CC), and extending to the Hippocampus (HC), greatly impairing neural functions (Knopman et al., 2021). Early diagnosis of AD is critical for maintaining good disease management and improving patient quality of life (Begum and Selvaraj, 2024). Modern diagnostic methods, such as PET and MRI scans, are crucial for the diagnosis of AD because they identify both structural and functional changes in the brain (Porsteinsson et al., 2021). The important information regarding the disease development from normal cognitive (NC) function through MCI to full-blown AD is provided by these imaging modalities in addition to other clinical data (Shukla et al., 2023).

New developments in machine learning technologies, especially techniques of deep learning like CNNs, have shown enormous potential in the early diagnosis and classification of AD (Wen et al., 2020). CNNs are superior in pattern recognition and image classification, which makes them ideal for large dataset image analysis, in this case, medical imaging data. Taking advantage of MRI, PET scans, and Diffusion Tensor Imaging (DTI) information, CNNs can help in the efficient and effective identification of AD and predict its progression from MCI to AD (Logan et al., 2021).

1.1 Research contribution

This study offers three key contributions to the field of AD diagnosis and multi-disease classification:

- We propose a novel framework based on Spectral Graph Convolutional Neural Networks (SGCNN) for diagnosing AD and categorizing multiple diseases by analyzing functional brain changes observed in Magnetic Resonance Images (MRI). Structural modifications to the SGCNN model are rigorously analyzed through ablation studies, and the performance of various Convolutional Neural Network (CNN) models, including SGCNN variants, Base CNN, Lean CNN, and Deep CNN, is systematically evaluated.
- Our study improves the reliability of AD classification tasks by implementing essential preprocessing steps such as image visualization, pixel value normalization, and precise dataset splitting. These processes ensure higher quality and consistency within the dataset, which directly enhances the performance and accuracy of the CNN models.
- Through extensive experimentation, the Ablation of SGCNN Model 1 achieves a classification accuracy of 95%, highlighting its superior potential for early detection and diagnosis of AD. This result demonstrates the effectiveness of the proposed model modifications in advancing the field of neurodegenerative disease diagnosis.

1.2 Research organization

This document is formatted as follows: In Section 2, the background information and relevant works are provided. Section 3 presents the proposed deep-learning approach for categorizing images associated with AD. We assess the performance of our technique and compare it with the baseline methods in Section 4. The article is concluded in Section 5, which offers suggestions for further reading.

2 Literature review

In the literature review, we address machine learning (ML) and deep learning (DL) methodologies for AD prediction. This section explores how these advanced methodologies contribute to improving diagnostic accuracy and understanding AD progression. Table 1 provided the overview of studies on AD prediction.

Biswas and Gini (2024) suggested an output-based multi-class categorization system ranging from Normal to Severe facilitates the early identification of AD. It starts by extracting hippocampal, gray and white matter from 3D MRI images and computing the volumes of each from the images using Analyze Direct and ITK Snap. Such volumes, besides other characteristics like age, gender

---

Abbreviations: AD, Alzheimer's disease; ADRD, Alzheimer's disease and related dementias; AUC, area under the curve; CNN, convolutional neural network; DTI, diffusion tensor imaging; DL, deep learning; ID3, iterative dichotomiser 3; LGBM, light gradient boost machine; ML, machine learning; MRI, magnetic resonance imaging; MCI, mild cognitive impairment; NLP, natural language processing; NC, normal control; PD, Parkinson's disease; PET, positron emission tomography; QVC, quantum variational circuit; RF, random forest; ROC, receiver operating characteristic; SGCNN, spectral graph convolutional neural network; SVM, support vector machine; VGNN, variational graph neural networks.

TABLE 1 Overview of studies on AD classification.

References	Approach	Dataset	Classes and descriptions	No. of images	Key findings
Rao et al. (2024)	3D convolutional neural networks with transfer learning	MRI brain images	AD, Mild Cognitive Impairment (MCI), Normal Control (NC)	1,686 images	ResNet50V2 achieved 92.15% training accuracy and 91.25% testing accuracy
Tripathi and Kumar (2024)	Speech-based cognitive impairment assessment using ML	DementiaBank's Pitt Corpus	Six classes based on cognitive impairment levels	292 recordings	Achieved 75.59% accuracy in six-class classification; XGBoost showed significant accuracy differences
Krishna et al. (2024)	DL with SMOTE data augmentation for MRI data	MRI datasets	AD, MCI, NC	2,453 images	Improved model accuracy and validity, effective for imbalanced data
Srividhya et al. (2024)	CNN-based multi-class classification	ADNI2 (sMRI)	AD, MCI, NC	1120 images	ResNet-50v2 achieved 91.84% mean accuracy, F1-score of 0.97 for AD class
Goenka and Tiwari (2023)	Multimodal DL for Alzheimer's classification	ADNI (T1-weighted MRI, AV-45 PET)	AD, MCI, NC	2,391 images	3D-Subject method achieved 93.01% accuracy, surpassing Patch-based (89.55%) and Slice-based (89.37%)
Francis and Pandian (2023)	Ensemble of pre-trained models for multi-class classification	ADNI (T1-weighted sMRI)	AD, MCI, NC	2,156 images	Achieved 85% accuracy in multi-class classification; outperformed other state-of-the-art methods
Venkatasubramanian et al. (2023)	MTDL for segmentation and classification	ADNI (structural MRI)	AD, MCI, NC	2128 images	Achieved 97.1% accuracy, 93.5% Dice coefficient, 96% accuracy for binary, 93% for multi-class classification

and MMSE scores, are used to feed machine learning algorithms such as random forest, gradient boost, decision tree and KNN for the detection of Alzheimer's and the classification of the severity level of Alzheimer's. Additionally, the collected traits are arbitrarily mixed in every feasible way, including feature-level fusion, and further analyzed. The methodology is tested on two datasets, OASIS and ADNI, which were introduced in earlier sections. In the OASIS dataset, a 99% accuracy is achieved by random forest when using only white matter volume and 98% when all three volumes are integrated. For the ADNI data set, for white matter volume, the accuracy was found to be 92% for gradient boost, and for the combination of all three volumes when fused, the accuracy was 91% for both databases.

Rao et al. (2024) deal with AD by creating a new deep-learning approach that generalizes convolution networks in the third dimension to model spatial characteristics of the 3D MRI scans. The proposed classification system also uses attributes that are taken from the 3D convolutional network's several layers; however, it gives distinction importance to each layer. Using brain MRI scans from three classes (Mild Cognitive Impairment, Normal Control, AD and probability controls), the system combines transfer learning with fine-tuning. In regards to AD classification, the researchers also tried using pre-trained deep learning models such as ResNet50V2 and InceptionResNetV2, of which ResNet50V2 performed better. According to their results, ResNet50V2 achieved a testing accuracy of 91.25% and a training accuracy of 92.15%. The authors observed that the effective detection of AD utilizing 3D MRI brain images can be achieved using deep learning, particularly transfer learning with ResNet50V2.

Tripathi and Kumar (2024) suggest a method for the speech-based assessment of six kinds of cognitive impairment. After pre-processing the speech data from DementiaBank's Pitt Corpus to extract pertinent acoustic features, they train five machine learning algorithms (KNN, DT, SVM, XGBoost, and RF). Consequently, the work's output demonstrates a 75.59% accuracy rate in the six-class classification task. Besides, the significance of differences in the accuracy of XGBoost as compared to the other algorithms except the random forest is proved by the statistical tests. This approach has the potential to be used as cost- and time-effective compared to a provision of a medical diagnosis that is easily accessible in the early phases of the disease. Krishna et al. (2024) present a method that combines DL methodologies, including Deep Learning (DL), with methods of data augmentation of the SMOTE type for any MRI dataset to enhance the detection of Alzheimer's disorder. They are being used here because this approach can enhance the accuracy as well as the validity of the classification model due to the great management of the problems associated with the imbalanced data. Based on the present interdisciplinary analysis, the integration of DL with SMOTE improves the model's ability to identify AD, and this improvement was also observed when expanding its application to other forms of neurodegenerative diseases.

Srividhya et al. (2024) put forward a framework for the clustering of the stages of AD based on the AD Neuroimaging Initiative (ADNI2—Structural Magnetic Resonance Imaging—sMRI) image database. The approach entails the use of deep learning techniques, especially CNN, for a multi-class classification of AD MRI images. The emphasis is placed on choosing the most suitable pre-trained model that will be able to provide the best prediction for the AD stage of a particular patient. ResNet-50v2 was the overall best model, with a mean 91.84% accuracy and an F1-score value of 0.97 for the AD class. They used Grad-CAM and Saliency Map to visualize the highest accuracy model to know which part of the image the algorithm concentrated on for classification. Kaya and Çetin-Kaya (2024) put forward a framework that involves the use of PSO to adjust the hyperparameters of CNN for the detection of AD from MRI data. The approach comes in handy to fine-tweak hyperparameters like a number of convolution layers and filters and other issues like lack of labeled data, high inter-class similarity, and overfitting. As for the proposed lightweight model, it attains a test accuracy of 99.53%, and an F1-score of 99.63% of the tests were performed on a public dataset, which was higher than those obtained in prior studies and could be highly useful to help clinicians in the diagnosis and decision-making process.

By utilizing MRI data from the ADNI dataset, the El-Assy et al. (2024) provide a cutting-edge CNN design for the classification of AD. The two CNN models used by the network have varying filter widths and pooling layers. Nonetheless, these two models are combined for classification purposes because the system handles three, four, and five categories. With 99.43%, 99.57%, and 99.13% accuracy, respectively, the suggested CNN architecture produces comparatively high results. These results demonstrate the suggested network's ability to extract features from MRI scans and differentiate between various AD subtypes and stages, assisting medical professionals in accurately and promptly diagnosing AD patients. Khan et al. (2024) introduce their new multimodal fusion-based approach called Dual-3DM3-AD to diagnose AD from the MRI and PET image scans accurately and in the early stages. The management starts with the pre-processing of both image types: For the noise reduction of the raw data, a Quaternion Non-local Means Denoising Algorithm (QNLM) is applied. Subsequently, the Morphology function is used for skull stripping, resulting in an improved image quality further refined with the help of a Block Divider Model (BDM) to convert the 2D image into a 3D image. The model incorporates semantic segmentation using a Mixed-transformer with Furthered U-Net with Complexity Minimization. It employs the Densely Connected Feature Aggregator Module (DCFAM) for feature aggregation and implements a multi-scale feature extraction to extract features from the segmented images it obtains. There is then feature dimensionality reduction by applying multi-head attention, wherein a softmax layer is used, covering multi-class diagnosis of Alzheimer's. The proposed Dual-3DM3-AD achieves a high accuracy of 98% and a high sensitivity of 97.8%, specificity of 97.5%, F-measure of 98.2%, and ROC curves that are statistically significantly better than any other existing model for multi-class Alzheimer diagnosis.

Hu et al. (2024) study leverage Graph Neural Networks (GNNs) with claim data to predict AD and Related Dementia (ADRD) risk. A variational GNN (VGNN) with a relation importance method was used to estimate ADRD likelihood and provide explanations of feature importance. Three prediction scenarios (1-, 2-, and 3-year windows) were analyzed, and VGNN performance was compared to the Random Forest (RF) and Light Gradient Boost Machine (LGBM) models. Across all scenarios, the VGNN outperformed RF and LGBM models, with AUROC improvements of over 9%–10%. The VGNN showed strong predictive ability, with AUROC scores ranging from 0.7001 to 0.7480, highlighting its efficacy in ADRD risk prediction. In Amini et al. (2024), Natural Language Processing (NLP) techniques combined with machine learning methods were utilized to develop an automated approach for predicting the progression from MCI to AD within a 6-year timeframe based on speech data. The study analyzed neuropsychological test interviews of 166 participants from the Framingham Heart Study, comprising 90 cases of progressive MCI and 76 cases of stable MCI. The best-performing models incorporated speech-derived features along with demographic factors such as age, sex, and education level, achieving an accuracy of 78.5% and a sensitivity of 81.1% in predicting MCI-to-AD progression.

Goenka and Tiwari (2023) use T1-weighted MRI and AV-45 PET images from the ADNI database to provide a unique multimodal deep-learning model for the categorization of AD. They use three cutting-edge approaches: 3D-Subject, 3D-Patches, and 3D-Slices. The 3D-Patches, a unique feature, include patches of different sizes from 32 to 88 for feature extractions. In contrast, the 3D-Slices, another novel approach, include uniform slicing interpolation zoom and subset slicing to generate slices from 8 to 64. With the aid of the Ensembled Volumetric ConvNet, the model achieves an impressive accuracy of 93.01% for AD vs. NC vs. MCI. Notably, the 3D-subject-based method, a pioneering approach, yields the highest accuracy, 93.01%, surpassing the Patch-based (89.55%) and Slice-based (89.37%) methods. Using T1-weighted structural MRI images of the brain from the AlzhAlzheimer'sease Neuroimaging Initiative database, the authors in Francis and Pandian (2023) present an algorithm that integrates the last layers of pre-trained models Xception, Inception V3 and MobileNet for the AD and related cognitive states classification. The algorithm is tested with a multi-class classification problem, and the accuracy obtained is about 85%. It provides specific accuracies of 85% for distinguishing Mild Cognitive Impairment convertible (MCI) from Mild Cognitive Impairment non-convertible (MCInc), 94% for classifying AD from cognitively normal (CN), and 92% for differentiating MCIc from CN. The results demonstrate that the proposed algorithm surpasses other state-of-the-art methods in multi-class classification and in differentiating MCIc from MCInc.

Adaobi et al. (2023) employed a fine hybrid of Xception and Fractalnet-based deep learning techniques for the classification of the phases of AD into five stages. MRI images were drawn from the ADNI dataset to enhance the performance of the model, and an attempt was made to utilize appropriate pre-processing techniques together with segmentation procedures based on Unet++ algorithms. Recall, precision, and accuracy are established as the evaluation metrics of the performance of the proposed approach. These results of the investigation indicate that the proposed technique can achieve a level of accuracy of 98.30% recall, 99.72% precision and 99.6% accuracy in multi-class classification. To summarize, the findings point to the fact that the presented methods, when integrated with MRI images, can be useful in the classification and prediction of neurodegenerative diseases, such as AD. Venkatasubramanian et al. (2023) trained a deep learning model for the segmentation and automatic categorization of AD using structural MRI data. They adopted MTDL for the joint segmentation of the hippocampus in the given images, a

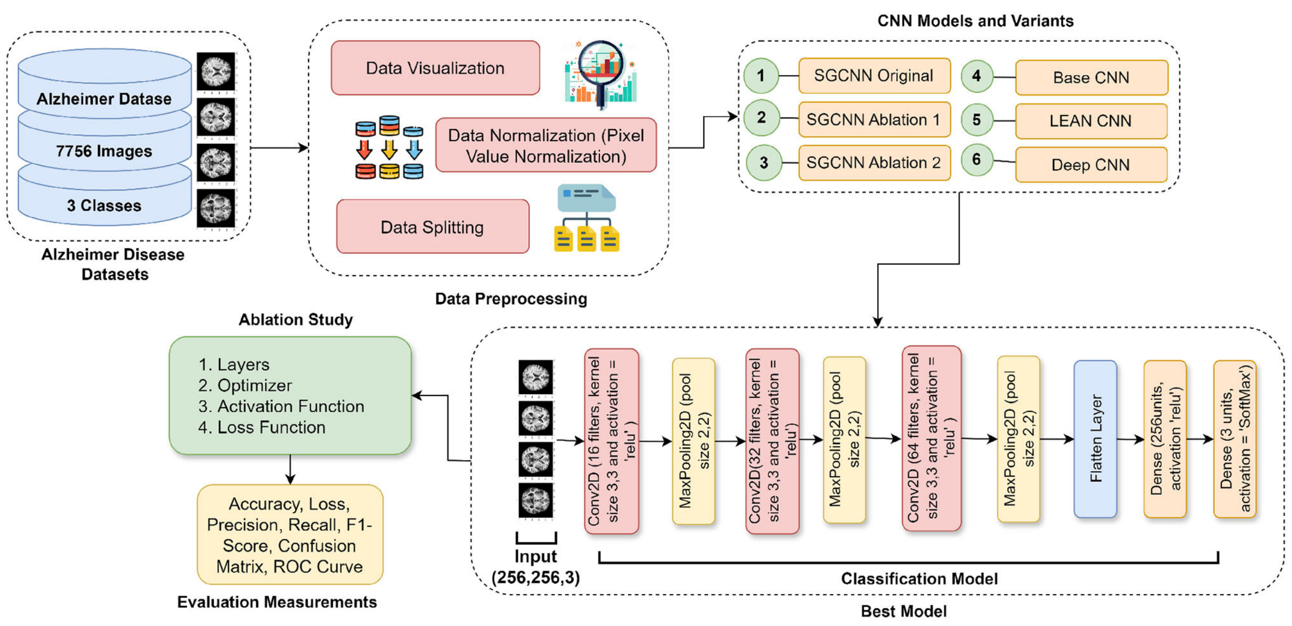


FIGURE 1
Proposed framework for Alzheimer's disease detection.

comprehensive approach. The deer hunting optimization (DHO) is then used to fine-tune the CNN model (capsule network) for the categorization of disease, guaranteeing a strong and trustworthy classification procedure. The typical method has been applied to ADNI-standardized MRI datasets, and it is effective, as suggested above. It is discovered that the proposed MTDL achieves 97.1% accuracy and 93.5% of the Dice coefficient. In comparison, the suggested MTDL model achieved a 96% accuracy for binary classification and a 93% accuracy for multi-class classification. These thorough evaluation results instill confidence in the validity and reliability of the proposed technique.

3 Proposed framework

The suggested methodology for utilizing deep learning models to identify AD is described in this section. In the classification of AD, Figure 1 presents a holistic view of deep learning models. The process starts with an experimental dataset of 7,756 images belonging to three categories. Data pre-processing covers visualization of data, normalization and data split. Next, the framework discusses model selection; SGCNN is compared to base CNN, Lean CNN, and deep CNN, as well as several ablation variants of SGCNN. The best-identified model is the sequential model of convolution, which is obtained by using a Conv2D layer of 16 filters followed by a series of layers of 32 and 64 filters, respectively, of a maximum pooling layer, then a flattening layer and two dense layers. Lastly, there is one final dense layer of three neurons with softmax activation to spit out the predictions for the three classes. The ablation study aims to determine the sensitivity of the model to hyperparameters, including loss function, learning rate, batch size, optimizer and activation function. In the last section, the results and analysis are presented with reference to the Receiver Operating Characteristic (ROC) curve, accuracy, loss, precision, recall, F1 score, confusion matrix, and accuracy. This elaborate work is meant to ensure the best deep learning model and hyperparameters that enable accurate classification of Alzheimer's disorders.

3.1 Experimental dataset

This paper focuses on categorizing participants into three groups: Alzheimer's disease (AD), Parkinson's disease (PD), and CONTROL. As a reference point for comparison, CONTROL stands for healthy people free of neurological conditions. Subjects with AD, a neurological illness depicted by a decline in cognition and memory, are included in the class. The participants in the PD class have been diagnosed with Parkinson's disease, which is typified by stiffness and tremors in the muscles. Data from clinical examinations, medical imaging, and other modalities that represent the neurological and physiological aspects of these illnesses are probably included in the dataset.

Two directories—training images and testing images—are included in the collection. In this study, we make a new directory to hold the combined photographs.

3.2 Data pre-processing

Preprocessing data is crucial for deep learning and data assessment systems. The data needs to be cleaned and altered to prepare it for additional analysis or training of deep learning models. For data preprocessing, this study used data visualization

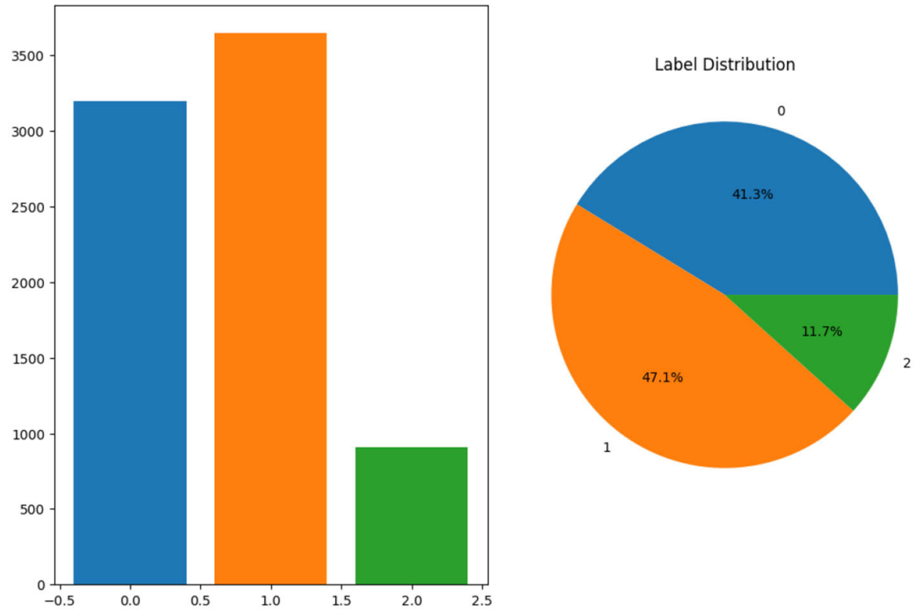


FIGURE 2
Label distribution.

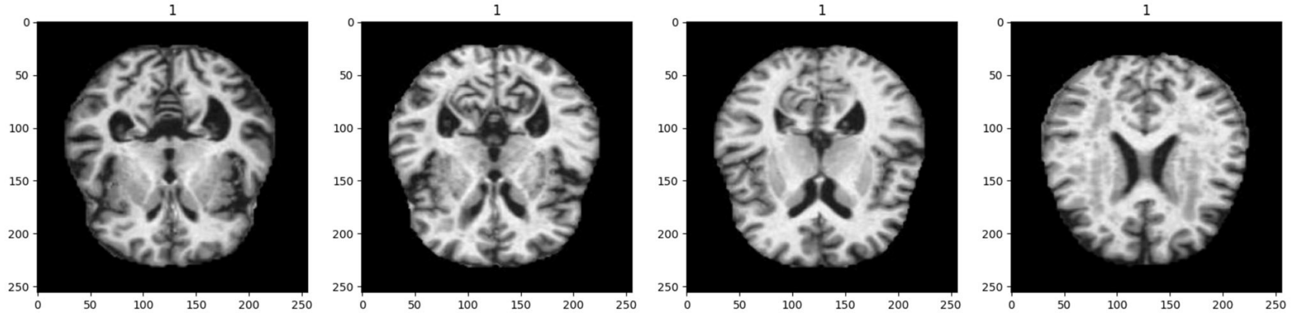


FIGURE 3
Sample MRI images.

and normalization, both of which enhanced the effectiveness of the suggested approaches.

Figure 2 shows the proportion of each label in the pie chart and a bar chart to demonstrate the distribution of labels. The bar chart on the left displays the frequency of each label, showing that label 1 has the highest count, with ∼3,500 samples, followed by label 0, with around 3,000 samples, and label 2, which has the smallest count of roughly 1,000 samples. On the right, the pie chart provides a proportional breakdown of the dataset. It shows that 41.3% of the data belongs to label 0, 47.1% to label 1, and the remaining 11.7% to label 2. Overall, label 1 dominates the dataset, while label 2 represents the least frequent category. Images from the dataset batch are shown in Figure 3, which gives a visual representation of the dataset.

An essential first step in getting data ready for deep learning models is normalization. A normalized function defined in the script accepts a picture x as input and its label y. The maximum pixel value in the image is determined by x_{max}. Function. By dividing the image by x_{max}, it normalizes it and guarantees that the values of pixels are scaled within the range of 0 and 1. Enhancing convergence rates and averting problems like gradient vanishing aids in the stabilization of the training process. After using the map method to apply this normalized function to the original dataset, the normalized dataset is created and saved in the variable data, ready to be fed into the model. Original Before normalization, the data range in the batch was from 0.0 to 254.42578. This represents the original pixel intensity values in the image data, where the maximum pixel value was close to 255, typical for 8-bit grayscale images. After applying the normalization function, the data range was scaled between 0.0 and 1.0. This was done by dividing each value of the pixel by the highest value in the batch, effectively normalizing the image data to a common scale suitable for neural network input. Data is separated into testing, validation, and training sets following data normalization to make sure the model is trained, validated, and tested on various subsets of the data. This segment is essential for assessing the model's functionality

TABLE 2 Comparison of original and normalized data batches: data size and range.

Batch	Data size	Data range (min−max)
Original batch	Varies, e.g., (batch_size, height, width)	0.0–254.42578
Normalized batch	Same as original	0.0–1.0

and generalization capacity. The training set is extracted using the take technique, which takes a predetermined piece of the dataset—usually the largest portion—while the testing and validation sets are extracted using the skip approach, which removes these training samples. Typically, 80% of these sets are used for training, 10% are used for validation, and 10% are used for testing. Now, the training size is 194, and the Validation size and test size are 24. Table 2 provides the comparison of original and normalized data sizes and ranges.

Algorithm 1 depicts the workflow that is followed in order to pre-process the data. The first step is the input of the dataset, which itself consists of training and testing images. It then makes a new directory for merged images and visualizes the dataset in order to see the distribution of labels and example images. By dividing each image by its greatest value, it first scales the pixel values of the image in the range of 0–1. Next, the dataset is divided into the following ratios: 80:10:10 for the training, validation, and test sets. Finally, the pre-processed dataset is presented in a form that is suitable for training a model.

```
Input: "Alzheimer dataset"
Output: Preprocessed dataset ready for model training
Step 1: Load Dataset
Load the dataset: training images $D_{\mathtt{train\_dir}}$ and testing images $D_{\mathtt{test\_dir}}$.
$D = \{D_{\mathtt{train\_dir}}, D_{\mathtt{test\_dir}}\}$
Step 2: Create New Directory
Create a new directory $D_{\mathtt{merged}}$ for storing merged images.
$D_{\mathtt{merged}} \leftarrow$ New Directory
Step 3: Visualize Dataset
Create figures to understand the distribution of labels and visualize sample images.
$V(D) \rightarrow$ Visualized images and label distribution
Step 4: Normalize Data
Define a normalization function to scale the values of image pixels within the range of 0 and 1.
For each image $x$ in the $D_{\mathtt{merged}}$:
$x_{\mathtt{normalized}} = \frac{x}{\max(x)}$
Apply the normalization function to the $D_{\mathtt{merged}}$:
$D_{\mathtt{normalized}} = N(D_{\mathtt{merged}})$
Step 5: $D_{\mathtt{normalized}}$
Assign training, validation, and testing sets to the normalized data.
Split the dataset as follows:
$D_{\mathtt{train}} = 0.80 \times D_{\mathtt{normalized}}$
$D_{\mathtt{val}} = 0.10 \times D_{\mathtt{normalized}}$
$D_{\mathtt{test}} = 0.10 \times D_{\mathtt{normalized}}$
Step 6: Output
Return the pre-processed dataset:
$D_{\mathtt{preprocessed}} = \{D_{\mathtt{train}}, D_{\mathtt{val}}, D_{\mathtt{test}}\}$
```

Algorithm 1. Experimental dataset and pre-processing steps.

3.3 Deep learning model

Deep learning models are sophisticated neural networks made to recognize and extract information from large, complex datasets automatically. These models are very effective for tasks like picture classification, audio recognition, and natural language processing because they are composed of numerous layers, each of which processes data to record increasingly abstract representations. This study conducted an ablation study on many deep learning models, including the Base CNN Model, LEAN CNN Model, Deep CNN Model, and SGCNN Original Model.

3.3.1 SGCNN original model

An effective deep learning model for graph-structured data is the Spectral Graph Convolutional Neural Network (SGCNN), which can identify intricate patterns and connections in non-Euclidean structures such as molecular graphs and social networks. Unlike conventional CNNs, SGCNNs use spectral-domain convolutional processes, which makes them useful for tasks like graph and node classification. Using the best features of both architectures, the hybrid deep learning model combines a segmentation and a classification model. Utilizing a U-Net architecture, the segmentation model processes images through convolutional layers activated by ReLU after first utilizing an input layer for 256 × 256 RGB images. Max pooling is then utilized to minimize the spatial dimensions of the processed images. The last convolutional layer creates the segmentation mask, while an upsampling layer recovers the image size.

A CNN is used in combination with this classification model to classify images such as PD, AD, and CONTROL. Several convolutional and max pooling layers are added after the input layer in order to extract features. The final softmax layer for classification is reached after the feature maps have been flattened and dense layers with ReLU activation have captured complex patterns. The hybrid model incorporates the classification model with the segmentation output of the U-Net network. With the use of precise spatial data, this method improves classification accuracy while enabling the model to execute segmentation and classification tasks. Overall performance in identifying the input images is improved by the combined model's excellent integration of segmentation characteristics. Figure 4 visualizes the architecture of the original SGCNN model.

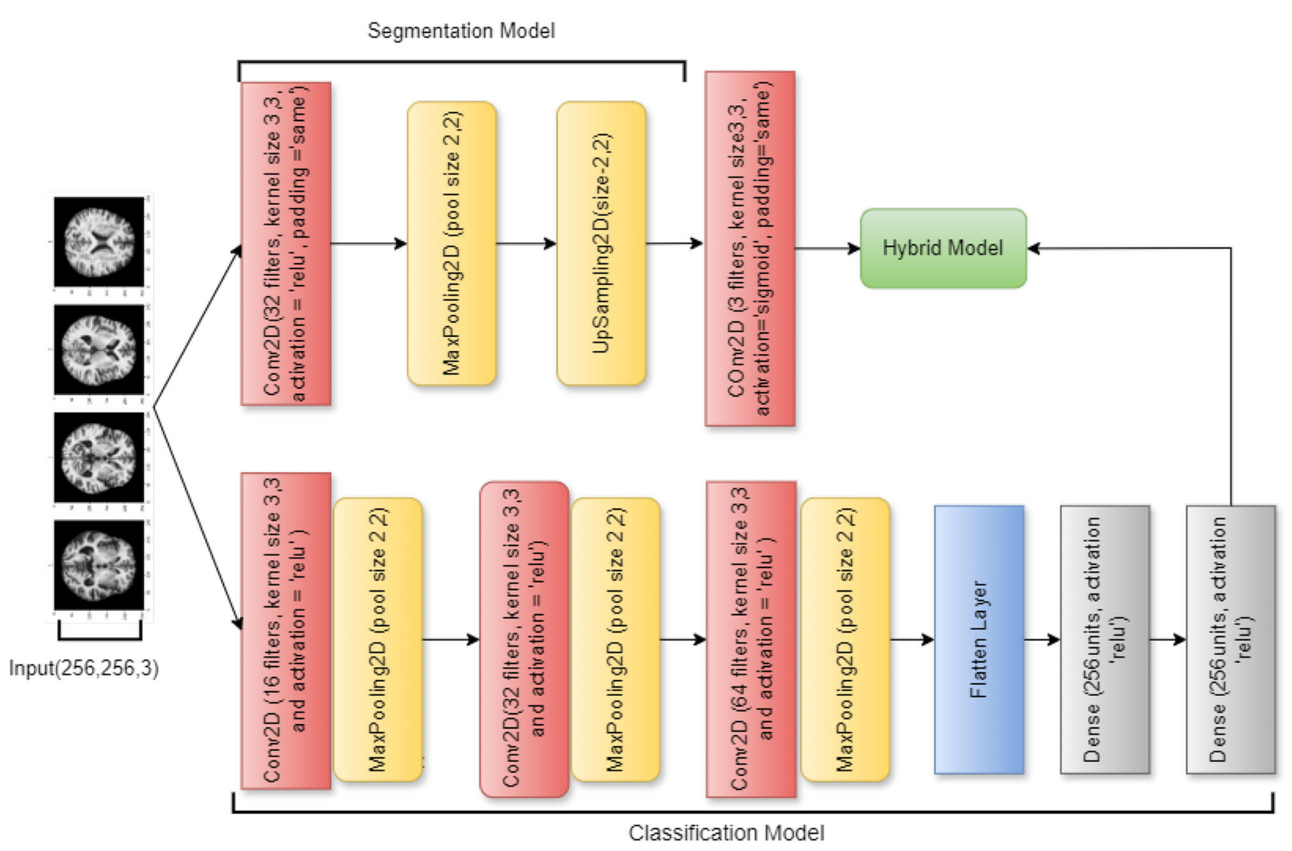


FIGURE 4
SGCNN model architecture.

3.3.2 Ablation of SGCNN model 1

An ablation study of the SGCNN model is the second model, with the primary goal of focusing only on classification problems and streamlining the architecture by eliminating the segmentation component. This model simplifies the network into a more conventional CNN by removing the segmentation layers while keeping a topology resembling that of the original SGCNN. It starts with several convolutional layers and moves on to pooling layers and dense layers for classification. An input layer for 256 × 256 RGB images is the first layer in the design. Three convolutional layers with progressively larger filter sizes (16, 32, and 64) come next. To lower the spatial dimensions, a max-pooling layer is paired with each convolutional layer.

The elimination of the segmentation network, which was a feature of the SGCNN, is the most notable modification to this model. Because of the network's amplification brought forth by this ablation, the performance of the classification component can be examined more closely. The model is simpler now that the segmentation layers have been eliminated, and it only concentrates on classifying the input images into three groups (e.g., CONTROL, AD, and PD). Like the original model, the model is compiled using the Adam optimizer with category cross-entropy loss. It is trained with a batch size of 32 across 15 epochs. Figure 5 visualizes the architecture of the ablation of the SGCNN model 1.

3.3.3 Ablation of SGCNN model 2

The architecture of this third model, which is an additional ablation study of the SGCNN model, is largely unchanged from the earlier iterations. However, there are a few significant changes. The model starts with a modified segmentation model that has an upsampling layer to boost spatial dimensions and a 32 filters-convolutional layer. In contrast to the original segmentation model, this variant generates the segmentation output in the last layer using a sigmoid activation function.

In contrast to the earlier models, the classification model component adds a 64-filter convolutional layer and removes the batch normalization and dropout layers from the dense layer. With these modifications, the classification model becomes more simplified and produces the classification output with softmax activation by connecting the flattened feature maps straight to the final dense layer. The design is put together utilizing the Adam optimizer and category cross-entropy loss, the same as the earlier models. The metrics used to evaluate the model's performance are test accuracy, validation, and training. Figure 6 visualizes the architecture of the ablation of the SGCNN model 2.

3.3.4 Base CNN model

The purpose of this base convolutional neural network (CNN) model is to categorize images. The design is simple, with layers

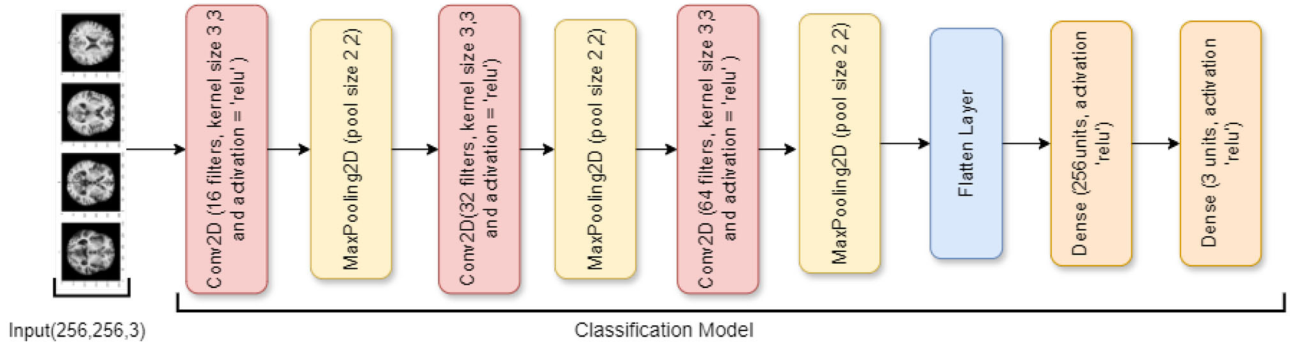


FIGURE 5
Ablation of SGCNN model 1 architecture.

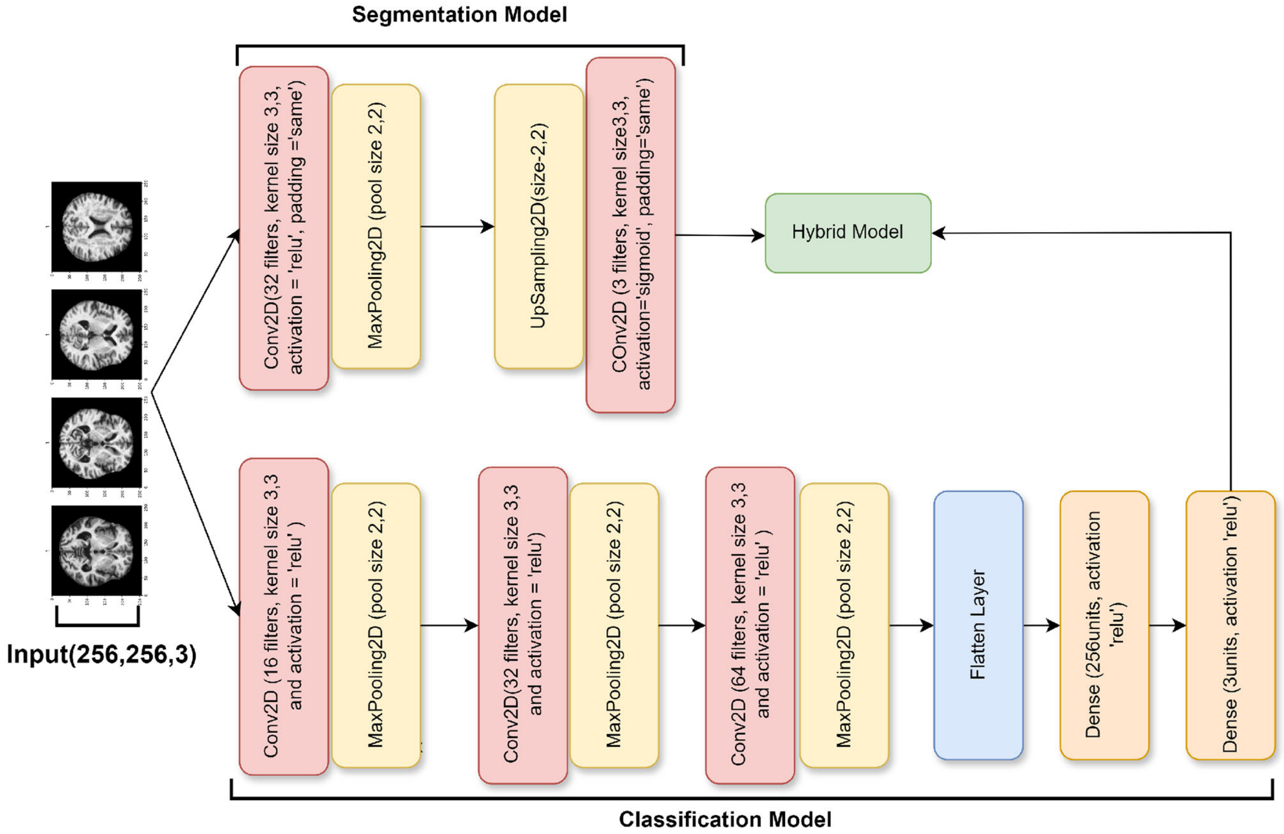


FIGURE 6
Ablation of SGCNN model 2 architecture.

processing and classifying input photos in a stack order. The model begins with a convolutional layer that applies 16 3 × 3 filters to the 256 × 256 input images, each of which has three RGB color channels. This layer introduces non-linearity using the ReLU activation function. Then, by normalizing the convolutional layer's output, batch normalization stabilizes the training process. Subsequently, the model comprises a max-pooling layer that reduces the computational complexity and concentrates on the most prominent characteristics by downsampling the feature maps' spatial dimensions by a factor of two.

The model then incorporates a dropout layer, which randomly removes 25% of the neurons during training in order to avoid overfitting. In order to prepare it for the thick layers that come next, the output is then flattened into a 1D feature vector. Using the ReLU activation function once more, the dense layers begin with a completely connected layer comprising 32 neurons. Batch normalization and an additional dropout layer—which removes 50% of the neurons—come after this layer to further lessen the possibility of overfitting. The classification output is produced by an output layer in the model's stages, which employs the softmax

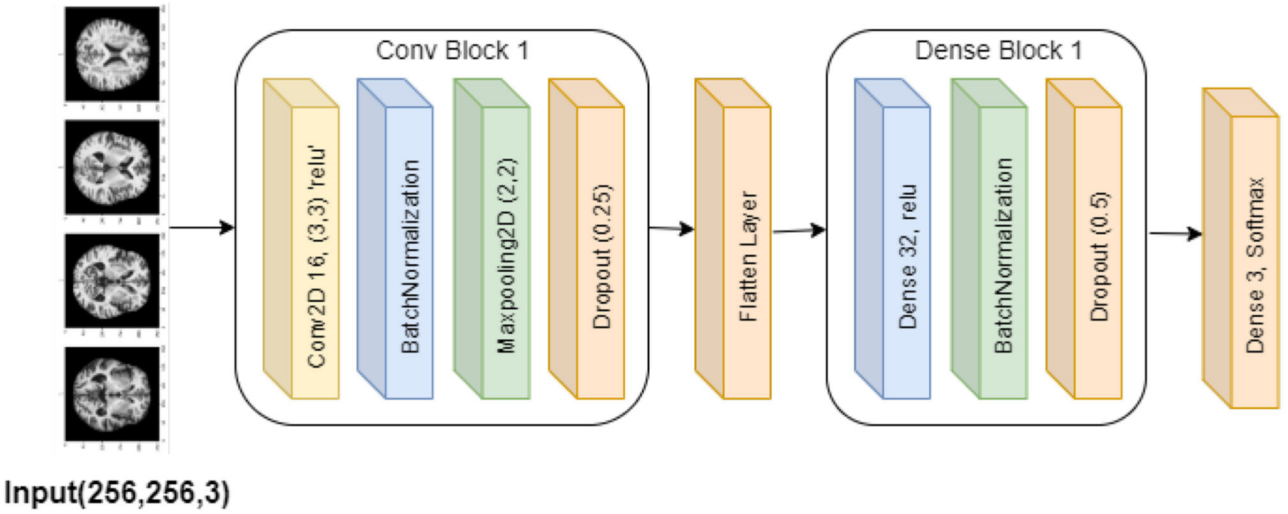


FIGURE 7
Base CNN model architecture.

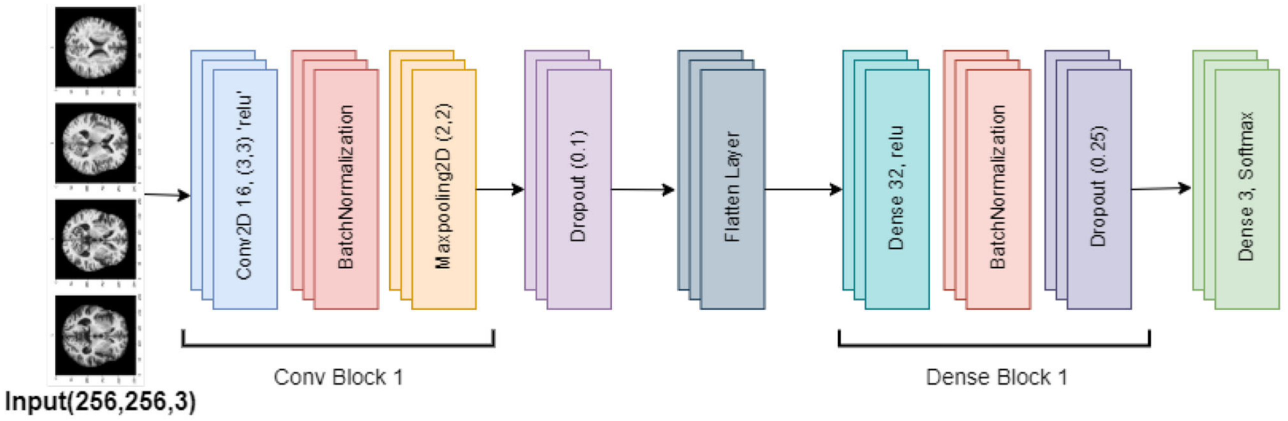


FIGURE 8
Lean CNN model architecture.

activation function to generate a probability distribution over three classes. The model is constructed using the standard techniques for classification tasks: the Adam optimizer and categorical cross-entropy loss. Throughout the training, the model's accuracy is monitored to make sure it picks up the picture classification skills correctly. Figure 7 visualizes the Base CNN model's architecture.

3.3.5 LEAN CNN model

The Lean CNN model is based on the Base CNN model, which incorporates adjustments to minimize overfitting. It introduces particular adjustments to dropout rates while maintaining a similar design. This model is composed of a stack of successive layers, starting with a convolutional layer that processes 256 × 256 × 3 (RGB) input pictures using 16 filters of size 3 × 3. The convolutional layer also uses ReLU activation. Next, the training is stabilized using batch normalization and the feature map dimensions are minimized using max pooling. Lower dropout rates are the primary change made to the Lean CNN model. In particular, compared to 0.25 and 0.5 in the Base CNN model, the dropout rate is reduced to 0.1 in the first dropout layer and 0.25 in the second dropout layer. By lowering the chance of overfitting, this modification is intended to lessen the degree of regularization, which could enhance the model's performance. Figure 8 visualizes the architecture of the Lean CNN model.

The feature maps are processed via dense layers after being flattened into a 1D vector. After the 32 units of dense layer with ReLU activation, there is a 0.25 dropout rate and batch normalization. The model's final layer uses softmax activation to create a dense output layer with three neurons that produce the classification probabilities. Overall, the lower dropout rates indicate a deliberate change meant to improve model performance by reducing overfitting, even if the Lean CNN retains the Base CNN's fundamental architecture.

Algorithm 2 explains the design of some SGCNN and CNN models for image classification, which are as follows. The algorithm initializes a sequential model and defines five different models, Including SGCNN Model 1, SGCNN Model 2, SGCNN Model 3, Base CNN Model, Lean CNN Model, and Deep CNN Model. Each model has six layers of convolution, batch normalization, max pooling, dropout, density layers, and one output layer that uses softmax in the form of probabilistic distribution. One model has fewer filters and smaller kernel sizes and uses a dropout rate of 0.2. In comparison, the second model has more filters, larger kernel sizes and a dropout rate of 0.3. The algorithm also defines the base layers applicable to all models such as a flattening layer, dense layers as well as the output layers. The model is then trained using the training data and verified using the validation set. Finally, it is stated that the prior model was assembled using the Adam optimizer and categorical crossentropy loss function.

3.3.6 Deep CNN model

The Base and Lean CNN models are built upon the Deep CNN model, which increases the architecture's complexity by using more layers and units. More convolutional layers in this model improve its capacity to extract fine-grained characteristics from the input images, which makes it more appropriate for challenging classification tasks. Two extra convolutional layers, each with 32 filters, are added to the original 16-filter convolutional layer in the Deep CNN model. These additional layers let the model recognize more complex patterns and enable it to extract deeper characteristics from the incoming data. Each convolutional layer is followed, like in the earlier models, by batch normalization and max pooling, which downsamples the feature maps.

Additionally, the Base and Lean CNN models' units in the dense layer are increased to 64 units in the Deep CNN model. With more neurons, the model can process the bigger feature set generated by the further convolutional layers. With a 25% dropout after the convolutional layers and a 50% dropout after the dense layer to avoid overfitting, the dropout rates are still in line with the earlier models. Overall, the extra convolutional layers and the larger dense layer distinguish the Deep CNN model from the Base and Lean CNN models. By extracting more precise features from the data, these improvements should increase the model's accuracy in classifying photos. Figure 9 visualizes the Deep CNN model's architecture.

3.4 Ablation study

In this study, ablation analysis was conducted to analyze the effect of structural modifications on the performance of the SGCNN architecture for Alzheimer's image classification. The SGCNN Original Model served as the baseline, achieving a classification accuracy of 93%. Two variants of the SGCNN model were developed to assess the effects of different structural changes: Ablation of SGCNN Model 1 and Ablation of SGCNN Model 2. The first variant, Ablation of SGCNN Model 1, incorporated specific architectural adjustments that led to a notable improvement in classification accuracy, reaching 95%. This significant enhancement

```
Initialize $S_{model}$ ← Sequential Model
SGCNN Model 1:
  $S_{\text{model}}$ ← Sequential Model {Initialize a Sequential model}
  $A_1$ ← Conv2D($F = 8, K = (3 \times 3)$, activation = ReLU, input_shape = $(256, 256, 3)$) {First convolutional layer with 8 filters}
  $A_2$ ← BatchNormalization {Apply batch normalization}
  $A_3$ ← MaxPooling2D($P = (2 \times 2)$) {Max pooling layer to reduce spatial dimensions}
  $A_4$ ← Dropout($D = 0.2$) {Dropout layer to prevent overfitting}
SGCNN Model 2 (Second Ablation Study):
  $A_5$ ← Conv2D($F = 8, K = (3 \times 3)$, activation = ReLU) {Convolutional layer with 8 filters}
  $A_6$ ← Conv2D($F = 8, K = (3 \times 3)$, activation = ReLU) {Additional convolutional layer}
  $A_7$ ← BatchNormalization {Batch normalization for stability}
  $A_8$ ← MaxPooling2D($P = (2 \times 2)$) {Pooling layer for down-sampling}
  $D_{\text{adjust}}$ ← $D$ (adjustable rate) {Adjustable dropout rate}
SGCNN Model 3 (Third Ablation Study):
  $A_9$ ← Conv2D($F = 8, K = (3 \times 3)$, activation = ReLU) {Convolutional layer for feature extraction}
  $A_{10}$ ← BatchNormalization {Batch normalization for improved convergence}
  $A_{11}$ ← MaxPooling2D($P = (2 \times 2)$) {Max pooling to reduce dimensionality}
  Simplified architecture focused on classification to reduce overfitting {Further architecture simplification}
Base CNN Model:
  $A_{12}$ ← Conv2D($F = 16, K = (3 \times 3)$, activation = ReLU, input_shape = $(256, 256, 3)$) {Base model with 16 filters}
  $A_{13}$ ← BatchNormalization {Normalize the output of the previous layer}
  $A_{14}$ ← MaxPooling2D($P = (2 \times 2)$) {Pooling layer for down-sampling}
  $A_{15}$ ← Dropout($D = 0.25$) {Dropout to reduce overfitting}
Lean CNN Model (Variant of Base CNN):
  $D_1$ ← Dropout($D = 0.1$) {First Dropout layer with reduced rate}
  $D_2$ ← Dropout($D = 0.25$) {Second Dropout layer with adjustable rate}
Deep CNN Model (Extension of Base CNN):
  $A_{16}$ ← Conv2D($F = 32, K = (3 \times 3)$, activation = ReLU) {Extended model with more filters}
  $A_{17}$ ← BatchNormalization {Normalize to stabilize training}
  $A_{18}$ ← MaxPooling2D($P = (2 \times 2)$) {Pooling for feature reduction}
  $U_1$ ← 64 {Dense layer units increased to 64}
```

```
Common Layers for All Models:
$A_{flatten} \leftarrow$ Flatten
$U_{common} \leftarrow$ Dense($U = 32$ or $64$)
$A_{19} \leftarrow$ BatchNormalization
$D_{common} \leftarrow$ Dropout (model-specific rates)
$A_{output} \leftarrow$ Dense($U = 3$, activation $= Softmax$)
```

Algorithm 2. SGCNN and CNN model variants.

demonstrates the effectiveness of these modifications in refining the model's capacity to distinguish between AD stages.

Conversely, the Ablation of SGCNN Model 2 did not exhibit an improvement over the baseline, maintaining the same 93% accuracy as the original model. This finding emphasizes how crucial it is to choose suitable structural changes to achieve performance gains. The ablation study underscores the potential of targeted architectural adjustments in optimizing model accuracy for medical image classification tasks. By isolating and analyzing these modifications, the study provides valuable insights into effective strategies for enhancing diagnostic tools for AD, offering promising avenues for further investigations and advancements in this critical field of medicine. Among the models, the Ablation of SGCNN Model 1 performed the best, achieving the highest accuracy of 95%.

4 Experimental result and discussion

This section provides the evaluation measurements used in this study and the experimental results of all the models.

4.1 Evaluation measurements

The effectiveness of the suggested methodology is calculated in this study utilizing a variety of evaluation measures, including F1-score, recall, accuracy, and precision. These crucial assessment metrics offer comprehensive details regarding how the suggested technique should be interpreted. The first metric that is frequently seen as the foundation of performance evaluation is accuracy. By considering the total number of instances, the percentage of correctly detected outcomes is calculated using the accuracy metric, which is defined as the ratio of correctly predicted positive and negative cases (true positives and true negatives) to the total number of instances, including false positives and false negatives as shown in Equation 1.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Precision is important in situations where the cost of false positives is high, as it measures the model's ability to avoid incorrectly predicting negative instances as positive. However, precision does not account for how many actual positive instances the model missed. Equation 2 explains the precision. Recall is especially important in scenarios where missing positive instances (false negatives) has serious consequences, such as in medical diagnoses. A high recall indicates that the model captures most of the positive instances, but it may come at the expense of higher false positives. Equation 3 defines the precision. To balance the trade-offs between precision and recall, the F1-score is used. The F1-score is the harmonic mean of precision and recall, ensuring that both are taken into account.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - score = 2 \times \frac{Precision + Recall}{Precision + Recall} \tag{4}$$

The F1-score is particularly useful when the dataset is imbalanced and when both false positives and false negatives are important to consider. It provides a single metric that captures a balance between precision and recall, allowing for more informed model performance evaluations. Equation 4 demonstrated its computation. A classification model's performance can be categorized and assessed using a confusion matrix, which provides a list of counts for true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). It offers details about the model's capacity for learning and differentiating between classes. While genuine positives and true negatives demonstrate appropriate classifications, false positives and false negatives highlight instances in which the model misclassifies. Identifying specific types of errors the model makes and guiding modifications to enhance its performance are made possible in large part by this matrix. A binary classification model's performance at different thresholds is represented graphically by a curve known as a Receiver Operating Characteristic (ROC). True positive rate (TPR) against false positive rate (FPR) plotting displays the proportion of correctly labeled positive instances on the y-axis and the proportion of falsely identified positive cases on the x-axis. Curves further from the diagonal line indicate better model performance, which is the representation of plotting random guesses. The left-hand corner of the plot denotes improved performance and accuracy. Increases in the model's performance are indicated by higher values of the area under the curve (AUC).

4.2 Result and findings

Table 3 shows the classification report of several pre-trained CNNs, each one evaluated with a different configuration on a dataset. In the table, values in terms of precision, recall, F1-score, support for each class and the weighted average over all classes (Wei. Avg) are provided. The SGCNN Original Model shows strong performance, particularly in Class 1, with a precision value of 0.99 and a recall value of 0.99, resulting in an F1-score value of 0.99 with support of 355 instances. Class 2 exhibits a slightly lower performance, with a precision value of 0.95, a recall value of 0.91, and an F1-score value of 0.93, supported by 343 instances. The

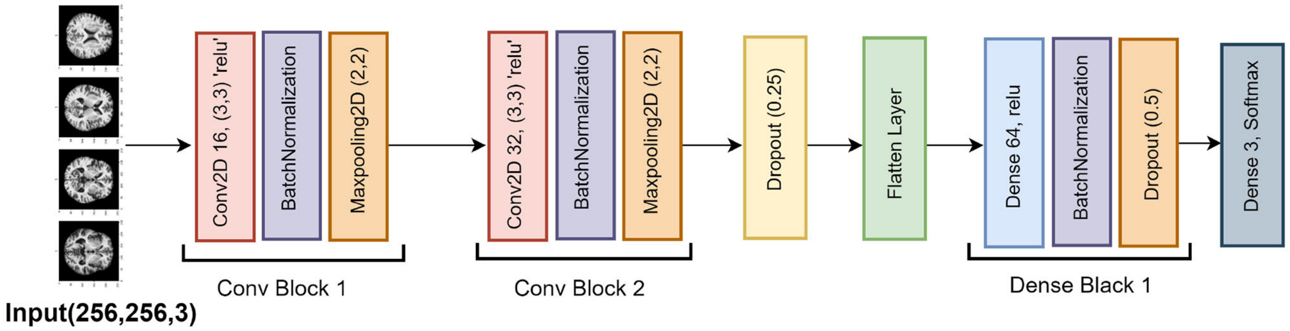


FIGURE 9
Deep CNN model architecture.

model struggles more with Class 3, achieving a precision value of 0.68, a recall value of 0.83, and an F1-score value of 0.75, with a support of 70 instances. The precision, recall, and F1-score weighted averages are 0.95, 0.94, and 0.94, respectively.

The SGCNN Model 1 performs excellently in Class 1, with a precision value of 0.99, a recall value of 1.00, and an F1-score value of 0.99, supported by 346 instances. In Class 2, it achieves a precision value of 0.96, a recall value of 0.93, and an F1-score value of 0.94, supported by 342 instances. Class 3 is handled well, with a precision value of 0.77, a recall value of 0.85, and an F1-score value of 0.81, with support of 80 instances. This model has a weighted average precision, recall, and F1-score of 0.95 for all measures. The SGCNN Model 2 shows robust performance in Class 1, with a precision value of 0.95, a recall value of 0.99, and an F1-score value of 0.97, supported by 343 instances. However, in Class 2, there is a slight drop with a precision value of 0.95, a recall value of 0.88, and an F1-score value of 0.91, with a support of 347 instances. Class 3 is reasonably well handled, with a precision value of 0.73, a recall value of 0.85, and an F1-score value of 0.78, with support of 78 instances. This model has a weighted average precision, recall, and F1-score of 0.93 for all measures. The Base CNN Model demonstrates excellent performance in Class 1, with a precision value of 0.99, a recall value of 0.99, and an F1-score value of 0.99, supported by 346 instances. Class 2 exhibits good performance with a precision value of 0.94, a recall value of 0.89, and an F1-score value of 0.91, supported by 341 instances. The model performs less effectively in Class 3, with a precision value of 0.67, a recall value of 0.81, and an F1-score value of 0.73, supported by 81 instances. This model has a weighted average precision, recall, and F1-score of 0.93 for all measures. The Lean CNN Model maintains strong performance in Class 1, with a precision value of 0.98, a recall value of 1.00, and an F1-score value of 0.99, supported by 361 instances. Class 2 shows a good performance, with a precision value of 0.96, a recall value of 0.90, and an F1-score value of 0.93, with support of 335 instances. The model achieves a precision value of 0.70, a recall value of 0.81, and an F1-score value of 0.75 in Class 3, supported by 72 instances. This model has a weighted average precision, recall, and F1-score of 0.94 for all measures. The Deep CNN Model also shows strong results, particularly in Class 1, with a precision value of 0.99, a recall value of 0.99, and an F1-score value of 0.99, supported by 336 instances. Class 2 maintains good results, with a precision value of 0.94, a recall value of 0.93, and an F1-score value of 0.93, supported by 351 instances. For Class 3, the model achieves a precision value of 0.72, a recall value of 0.78, and an F1-score value of 0.75, supported by 81 instances. The weighted average value of precision, recall, and f1-score is 0.94. Overall, the various CNN models demonstrate strong performance in the frequently observed classes, with some variations in the less frequent Class 3, highlighting different aspects of model robustness and generalization capabilities.

4.2.1 Result of SGCNN original model

Figure 10 represented the graphical representation of the SGCNN Orginal Model. Figure 10A presents the model loss and accuracy graph of an SGCNN Original Model. The training accuracy starts from 0.50% value at 0*th* epoch, and it increases up to 0.68% at 1*st* epoch. Then, it increases upward, and the training accuracy stops at 1.0% at 14*th* epoch. The validation accuracy starts from 0.67% value at 0*th* epoch. After some fluctuation of increases and decreases, testing accuracy attained is 0.93% at 14*th* epoch. The training loss starts from a 0.8 value at 0*th* epoch, and it decreases downward up to a 0.0 value at 14*th* epoch. The validation loss starts from a 0.6 value at 0*th* epoch and decreases up to 0.2 at 14*th* epoch after going through some fluctuation of increases and decreases. The confusion matrix in Figure 10B gives the percentage of each class that is correctly and incorrectly classified. The elements on the diagonal consist of the classes that have been well-classified, while the non-diagonal elements on the matrix consist of the samples that are misclassified. For class 1, the model correctly classified 99.14% instances and 0.86% of class 1 instances misclassified as class 2. For class 2, the model accurately classified 91.59% instances and 1.16% instances of class 2 incorrectly categorized as class 1 and 7.25% instances of class 2 incorrectly categorized as class 3. For class 3, the model correctly classified 78.08% instances and 21.92% of class 3 instances misclassified as class 2. The performance of the model for each class is shown by the ROC graph in Figure 10C. For different threshold settings, the graph plots the real positive rate against the false positive rate. AUC values for all three classes are very high, as can be seen from the ROC curve, indicating that the model is doing exceptionally well. Class 0 has an AUC of 1.00, class 1 of 0.98, and class 2 of 0.98. For

TABLE 3 Classification reports of experimented models.

	Labels	Precision	Recall	F1-score	Support
SGCNN original	Class 1	0.99	0.99	0.99	355
	Class 2	0.95	0.91	0.93	343
	Class 3	0.68	0.83	0.75	70
	Wei. Avg	0.95	0.94	0.94	768
SGCNN model 1	Class 1	0.99	1.00	0.99	346
	Class 2	0.96	0.93	0.94	342
	Class 3	0.77	0.85	0.81	80
	Wei. Avg	0.95	0.95	0.95	768
SGCNN model 2	Class 1	0.95	0.99	0.97	343
	Class 2	0.95	0.88	0.91	347
	Class 3	0.73	0.85	0.78	78
	Wei. Avg	0.93	0.93	0.93	768
Base CNN model	Class 1	0.99	0.99	0.99	346
	Class 2	0.94	0.89	0.91	341
	Class 3	0.67	0.81	0.73	81
	Wei. Avg	0.93	0.93	0.93	768
Lean CNN model	Class 1	0.98	1.00	0.99	361
	Class 2	0.96	0.90	0.93	335
	Class 3	0.70	0.81	0.75	72
	Wei. Avg	0.94	0.94	0.94	768
Deep CNN model	Class 1	0.99	0.99	0.99	336
	Class 2	0.94	0.93	0.93	351
	Class 3	0.72	0.78	0.75	81
	Wei. Avg	0.94	0.94	0.94	768

all three classes, this demonstrates the model's extremely high accuracy level. Classes 0, 1, and 3 have cyan, orange, and blue ROC curves, respectively. The model has a low false positive rate and a high true positive rate, as indicated by the fact that all of its ROC curves are located close to the upper left corner. This attests to the model's high classification accuracy among the three classes.

4.2.2 Result of SGCNN model 1

Figure 11 represented the graphical representation of SGCNN Model 1. Figure 11A presents the model loss and accuracy graph of an SGCNN Model 1. The training accuracy starts from 0.65% value at 0*th* epoch, and it increases up to 0.90% at 1*st* epoch. Then, it increases upward, and the training accuracy stops at 1.00% at 14*th* epoch. The validation accuracy starts from 0.77% value at 0*th* epoch. After some fluctuation of increases and decreases, testing accuracy attained is 0.95% at 14*th* epoch. The training loss starts from a 0.8 value at 0*th* epoch, and it decreases downward up to a 0.0 value at 14*th* epoch. The validation loss starts from a 0.5 value at 0*th* epoch and decreases up to 0.3 value at 14*th* epoch after going through some fluctuation of increases and decreases. The confusion matrix in Figure 11B gives the percentage of each class that is correctly and incorrectly classified. The elements on the diagonal consist of the classes that have been well-classified, while the non-diagonal elements on the matrix consist of the samples that are misclassified. For class 1, the model correctly classified 99.72% instances and 0.28% of class 1 instances misclassified as class 2. For class 2, the model accurately classified 91.67% instances and 1.16% instances of class 2 incorrectly categorized as class 1 and 7.25% instances of class 2 incorrectly categorized as class 3. For class 3, the model correctly classified 78.08% instances and 21.92% of class 3 instances misclassified as class 2. The performance of the model for each class is shown by the ROC graph in Figure 11C. For different threshold settings, the graph plots the real positive rate against the false positive rate. AUC values for all three classes are very high, as can be seen from the ROC curve, indicating that the model is doing exceptionally well. Class 0 has an AUC of 1.00, class 1 of 0.98, and class 2 of 0.98. For all three classes, this demonstrates the model's extremely high accuracy level. Classes 0, 1, and 3 have cyan, orange, and blue ROC curves, respectively. The model has a low false positive rate and a high true positive rate, as indicated by the fact that all of its ROC curves are located close to the upper left corner. This attests to the model's high classification accuracy among the three classes.

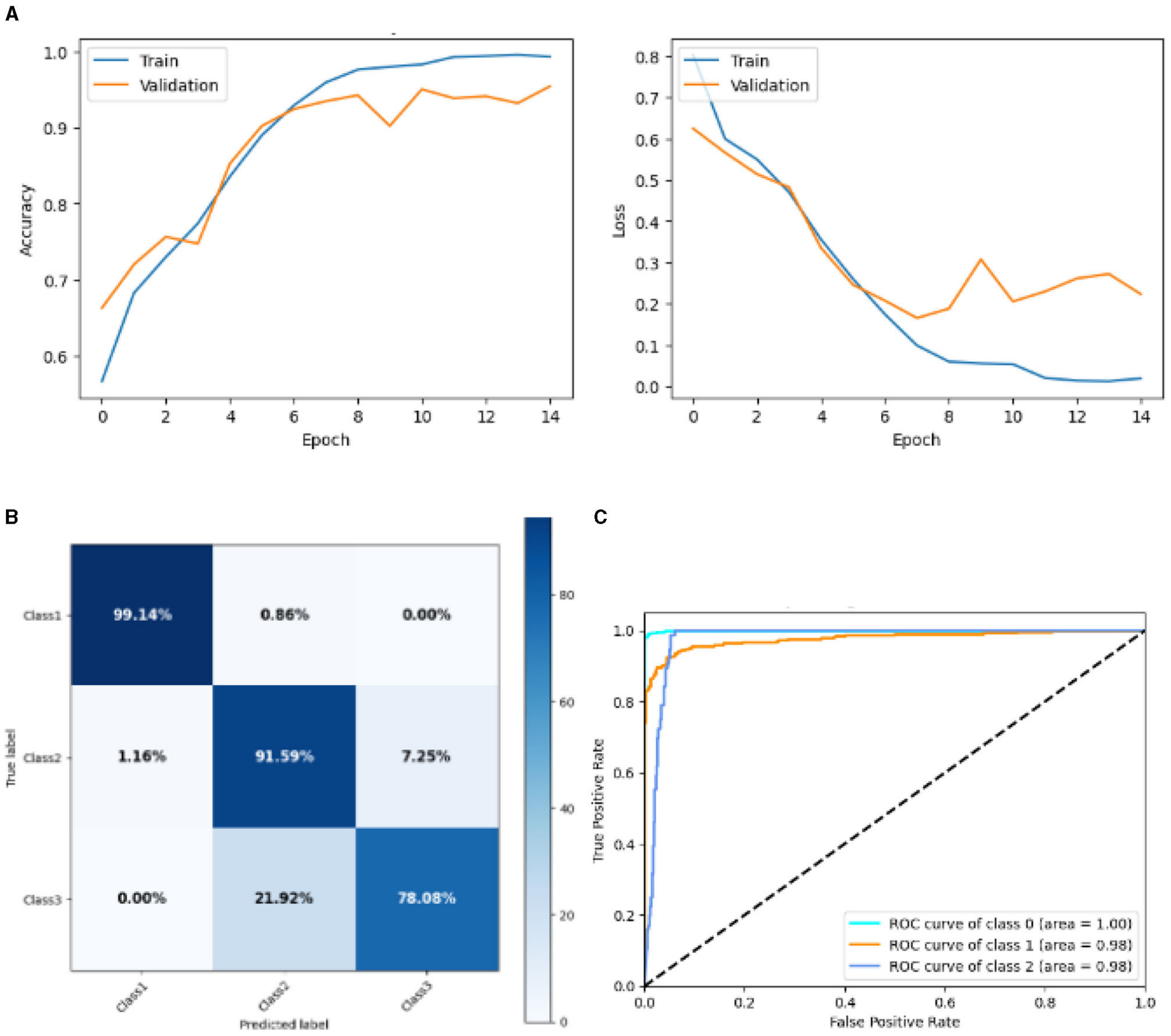


FIGURE 10
Graphical representation of SGCNN Orginal model's results. **(A)** Accuracy and loss. **(B)** Confusion matrix. **(C)** Receiver operating characteristics curve.

4.2.3 Result of SGCNN model 2

Figure 12 represented the graphical representation of the SGCNN Orginal model. Figure 12A presents the model's loss and accuracy graph of an SGCNN Model 2. The training accuracy starts from 0.5% value at 0*th* epoch. Then, it increases upward, and the training accuracy stops at 1.00% at 14*th* epoch. The validation accuracy starts from 0.67% value at 0*th* epoch. After some fluctuation of increases and decreases, testing accuracy attained is 0.95% at 14*th* epoch. The training loss starts from a 0.8 value at 0*th* epoch, and it decreases downward up to a 0.0 value at 14*th* epoch. The validation loss starts from a 0.68 value at 0*th* epoch, and it decreases up to a 0.29 value at 14*th* epoch. The confusion matrix in Figure 12B gives the percentage of each class that is correctly and incorrectly classified. The elements on the diagonal consist of the classes that have been well-classified, while the non-diagonal elements on the matrix consist of the samples that are misclassified. For class 1, the model correctly classified 99.16% instances and 0.84% of class 1 instances misclassified as class 2. For class 2, the model accurately classified 86.63% instances and 4.79% instances of class 2 incorrectly categorized as class 1 and 8.38% instances of class 2 incorrectly categorized as class 3. For class 3, the model correctly classified 85.33% instances and 14.67% of class 3 instances misclassified as class 2. The performance of the model for each class is shown by the ROC graph in Figure 12C. For different threshold settings, the graph plots the real positive rate against the false positive rate. AUC values for all three classes are very high, as can be seen from the ROC curve, indicating that the model is doing exceptionally well. Class 0 has an AUC of 1.00, class 1 of 0.98, and class 2 of 0.98. For all three classes, this demonstrates the model's extremely high accuracy level. Classes 0, 1, and 3 have cyan, orange, and blue ROC curves, respectively. The model has a low false positive rate and a high true positive rate, as indicated by the fact

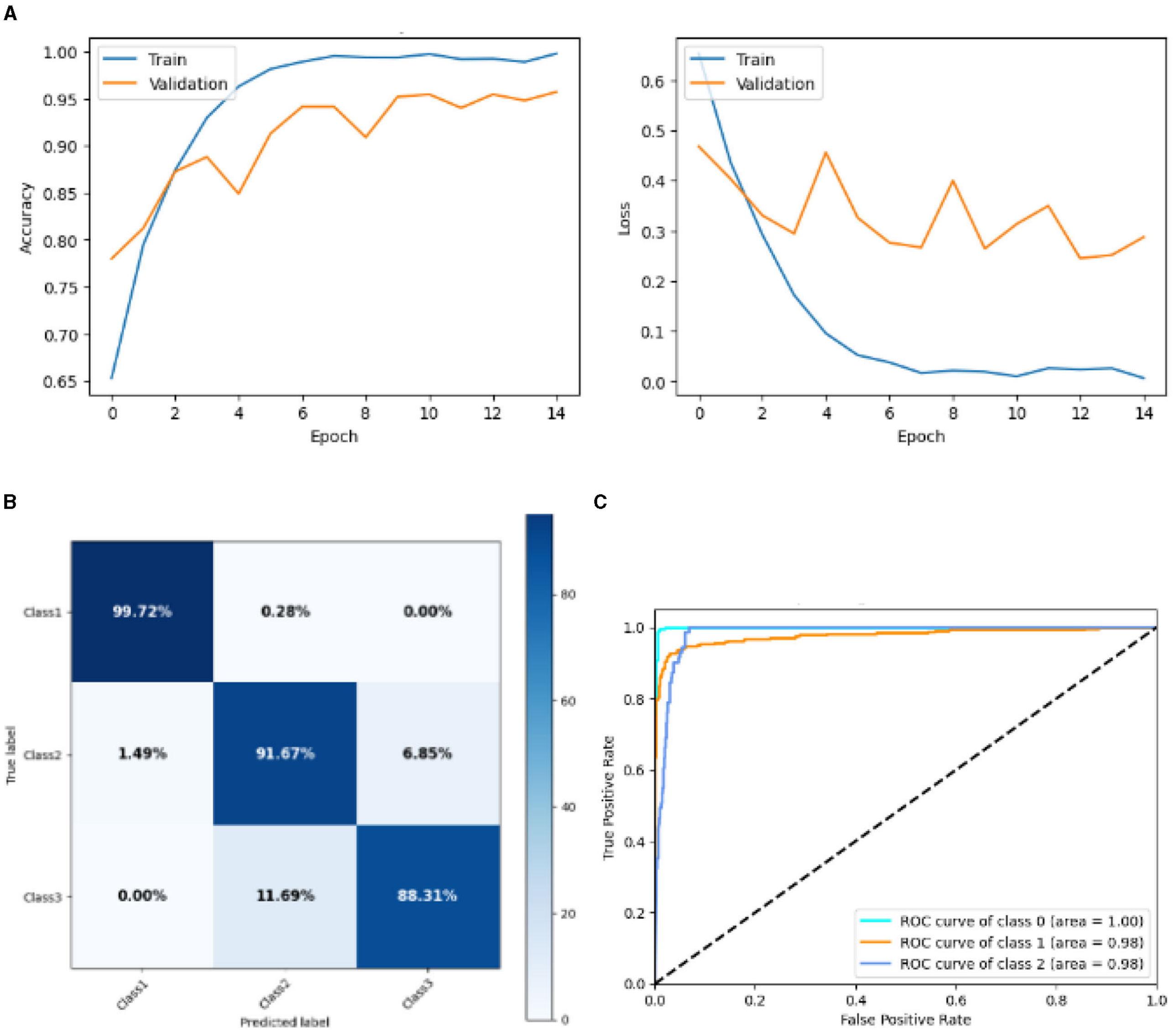


FIGURE 11
Graphical representation of SGCNN model 1's results. **(A)** Accuracy and loss. **(B)** Confusion matrix. **(C)** Receiver operating characteristics curve.

that all of its ROC curves are located close to the upper left corner. This attests to the model's high classification accuracy among the three classes.

4.2.4 Result of base CNN model

Figure 13 represented the graphical representation of the Base CNN Model. Figure 13A presents the model accuracy and loss of a Base CNN Model. The training accuracy starts from 0.62% value at 0*th* epoch. Then, it increases upward, and the training accuracy stops at 1.00% at 14*th* epoch. The validation accuracy starts from 0.53% value at 0*th* epoch, and it increases at 0.75% at 2*nd* epoch. After some fluctuation of increases and decreases, validation accuracy attained is 0.79% at 14*th* epoch. The training loss starts from a 0.9 value at 0*th* epoch, and it decreases downward up to a 0.0 value at 14*th* epoch. The validation loss starts from a 0.9 value at 0*th* epoch. It decreases up to 0.6 value at 2*nd*, and then it increases up to 1.1 value at 3*rd* epoch, and then it decreases up to 0.4 value at 4*th* then after going through some fluctuation of increases and decreases the validation loss stops at 0.8 value at 14*th* epoch. The confusion matrix in Figure 13B gives the percentage of each class that is correctly and incorrectly classified. The elements on the diagonal consist of the classes that have been well-classified, while the non-diagonal elements on the matrix consist of the samples that are misclassified. For class 1, the model correctly classified 98.02% instances and 1.98% of class 1 instances misclassified as class 2. For class 2, the model accurately classified 89.16% instances and 1.20% instances of class 2 incorrectly categorized as class 1 and 9.64% instances of class 2 incorrectly categorized as class 3. For class 3, the model correctly classified 84.15% instances and 15.85% of class 3 instances misclassified as class 2. The performance of the model for each class is shown by the ROC

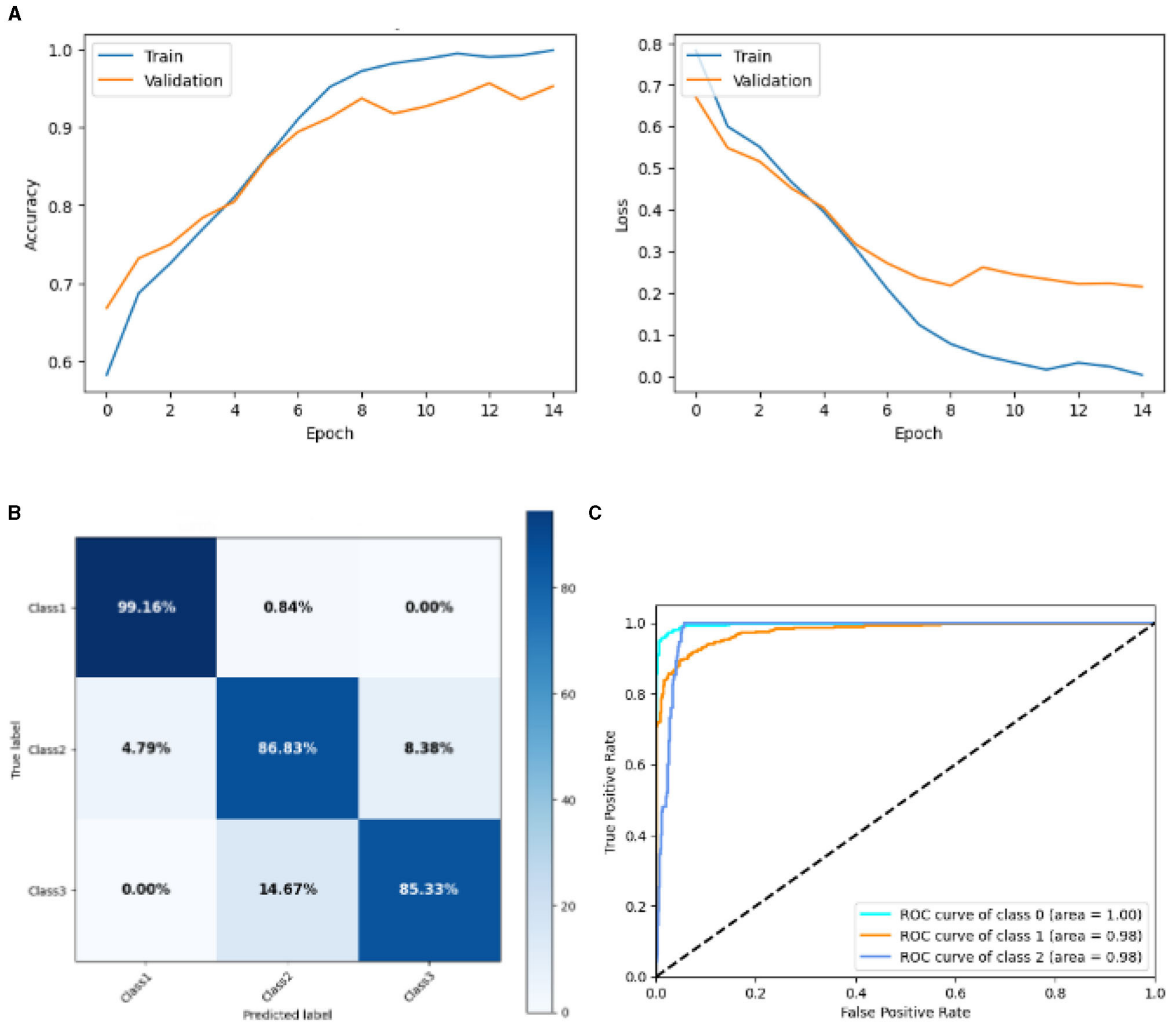


FIGURE 12
Graphical representation of SGCNN model 2's results. **(A)** Accuracy and loss. **(B)** Confusion matrix. **(C)** Receiver operating characteristics curve.

graph in Figure 13C. For different threshold settings, the graph plots the real positive rate against the false positive rate. AUC values for all three classes are very high, as can be seen from the ROC curve, indicating that the model is doing exceptionally well. Class 0 has an AUC of 1.00, class 1 of 0.97, and class 2 of 0.98. For all three classes, this demonstrates the model's extremely high accuracy level. Classes 0, 1, and 3 have cyan, orange, and blue ROC curves, respectively. The model has a low false positive rate and a high true positive rate, as indicated by the fact that all of its ROC curves are located close to the upper left corner. This attests to the model's high classification accuracy among the three classes.

4.2.5 Result of lean CNN model

Figure 14 represented the graphical representation of the Lean CNN Model. Figure 14A presents the model loss and accuracy of a Lean CNN Model. The training accuracy starts from 0.62% value at 0*th* epoch. Then, it increases upward, and the training accuracy stops at 1.00% at 14*th* epoch. The validation accuracy starts from 0.2% value at 0*th* epoch, and it increases at 0.75% at 1*st* epoch. After some fluctuation of increases and decreases, validation accuracy attained is 0.9% at 14*th* epoch. The training loss starts from a 0.9 value at 0*th* epoch, and it decreases downward up to a 0.0 value at 14*th* epoch. The validation loss starts from a 1.5 value at 0*th* epoch. It decreases up to 0.7 value at 1*st*, and then it increases up to 2.4 value at 7*th* epoch, and then it decreases up to 0.4 value at 14*th* epoch. The confusion matrix in Figure 14B gives the percentage of each class that is correctly and incorrectly classified. The elements on the diagonal consist of the classes that have been well-classified, while the non-diagonal elements on the matrix consist of the samples that are misclassified. For class 1, the model correctly classified 100.00% instances, and 0.00% instances of class 1 were misclassified. For class 2, the model correctly categorized 90.79%

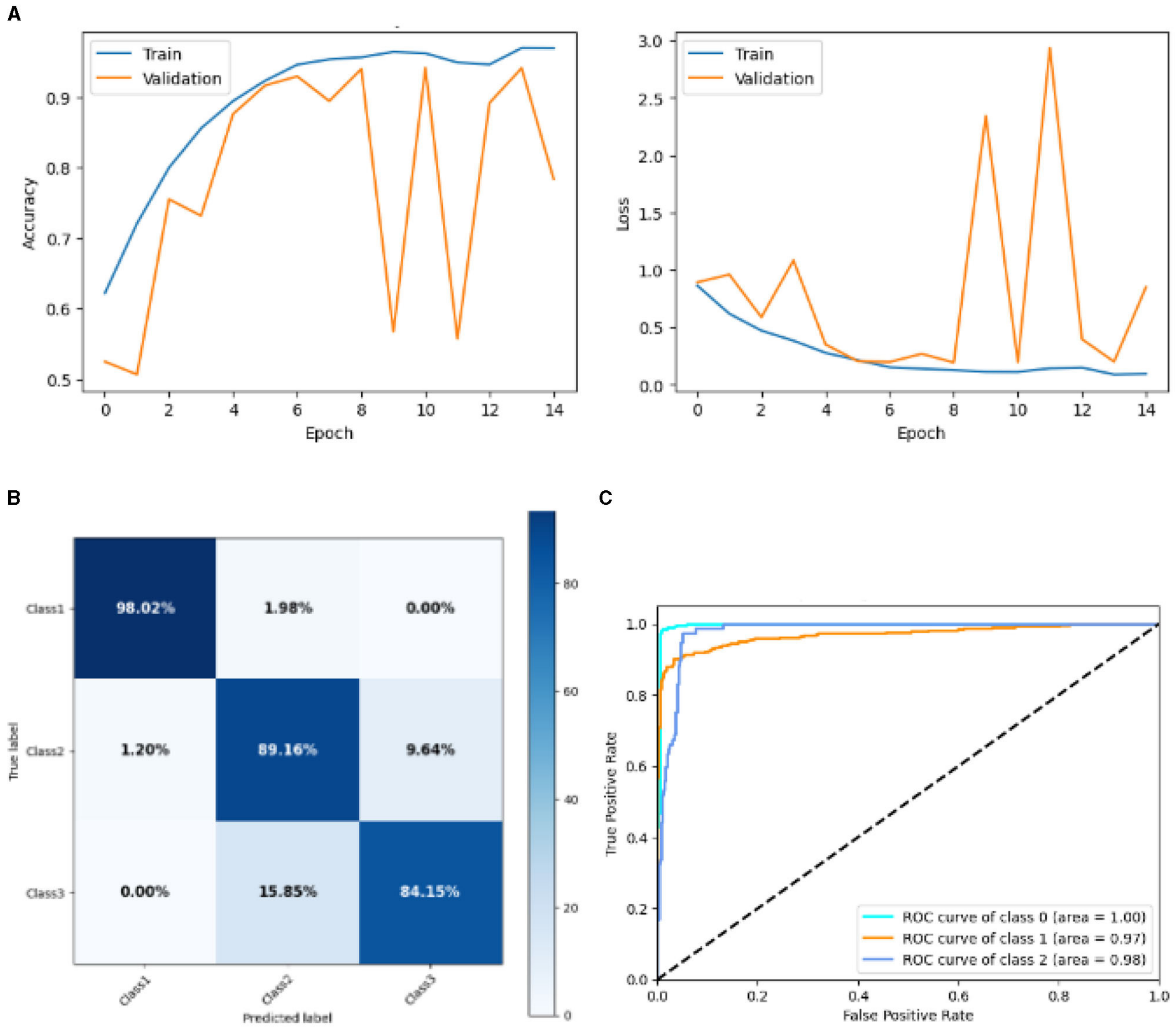


FIGURE 13
Graphical representation of base CNN model's results. **(A)** Accuracy and loss. **(B)** Confusion matrix. **(C)** Receiver operating characteristics curve.

instances and 2.31% instances of class 2 misclassified as class 1 and 7.51% instances of class 2 misclassified as class 3. For class 3, the model correctly classified 81.16% instances and 18.84% of class 3 instances misclassified as class 2. Overall, the Lean CNN Model maintains strong performance in Class 1. The performance of the model for each class is shown by the ROC graph in Figure 14C. For different threshold settings, the graph plots the real positive rate against the false positive rate. AUC values for all three classes are very high, as can be seen from the ROC curve, indicating that the model is doing exceptionally well. Class 0 has an AUC of 1.00, class 1 of 0.96, and class 2 of 0.95. For all three classes, this demonstrates the model's extremely high accuracy level. Classes 0, 1, and 3 have cyan, orange, and blue ROC curves, respectively. The model has a low false positive rate and a high true positive rate, as indicated by the fact that all of its ROC curves are located close to the upper left corner. This attests to the model's high classification accuracy among the three classes.

4.2.6 Result of deep CNN model

Figure 15 represented the graphical representation of the Deep CNN Model. Figure 15A presents the model accuracy and loss of a Deep CNN Model. The training accuracy starts from 0.7% value at 0*th* epoch. Then, it increases upward, and the training accuracy stops at 0.9% at 14*th* epoch. The validation accuracy starts from 0.3% value at 0*th* epoch, and it increases at 0.83% at 2*nd* epoch. After some fluctuation of increases and decreases, validation accuracy attained is 0.9% at 14*th* epoch. The training loss starts from a 0.9 value at 0*th* epoch, and it decreases downward up to a 0.0 value at 14*th* epoch. The validation loss starts from a 2.6 value at 0*th* epoch. It increases up to 4.0 value at 1*st*, and then it decreases up to 0.5 value at 2*nd* epoch, and then some fluctuation of increases and decreases the validation loss stops at 0.4 value at 14*th* epoch. The confusion matrix in Figure 15B gives the percentage of each class that is correctly and incorrectly classified. The elements on the diagonal consist of the classes that have been well-classified, while

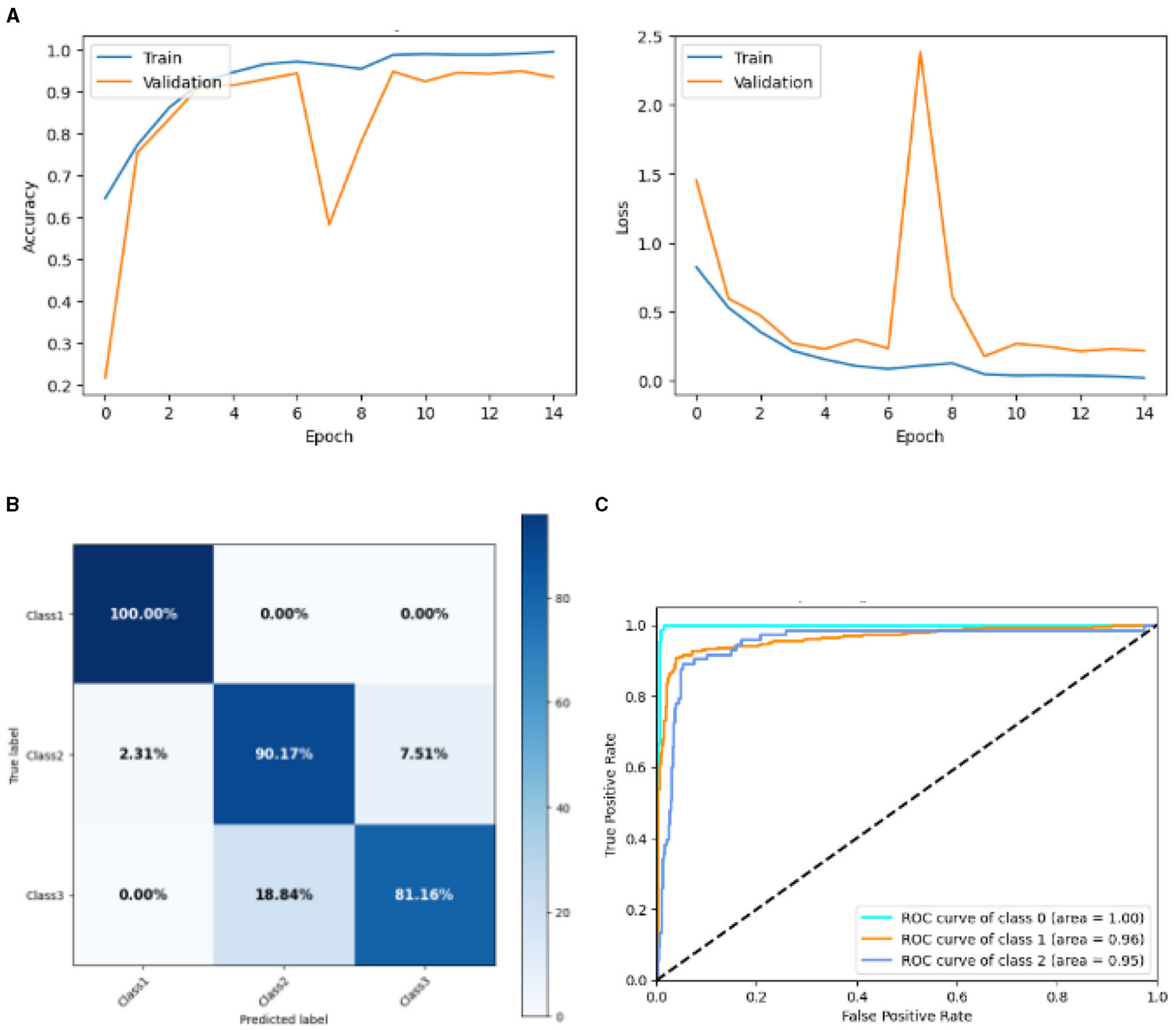


FIGURE 14
Graphical representation of lean CNN model's results. **(A)** Accuracy and loss. **(B)** Confusion matrix. **(C)** Receiver operating characteristics curve.

the non-diagonal elements on the matrix consist of the samples that are misclassified. For class 1, the model correctly classified 99.14% instances, and 0.51% instances of class 1 were misclassified as class 2. For class 2, the model correctly classified 92.75% instances and 0.29% instances of class 2 misclassified as class 1 and 6.96% instances of class 2 misclassified as class 3. For class 3, the model correctly classified 79.27% instances and 20.73% of class 3 instances misclassified as class 2. Overall, the deep CNN Model maintains strong performance in Class 1. The performance of the model for each class is shown by the ROC graph in Figure 15C. For different threshold settings, the graph plots the real positive rate against the false positive rate. AUC values for all three classes are very high, as can be seen from the ROC curve, indicating that the model is doing exceptionally well. Class 0 has an AUC of 1.00, class 1 of 0.98, and class 2 of 0.98. For all three classes, this demonstrates the model's extremely high accuracy level. Classes 0, 1, and 3 have cyan, orange, and blue ROC curves, respectively. The model has a low false positive rate and a high true positive rate, as indicated by the fact that all of its ROC curves are located close to the upper left corner. This attests to the model's high classification accuracy among the three classes.

4.3 Discussion and comparison

Table 4 presents a comparative analysis of various models employed for diagnosing and classifying brain-related disorders. The comparison includes studies from different years, showcasing the models used, the datasets involved, and the results achieved. One of the studies from 2022 (Shahwar et al., 2022) employed a Hybrid Classical–Quantum Transfer Learning approach

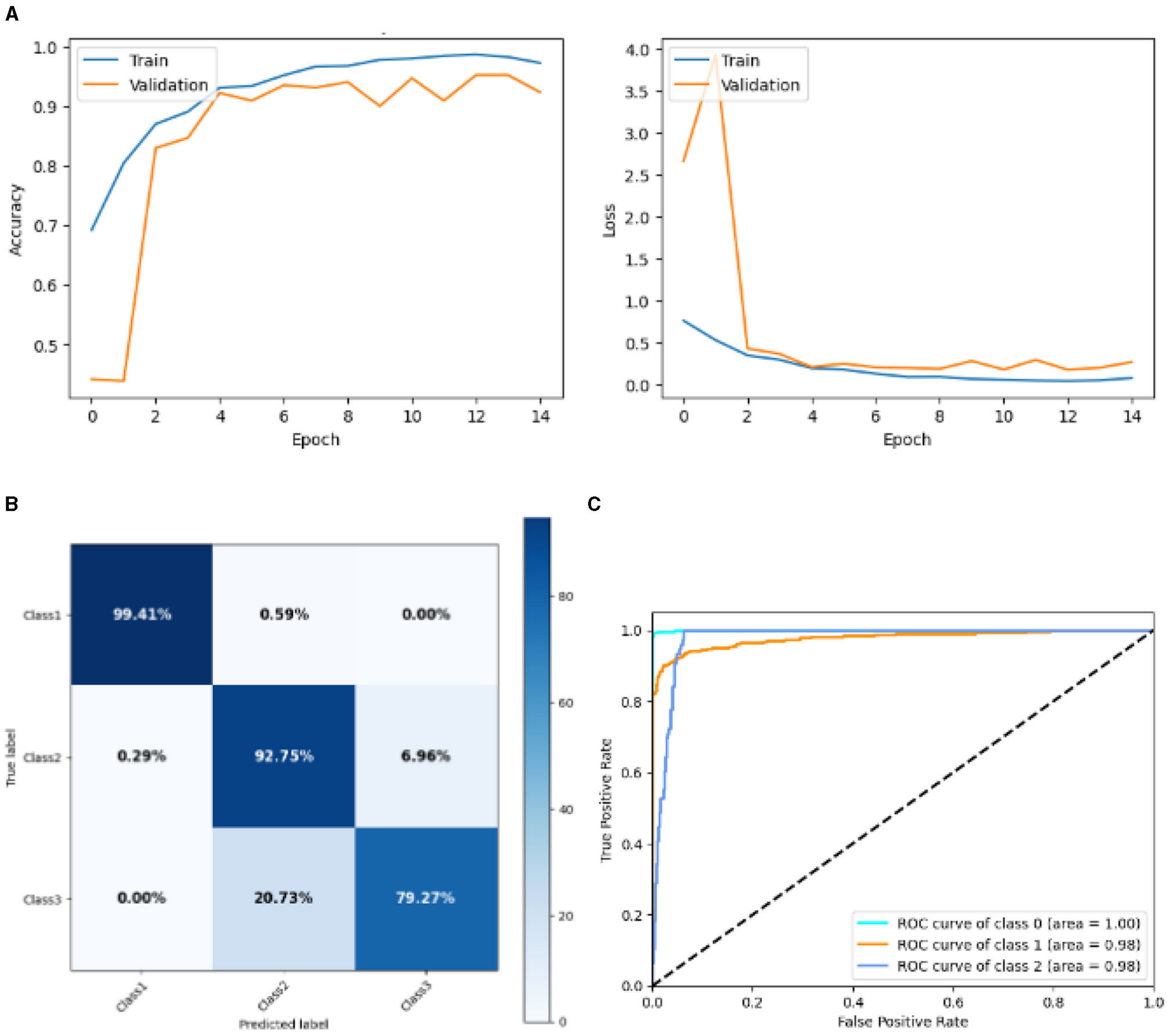


FIGURE 15
Graphical representation of deep CNN model's results. **(A)** Accuracy and loss. **(B)** Confusion matrix. **(C)** Receiver operating characteristics curve.

that combined ResNet34 with a Quantum Variational Circuit (QVC). This hybrid model was applied to a dataset focusing on dementia associated with AD. By leveraging the strengths of both classical and quantum computing, the model achieved an accuracy of 92%, demonstrating significant potential for enhancing the machine learning model's performance in dementia detection. In 2023, another study (Nancy Noella and Priyadarshini, 2023) explored multiple classifiers of machine learning, including Multi-class Support Vector Machine (SVM), Naive Bayes, ID3 and Bagged Ensemble. The dataset consisted of PET images representing AD, Parkinson's Disease, and healthy brains. The Bagged Ensemble classifier fared better than the others, obtaining a 90.3% accuracy rate, according to the study. This research demonstrates the precision with which complicated brain disorders can be classified using ensemble learning techniques.

The study (de Oliveira et al., 2024) utilized a logistic regression model with L1 and L2 regularization to diagnose AD. The performance of the model was evaluated using several measures, with the Area Under the Curve (AUC) reaching 94.75%, indicating a strong ability to generalize to unseen neuroimages. The proposed SGCNN model is specifically designed to classify AD images, Parkinson's disease images, and control subjects. This classifier attained an accuracy of 95%, surpassing the performance of previous studies and demonstrating its effectiveness in accurately diagnosing these conditions. The proposed model seems to be effective for AD detection from MRI images because they excel at capturing complex brain connectivity patterns, which are disrupted in AD. MRI data reflects the brain's structural and functional connectivity, which can be represented as a graph, where nodes correspond to brain regions, and edges represent connections between them.

TABLE 4 Comparison of proposed model's result with existing techniques.

References	Year	Models	Dataset	Results
Shahwar et al. (2022)	2022	Hybrid classical–quantum transfer learning with ResNet34 and Quantum variational circuit (QVC)	Dementia of AD	Accuracy 92%
Nancy Noella and Priyadarshini (2023)	2023	Bagged ensemble, ID3, Naive Bayes, multiclass support vector machine	Image dataset (AD, PD, healthy brain)	Accuracy 90.3%
de Oliveira et al. (2024)	2024	Logistic Regression with L1 and L2 regularization	Images dataset (AD, CN)	AUC 94.75%
Proposed model	**2024**	**SGCNN model 1**	**Image dataset (AD, PD, and CONTROL)**	**Accuracy 95%**

Traditional CNNs, designed for grid-like data such as images, struggle with such irregular structures. The proposed model, however, operates on graphs by applying spectral convolutions that capture intricate relationships in the brain's network, enabling them to identify subtle alterations in brain connectivity that are characteristic of Alzheimer's, improving the model's ability to detect the disease accurately. The outcomes demonstrate the elevated precision of the suggested approach and its potential for practical use in the prompt identification and diagnosis of neurodegenerative illnesses.

5 Conclusion

The purpose of this study was to increase the diagnostic precision of AD by proposing and evaluating many CNN models for picture categorization. We were able to provide a strong basis for model training and evaluation by carefully splitting our experimental dataset and using strict pre-processing. Further ablation investigations showed that structural alterations could improve performance, as illustrated by the Ablation of SGCNN Model 1, which achieved the maximum accuracy of 95%. The SGCNN Original Model served as a solid baseline with a 93% accuracy. Furthermore, with accuracies ranging from 93% to 94%, the BASE CNN, LEAN CNN, and Deep CNN models showed strong performance. Our results indicate the prospect of the ablation of SGCNN Model 1 as a very powerful tool for classifying AD images, underscoring its potential to support early diagnosis and therapy of AD. Additionally, the constant performance across different models suggests that CNN-based methods can be quite dependable for tasks involving the classification of medical images. To further advance the field of diagnosing neurodegenerative diseases, future studies could focus on enhancing these models by incorporating more diverse and larger datasets, integrating multi-modal data such as genetic or biochemical markers, and exploring real-time applications for early detection and continuous monitoring. Additionally, investigating the use of advanced techniques like transfer learning, ensemble methods, and model interpretability could help improve diagnostic accuracy and reliability in clinical settings.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

HA: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Validation, Writing – original draft, Writing – review & editing. RJ: Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. AA: Formal analysis, Methodology, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. S-jL: Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adaobi, C. C., Asaarik, M. J. A., Miezah, N. A., Odum, J. K., et al. (2023). A hybrid multi-class classification of Alzheimer disease based on operative deep learning techniques: Xception-fractalnet. *Converg. Chron*. 4, 796–812.

Amini, S., Hao, B., Yang, J., Karjadi, C., Kolachalama, V. B., Au, R., et al. (2024). Prediction of Alzheimer's disease progression within 6 years using speech: a novel approach leveraging language models. *Alzheimers Dement*. 20, 5262–5270. doi: 10.1002/alz.13886

Begum, A. P., and Selvaraj, P. (2024). Multiclass diagnosis of Alzheimer's disease analysis using machine learning and deep learning techniques. *Int. J. Image Graph*. 24:2450031. doi: 10.1142/S0219467824500311

Biswas, R., and Gini, J. R. (2024). Multi-class classification of Alzheimer's disease detection from 3D MRI image using ml techniques and its performance analysis. *Multimed. Tools Appl*. 83, 33527–33554. doi: 10.1007/s11042-023-16519-y

de Oliveira, C. E. G., de Araújo, W. M., de Jesus Teixeira, A. B. M., Gonçalves, G. L., Itikawa, E. N., Initiative, A. D. N., et al. (2024). Pca and logistic regression in 2-[18F] FDG pet neuroimaging as an interpretable and diagnostic tool for Alzheimer's disease. *Phys. Med. Biol*. 69:025003. doi: 10.1088/1361-6560/ad0ddd

El-Assy, A., Amer, H. M., Ibrahim, H., and Mohamed, M. (2024). A novel CNN architecture for accurate early detection and classification of Alzheimer's disease using MRI data. *Sci. Rep*. 14:3463. doi: 10.1038/s41598-024-53733-6

Francis, A., and Pandian, I. A. (2023). Ensemble learning approach for multi-class classification of Alzheimer's stages using magnetic resonance imaging. *Telkomnika* 21, 374–381. doi: 10.12928/telkomnika.v21i2.23352

Garg, N., Choudhry, M. S., and Bodade, R. M. (2023). A review on Alzheimer's disease classification from normal controls and mild cognitive impairment using structural MR images. *J. Neurosci. Methods* 384:109745. doi: 10.1016/j.jneumeth.2022.109745

Goenka, N., and Tiwari, S. (2023). Alzheimer's detection using various feature extraction approaches using a multimodal multi-class deep learning model. *Int. J. Imaging Syst. Technol*. 33, 588–609. doi: 10.1002/ima.22813

Hu, X., Sun, Z., Nian, Y., Wang, Y., Dang, Y., Li, F., et al. (2024). Self-explainable graph neural network for Alzheimer disease and related dementias risk prediction: Algorithm development and validation study. *JMIR Aging* 7:e54748. doi: 10.2196/54748

Karran, E., and De Strooper, B. (2022). The amyloid hypothesis in Alzheimer disease: new insights from new therapeutics. *Nat. Rev. Drug Discov*. 21, 306–318. doi: 10.1038/s41573-022-00391-w

Kaya, M., and Çetin-Kaya, Y. (2024). A novel deep learning architecture optimization for multiclass classification of Alzheimer's disease level. *IEEE Access* 12, 46562–46581. doi: 10.1109/ACCESS.2024.3382947

Khan, A. A., Mahendran, R. K., Perumal, K., and Faheem, M. (2024). Dual-3dm 3-ad: mixed transformer based semantic segmentation and triplet pre-processing for early multi-class Alzheimer's diagnosis. *IEEE Trans. Neural Syst. Rehabil. Eng*.32, 696–707. doi: 10.1109/TNSRE.2024.3357723

Knopman, D. S., Amieva, H., Petersen, R. C., Chételat, G., Holtzman, D. M., Hyman, B. T., et al. (2021). Alzheimer disease. *Nat. Rev. Dis. Primers* 7:33. doi: 10.1038/s41572-021-00269-y

Krishna, K. S., Kumar, B. D., Reddy, M. D., Saketh, C., and Ramasamy, G. (2024). "A multi-class classification framework with smote based data augmentation technique for Alzheimer's disease progression," in *2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE)* (Shivamogga: IEEE), 1–6. doi: 10.1109/AMATHE61652.2024.10582153

Logan, R., Williams, B. G., Ferreira da Silva, M., Indani, A., Schcolnicov, N., Ganguly, A., et al. (2021). Deep convolutional neural networks with ensemble learning and generative adversarial networks for Alzheimer's disease image data classification. *Front. Aging Neurosci*. 13:720226. doi: 10.3389/fnagi.2021.720226

Nancy Noella, R., and Priyadarshini, J. (2023). Machine learning algorithms for the diagnosis of Alzheimer and Parkinson disease. *J. Med. Eng. Technol*. 47, 35–43. doi: 10.1080/03091902.2022.2097326

Porsteinsson, A. P., Isaacson, R., Knox, S., Sabbagh, M. N., and Rubino, I. (2021). Diagnosis of early Alzheimer's disease: clinical practice in 2021. *J. Prev. Alzheimers Dis*. 8, 371–386. doi: 10.14283/jpad.2021.23

Rao, B. S., Aparna, M., Kolisetty, S. S., Janapana, H., and Koteswararao, Y. V. (2024). Multi-class classification of Alzheimer's disease using deep learning and transfer learning on 3d MRI images. *Trait. Signal* 41, 1397–1404. doi: 10.18280/ts.410328

Self, W. K., and Holtzman, D. M. (2023). Emerging diagnostics and therapeutics for Alzheimer disease. *Nat. Med*. 29, 2187–2199. doi: 10.1038/s41591-023-02505-2

Shahwar, T., Zafar, J., Almogren, A., Zafar, H., Rehman, A. U., Shafiq, M., et al. (2022). Automated detection of Alzheimer's via hybrid classical quantum neural networks. *Electronics* 11. doi: 10.3390/electronics11050721

Shukla, A., Tiwari, R., and Tiwari, S. (2023). Review on Alzheimer disease detection methods: Automatic pipelines and machine learning techniques. *Science* 5:13. doi: 10.3390/sci5010013

Srividhya, L., Sowmya, V., Ravi, V., Gopalakrishnan, E. A., and Soman, K. P. (2024). Deep learning-based approach for multi-stage diagnosis of Alzheimer's disease. *Multimed. Tools Appl*. 83, 16799–16822. doi: 10.1007/s11042-023-16026-0

Tripathi, T., and Kumar, R. (2024). Speech-based detection of multi-class Alzheimer's disease classification using machine learning. *Int. J. Data Sci. Anal*. 18, 83–96. doi: 10.1007/s41060-023-00475-9

Turer, B. Y., and Sanlier, N. (2024). Relationship of curcumin with aging and Alzheimer and Parkinson disease, the most prevalent age-related neurodegenerative diseases: a narrative review. *Nutr. Rev*. nuae079. doi: 10.1093/nutrit/nuae079

Vejandla, B., Savani, S., Appalaneni, R., Veeravalli, R. S., and Gude, S. S. (2024). Alzheimer's disease: the past, present, and future of a globally progressive disease. *Cureus* 16:51705. doi: 10.7759/cureus.51705

Venkatasubramanian, S., Dwivedi, J. N., Raja, S., Rajeswari, N., Logeshwaran, J., Praveen Kumar, A., et al. (2023). Prediction of Alzheimer's disease using dho-based pretrained cnn model. *Math. Probl. Eng*. 2023:1110500. doi: 10.1155/2023/1110500

Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., et al. (2020). Convolutional neural networks for classification of Alzheimer 's disease: Overview and reproducible evaluation. *Med. Image Anal*. 63:101694. doi: 10.1016/j.media.2020.101694

Zhao, J., Han, Z., Ding, L., Wang, P., He, X., Lin, L., et al. (2024). The molecular mechanism of aging and the role in neurodegenerative diseases. *Heliyon* 10:e24751. doi: 10.1016/j.heliyon.2024.e24751

frontiers | Frontiers in Neuroinformatics

TYPE Original Research
PUBLISHED 26 November 2024
DOI 10.3389/fninf.2024.1444650

OPEN ACCESS

EDITED BY
Claudio De Stefano,
University of Cassino, Italy

REVIEWED BY
Xianlong Zeng,
Ohio University, United States
Ramazan Incir,
Gümüşhane University, Türkiye

*CORRESPONDENCE
Khalaf Alsalem
✉ kosalem@ju.edu.sa

RECEIVED 06 June 2024
ACCEPTED 21 October 2024
PUBLISHED 26 November 2024

CITATION
Gasmi K, Ben Aoun N, Alsalem K, Ltaifa IB, Alrashdi I, Ammar LB, Mrabet M and Shehab A (2024) Enhanced brain tumor diagnosis using combined deep learning models and weight selection technique.
Front. Neuroinform. 18:1444650.
doi: 10.3389/fninf.2024.1444650

Enhanced brain tumor diagnosis using combined deep learning models and weight selection technique

Karim Gasmi[1], Najib Ben Aoun[2,3], Khalaf Alsalem[4]*, Ibtihel Ben Ltaifa[5], Ibrahim Alrashdi[1], Lassaad Ben Ammar[6], Manel Mrabet[6] and Abdulaziz Shehab[4]

[1]Department of Computer Science, College of Computer and Information Sciences, Jouf University, Sakkaka, Saudi Arabia, [2]College of Computing and Information, Al-Baha University, Alaqiq, Saudi Arabia, [3]REGIM-Lab: Research Groups in Intelligent Machines, National School of Engineers of Sfax (ENIS), University of Sfax, Sfax, Tunisia, [4]Department of Information Systems, College of Computer and Information Sciences, Jouf University, Sakaka, Saudi Arabia, [5]STIH: Sens Texte Informatique Histoire, Sorbonne University, Paris, France, [6]Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia

Brain tumor classification is a critical task in medical imaging, as accurate diagnosis directly influences treatment planning and patient outcomes. Traditional methods often fall short in achieving the required precision due to the complex and heterogeneous nature of brain tumors. In this study, we propose an innovative approach to brain tumor multi-classification by leveraging an ensemble learning method that combines advanced deep learning models with an optimal weighting strategy. Our methodology integrates Vision Transformers (ViT) and EfficientNet-V2 models, both renowned for their powerful feature extraction capabilities in medical imaging. This model enhances the feature extraction step by capturing both global and local features, thanks to the combination of different deep learning models with the ViT model. These models are then combined using a weighted ensemble approach, where each model's prediction is assigned a weight. To optimize these weights, we employ a genetic algorithm, which iteratively selects the best weight combinations to maximize classification accuracy. We trained and validated our ensemble model using a well-curated dataset comprising labeled brain MRI images. The model's performance was benchmarked against standalone ViT and EfficientNet-V2 models, as well as other traditional classifiers. The ensemble approach achieved a notable improvement in classification accuracy, precision, recall, and F1-score compared to individual models. Specifically, our model attained an accuracy rate of 95%, significantly outperforming existing methods. This study underscores the potential of combining advanced deep learning models with a genetic algorithm-optimized weighting strategy to tackle complex medical classification tasks. The enhanced diagnostic precision offered by our ensemble model can lead to better-informed clinical decisions, ultimately improving patient outcomes. Furthermore, our approach can be generalized to other medical imaging classification problems, paving the way for broader applications of AI in healthcare. This advancement in brain tumor classification contributes valuable insights to the field of medical AI, supporting the ongoing efforts to integrate advanced computational tools in clinical practice.

KEYWORDS

brain tumor prediction, ensemble learning, vision transformer, genetic algorithm, parameter selection

1 Introduction

Brain tumor is the most prevalent condition in children and also the most challenging sickness to identify. Despite the advancements in technology and the vast research being conducted to detect and categorize brain tumors, it remains a difficult endeavor due to the varied appearance of tumors and their similarity to normal brain structures. Magnetic resonance imaging (MRI) is the recommended modality for the detection of brain tumors. This procedure is remarkably effective despite being quite time-consuming, and the results are of great quality. The main goal of this study is to identify and categorize the tumor based on the provided MRI imaging. The implementation of automated tumor detection algorithms on MRI images would facilitate the identification of tumors at the earliest possible stage, a critical factor in the treatment of brain tumors. Once the tumor is detected, it is classified according to its severity. This task is intricate and requires a significant amount of time, as it is performed by radiologists who analyze a vast collection of MRI images. The implementation of automated brain tumor identification and classification will mitigate human error and expedite the detection process. The initial step in detecting and classifying brain tumors is to establish a system that assists in segmenting MRI images. This stage include the initial processing of MRI images, the extraction and reduction of features, and the classification of the tumor. The next step of the research is assisting the radiologist with the extensive database. Automated brain tumor classification is essential for alleviating the burden on radiologists and offering an effective tool for tumor categorization.

Brain tumor classification is frequently discussed to assist radiologists in accurately interpreting brain MRI images for diagnosis (Zahid et al., 2022). Prior research has demonstrated the enhancement of different cutting-edge deep learning models in the categorization of brain tumors. Post-secondary tests and categorization rely on achieving accuracy and evaluating image quality, which is a laborious task. Upon the object's discovery, it is imperative to promptly identify all potential attributes and connections among its fundamental elements. This is a broad and abstract idea. Convolutional neural networks have demonstrated success in image processing applications within the realm of deep learning models. The typical architecture comprises many convolutional layers. These layers capture and isolate different characteristics found in the input image. However, these qualities can be more accurately detected by further discerning the most crucial characteristics of the layers. The pooling layer is typically used to achieve significant features, but in the current day, there is a pressing need to identify important features at each layer.

Classical machine learning methods have demonstrated little efficacy in addressing real-world problems across several domains, despite their ability to generate scores for tasks using predefined knowledge sets. Furthermore, the process of extracting features from the highly-dimensional inputs of advanced medical devices becomes insignificant when compared to the constant emergence of new techniques in the field of deep learning. Hence, in this study, ensemble learning methods are favored since they allow for the simultaneous utilization of two or more classifiers and generally exhibit superior performance compared to deep learning algorithms. This work introduces a new ensemble learning technique dubbed BT-ViTEff, specifically developed for the classification of brain tumors in medical MRI scans. The approach seeks to address difficult limits, such as intricate and ever-changing backgrounds, which arise due to the presence of varying backgrounds in the input medical photos. ViTEff combines the Convolutional Neural Network EfficientNet v2 (EfficientNetV2) with the vision transformer V2 model. EfficientNet v2 is a newly developed convolutional neural network (CNN) structure. The approach seeks to address the training limitation of EfficientNet models by showcasing enhanced parameter efficiency and accelerated learning speed as compared to similar models. The first steps of preprocessing consist of resizing the receiving MRI images to dimensions of 224 × 224 pixels. Afterwards, a range of data augmentation techniques, including rotation, shearing, shifting, and zooming, are used to enhance the diversity of the training data. This improves the capacity of our model to precisely categorize brain tumors into 44 different groups and avoids overfitting. Afterwards, the MRI images are fed into the EfficientNet v2 and ViT v2 models to extract complex characteristics, resulting in a comprehensive representation of various brain tumor classifications. The fusion model utilizes Weighted Average Ensembling and Simple Average Ensembling approaches to combine the feature maps produced by the EfficientNet v2 and ViT v2 models. Ultimately, we utilize the genetic algorithm to establish the optimal and superior weight for each deep learning model employed.

The subsequent sections of the paper are structured in the following manner. Section 2 provides an overview of previous studies, while Section 3 provides a detailed explanation of the materials and procedures used in our research. The experimental findings are analyzed in Section 4, and Section 5 presents the final conclusions of this study.

2 Related work

Recent advances in medical image classification have leveraged various machine learning and deep learning models to enhance diagnostic accuracy. Studies have shown that ensemble learning, which combines multiple model predictions, often outperforms individual models by capturing complementary strengths. In this section, we aim to describe same previous work proposed.

2.1 Traditional machine learning techniques

Various machine learning models, including decision tree, K-nearest neighbor (K-NN), logistic regression, and multiple support vector machine (SVM) models, were created using the suggested features to classify brain tumors.

Wisaeng and Sa-Ngiamvibool (2023) proposed a novel method, known as fuzzy Otsu thresholding morphological approach, for segmenting brain tumors. The values derived from each histogram in the original MRI image were modified by the implementation of a color normalizing preprocessing approach, together with histogram specification. The data indicates that the accuracy rates for images of gliomas, meningiomas, and pituitaries are 93.77 percent, 94.32 percent, and 94.37 percent, respectively. This unequivocally confirms that these occurrences can be precisely identified.

Jena et al. (2022) introduced a method for classifying and segmenting brain tumors by utilizing textural data and employing various machine learning methods. The technique is comprised of two distinct stages: tumor categorization and tumor segmentation. During the tumor classification stage, the MRI scans undergo pre-processing, and texture features are taken from the images using several techniques for texture extraction. The retrieved features were merged to create a feature vector matrix with dimensions of 200 × 471. Afterwards, the feature vector matrix was employed to train five machine learning algorithms: Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), binary decision trees, Random Forest (RF), and ensemble approaches. The experimental findings indicate that the ensemble approaches yielded the most favorable outcome, attaining a classification accuracy of 96.98% and 97.01% for BraTS2017 + TCIA and BraTS2019 + TCIA, respectively.

Varuna Shree and Kumar (2018) have proposed a method that entails the classification of a tumor into one of two categories: "malignant" or "benign." To extract features, they employ the discrete wavelet transform, and they apply a Support Vector Machine (SVM) for classification. Compared to the other classification algorithms, the Support Vector Machine (SVM) consistently outperformed them in terms of accuracy.

Furthermore, Dev et al. (2019) developed a novel approach to differentiate between malignant and non-cancerous tumors. The term "cancerous" denotes a malignant tumor that has the potential to spread and cause damage. Conversely, "non-cancerous" denotes a benign tumor that is incapable of disseminating and is not detrimental. The segmentation technique was employed to recover segments of the tumor images. Additionally, the authors implemented a median filter to completely eradicate any extraneous noise that was present in the background. Their model achieves an accuracy of 92.31% by employing a Classification and Regression Tree (CART) and a Support Vector Machine (SVM).

Within a comparable framework, Williams and Li (2018) presented a classification technique that employs wavelet pooling. The researchers found that wavelet pooling produced better results in comparison to other pooling methods. They obtained positive results, but, the amount of time required was significant. The effectiveness of a pooling strategy cannot be determined due to its reliance on multiple factors, such as the dataset and the number of levels used in different models.

The researchers in Zacharaki et al. (2009) proposed a technique that use a Support Vector Machine (SVM) to classify gliomas into distinct categories. Their multi-classification accuracy rate was 85%, while their binary classification accuracy rate was 88%. Furthermore, the authors in Machhale et al. (2015) presented a model that employs Support Vector Machines (SVM) to classify brain cancers. Additionally, they performed a comparison between two Convolutional Neural Network (CNN) models to determine the most efficient one in terms of attaining ideal results.

Babu et al. (2023) employed MRI images to develop a method for classification and segmenting brain tumors. The technique comprises four procedures: image denoising, tumor segmentation, feature extraction, and hybrid classification. After applying the thresholding technique to remove malignancies from brain MRI scans, they next utilized a wavelet-based approach to extract unique characteristics from the images. A Convolutional Neural Network (CNN) was utilized to carry out the conclusive hybrid categorization. The trial resulted in a segmentation accuracy of 95.23 percent for the technique, whereas the suggested optimized CNN attained a classification accuracy of 99 percent.

An advanced version of the Support Vector Machine (SVM) was proposed by Ansari (2023) as a novel approach. In order to identify and categorize brain cancers using MRI data, they suggested the implementation of the following four stages: preprocessing, image segmentation, feature extraction, and image categorization. The tumors were divided using a fuzzy clustering technique, and the fundamental characteristics were retrieved using GLCM. Improvements to the Support Vector Machine (SVM) were eventually integrated into the categorizing procedure. The technique used resulted in an accuracy rate of 88%.

2.2 Deep learning techniques

In order to gain a comprehensive understanding of deep learning algorithms in the specific context of brain tumor detection and diagnosis, it is crucial to consider the four fundamental deep learning tasks: single label image classification (İncir and Bozkurt, 2024b), multi-label/multi-class image classification, object detection in images, and dense prediction at the pixel level. The referenced study focuses on single label image classification. However, our objective is to achieve multi-label/multi-class image classification. This means that an input MRI scan can be assigned many tumor kinds or no tumor at all. An expansion of this work would involve identifying things inside an image, specifically cancers. Ultimately, dense prediction at the pixel level would entail accurately segmenting a tumor using MRI data.

Deep learning is a newly emerged field in machine learning that involves a classifier which receives an input x and converts it into a separate domain with the same dimensions. When it comes to annotated data classification tasks, the input domain refers to the collection of photos, while the output domain refers to the collection of class labels. This classifier can be seen as a combination of numerous elementary feature extraction and mapping modifications. Deep learning approaches have been minimally utilized in the domain of brain tumor detection using patient MRI data, but considerable progress has been made. This work employs a fusion between tow deep learning models, to classify input data into one of 44 distinct classes. Although this method achieves generally good classification outcomes, more sophisticated deep learning systems have the potential to effectively categorize complex and noisy data with high dimensionality into a reduced dimensional feature space, which could potentially lead to improved classification results.

In this context, Abd El Kader et al. (2021) introduced a convolutional neural network (CNN) method for classifying MR brain images. Their approach involved utilizing differential deep CNNs. In conventional Convolutional Neural Networks (CNNs), standard feature maps are generated through either random initialization or transfer learning. Nevertheless, this study generated distinct feature maps by utilizing user-defined hyperactive values and a differential operator proposed by Lei et al. (2018). The generated differential convolution maps are utilized to examine the directional patterns of voxels and their surrounding areas by computing the disparity in pixel activations. The study presented multiple data augmentation approaches to enhance the classification model's generalization performance. The process of data augmentation resulted in an expansion of the dataset, increasing its size to 25,000. The results indicate that the inclusion of differential feature maps enhanced the model's performance. The results additionally demonstrate that the suggested methodology may accurately classify a significant number of MR images. The approach attained a classification accuracy of 99.25%, sensitivity of 95.89%, and specificity of 93.75%.

A recent study conducted by Tanvir Rouf Shawon et al. (2023) introduced a cost-sensitive deep neural network (CS-DNN) designed to detect brain cancers from MRI (Magnetic Resonance Imaging) images. The suggested model employed a DenseNet architecture, a type of convolutional neural network, to automatically extract intricate and profound characteristics. Additionally, cost-sensitive learning was incorporated into the DenseNet to address the issue of class imbalance in the radiology dataset. The researchers employed both binary and multi-class cost-sensitive learning techniques to identify brain tumors from the MRI imaging dataset. CS-DNN outperformed seven alternative models in terms of sensitivity, specificity, precision, and accuracy for Tumor 3MRI images, making it the top performer.

Ge et al. (2020) addressed the issue of limited datasets by utilizing augmented brain MR images. A paired generative adversarial network (GAN) was employed to produce synthetic MR images for four different MRI techniques. The work involved extracting 2D MRI slices from three different perspectives of 3D volume images: coronal, axial, and sagittal. The 2D MRI slices that were obtained were separated into subgroups for training, validation, and testing purposes. Moreover, a paired Generative Adversarial Network (GAN) model was employed to produce artificial Magnetic Resonance Imaging (MRI) for the subset utilized in training. The pairwise GAN employed a dual input system with two separate streams. The purpose of this system is to address two specific situations: (a) generating artificial images of non-existent patients in order to expand the training dataset, and (b) generating artificial images for patients who are lacking certain MRI modalities. The study employed the U-Net architecture. The ultimate result of the architecture is the classification of glioma for each individual slice in a magnetic resonance (MR) image. Therefore, for every patient, the subtype of each MRI slice was taken into account, and the ultimate diagnostic or subtype categorization for the patient will be determined by a majority consensus. The experiments evaluated many case studies, and the case study that yielded the most favorable outcome attained an average classification accuracy, sensitivity, and specificity of 88.82%, 81.81%, and 92.17%, respectively.

2.3 Ensemble learning approach

Ensemble learning is employed to enhance the classification performance. Ensemble learning approaches leverage the power of several learning algorithms to achieve superior predicted performance compared to individual algorithms. Currently, there is a scarcity of research studies on brain tumor classification utilizing ensemble learning. Bansal and Jindal (2022) employed decision tree, random forest, and k-NN algorithms to classify brain tumors in MRI images. The classification of four types of cancers was performed using intensity, texture, and wavelet data. The highest level of accuracy achieved was 83.33% by utilizing the random forest algorithm.

Sekhar et al. (2021) introduced a tumor classification model that utilizes a modified GoogleNet pre-trained CNN model together with two machine learning algorithms: Support Vector Machine (SVM) and k-Nearest Neighbors (k-NN). The project involved modifying and fine-tuning the last three fully connected layers of the GoogleNet network using brain tumor photos. The 1,024 feature vector obtained from the final average pooling layer was recovered after fine-tuning and utilized for training SVM and k-NN classifiers. The method was assessed using the CE-MRI dataset, which consists of 3,064 T1w post GBCA brain MR images obtained from 233 patients. The experimental findings indicate that GoogleNet achieved a precision of 96.02% and a recall of 97.00% for glioblastoma while employing the softmax activation function. The utilization of the SVM classifier resulted in a performance enhancement of more than 2.5% for the model.

Jena et al. (2022) proposed a technique to classify and segment brain tumors by leveraging textural data and employing diverse machine learning algorithms. The technique comprises two separate stages: tumor categorization and tumor segmentation. During the tumor classification step, the MRI scans are subjected to pre-processing, and texture features are gathered from the images using several texture extraction methods. The study examined many features that are based on texture. Data was gathered from a combined set of 100 images of tumors and 100 images without tumors to extract features. The extracted characteristics were combined to form a feature vector matrix measuring 200 × 471. Subsequently, the feature vector matrix was employed to train five machine learning algorithms: Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), binary decision trees, Random Forest (RF), and ensemble techniques. The ensemble methods consist of seven distinct algorithms: Adaboost, Gentleboost, Logitboost, LPboost, Robustboost, RUSboost, and Totalboost. After the training was finished, the study used the photos that had tumors to construct a hybrid technique for segmenting tumors. The hybrid technique involves combining the k-NN and fuzzy C-means clustering techniques. The hybrid approach was used to divide the tumor regions in the images. The dataset used for model evaluation consists of the BraTS2017 and BraTS2019 datasets, along with the

Cancer Imaging Archive (TCIA). The experimental results indicate that the ensemble techniques achieved the maximum outcome.

Kang et al. (2021) introduced a technique for categorizing brain tumors by employing a combination of deep characteristics. The methodology comprises three distinct stages. During the initial phase, input photos undergo pre-processing, and additional images are created by data augmentation techniques. The preprocessed photos are subsequently utilized as input for 13 pre-trained convolutional neural network (CNN) models. Pre-trained convolutional neural network (CNN) models are employed to extract features from the images. Specifically, the characteristics are derived from the fully linked layers of the pre-trained models. The collected features are utilized for training nine machine learning classifiers, specifically: Gaussian Nave Bayes, Extreme Learning Machine (ELM), Adaptive Boosting (AdaBoost), k-NN, RF, SVM, and neural networks with a fully connected layer. Furthermore, the three most successful pre-trained models are identified, and the retrieved features from these models are merged into a single sequence. Ultimately, the amalgamated characteristics are employed to train the nine machine learning classifiers. The method was assessed using three brain MRI datasets obtained from Kaggle websites. The findings indicated that DenseNet-169, Inception-v3, and ResNeXt-50 offered the most favorable characteristics.

Deepak and Ameer (2021) introduced an automated technique for classifying brain tumors using Support Vector Machines (SVM) and Convolutional Neural Networks (CNN). A Convolutional Neural Network (CNN) was employed in the research to extract picture attributes from Magnetic Resonance Imaging (MRI) scans. The Convolutional Neural Network (CNN) comprises of five convolutional layers and two fully linked layers. The feature maps obtained from the fifth convolution layer and the first fully connected layer are isolated and utilized independently to train a Support Vector Machine (SVM) for the purpose of multiclass classification. The fifth convolution layer has 3,136 feature vectors, while the first fully linked layer has 10 feature vectors. When trained on the 10 feature vectors retrieved from the fully connected layer, the suggested approach attained an accuracy of 95.82%. The accuracy decreased to 93.83% when the model was trained using the 3,136-feature set retrieved from the fifth convolution layer. These findings indicate that Support Vector Machine (SVM) models trained on smaller feature sets have the capacity to yield superior outcomes compared to models trained on larger feature sets.

Multiple authors (Pereira et al., 2016) have investigated the use of tiny kernels to construct deeper networks that mitigate overfitting. The trials have shown that small kernels produce effective brain segmentation results. Furthermore, specific research has established effective and adaptable systems for the detection of brain cancers. Sharif et al. (2021) devised a versatile framework that can do multiple functions, including enhancing tumor visibility, extracting and selecting characteristics, localizing tumors, and segmenting tumors. The study included many advanced techniques such as the homomorphic wavelet filter, inception-v3 model, non-dominated sorted genetic algorithm (NGSA), YOLOv2, and ML algorithms to achieve specific objectives. These objectives included enhancing tumor visibility, extracting relevant features, selecting important features, localizing tumors, and segmenting tumors. The study's studies illustrate that the adaptable framework produced positive results. This method will be incredibly beneficial for medical practitioners as it allows them to effectively handle several responsibilities.

Binary classification is a significant obstacle for current classification algorithms. The majority of the current methodologies were devised to classify brain tumors into two distinct groups: benign and malignant. The study conducted by Sajjad et al. (2019) is among the limited number of research works that have devised a technique for classifying multiple classes simultaneously. Advanced multi-grade classification techniques have the potential to enhance the decision-making and diagnosing abilities of radiologists and other medical practitioners.

3 Proposed model

Convolutional neural networks (CNNs) have consistently achieved the highest level of performance in computer vision tasks, specifically in the areas of brain tumor segmentation and classification, in recent years. Nevertheless, Convolutional Neural Networks (CNNs) are limited in their ability to effectively collect extensive information or interconnections because of their tiny kernel size (Hatamizadeh et al., 2021). Long-range dependencies refer to situations when the desired outcome is influenced by visual sequences that were displayed at significantly earlier or later dates. Medical images often display a series of visual representations due to the resemblance of human organs (Dai et al., 2021). The elimination of these sequences will have a substantial impact on the performance of a CNN model. The reason for this is that the interconnections among medical picture sequences, such as modality, slice, and patch, provide substantial information (Dai et al., 2021). Sequences that have long-range dependencies can be effectively managed using techniques that are capable of processing sequence relations. The self-attention mechanism employed in ViTs (Dosovitskiy et al., 2020) possesses the ability to effectively capture long-range dependencies, a crucial factor in achieving accurate brain tumor segmentation. ViT-based models are able to learn local and global feature representations by modeling pairwise interactions between token embeddings, as described in Raghu et al. (2021)'s work on vision. ViT has exhibited encouraging performance on diverse benchmark datasets (Hatamizadeh et al., 2021; Wenxuan et al., 2021).

Although CNNs have achieved notable success, they do have certain limitations. Initially, Convolutional Neural Networks necessitate extensive datasets for the purpose of training. Furthermore, Convolutional Neural Networks (CNNs) generally lack resilience when it comes to affine rotations and transformations, as stated by Rodriguez et al. (2019) in their study on rotation. Moreover, the routing strategy utilized by CNN's pooling layers differs from the routing mechanism utilized by the human visual system. The CNN pooling layer distributes all the information acquired from the image to every neuron in the following layer, disregarding crucial details or little objects in the image (Aziz et al., 2021).

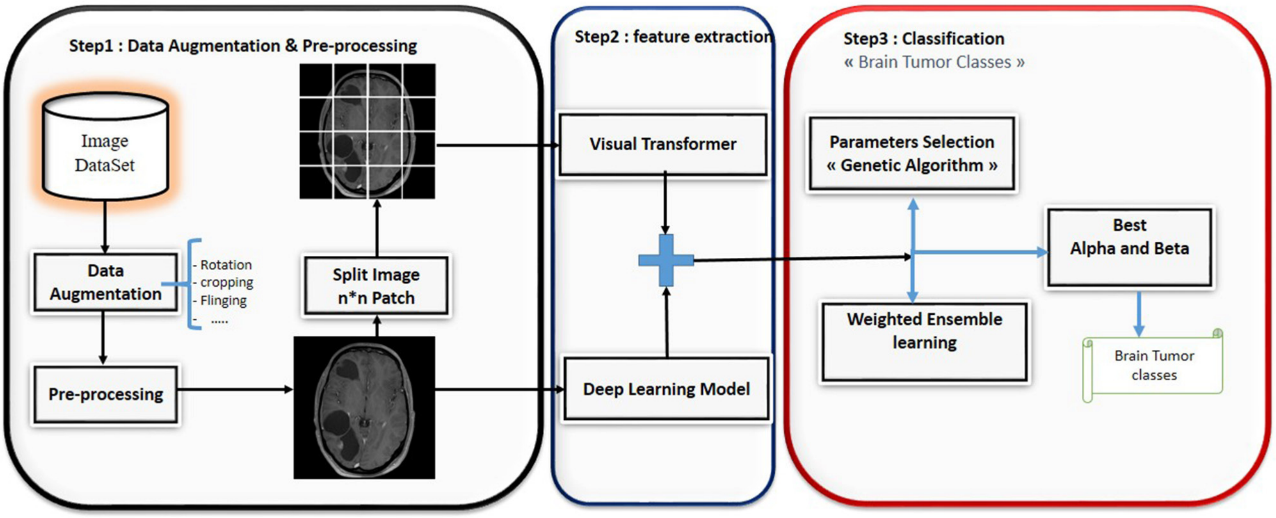


FIGURE 1
Hybrid deep learning approach for brain tumor classification.

This paper presents a novel ensemble learning method called BT-ViTEff, presented in Figure 1, which is designed for classifying brain tumors in medical MRI images. The approach aims to overcome challenging limitations such as complex and dynamic backgrounds that are caused by the presence of changing backgrounds in the input medical images. ViTEff integrates the advanced Convolutional Neural Network EfficientNet v2 (EfficientNetV2M) with the second version of the vision transformer. EfficientNet v2 is a novel convolutional neural network (CNN) architecture. The proposal aims to overcome the training constraint of EfficientNet models by demonstrating improved parameter efficiency and faster learning speed in comparison to comparable models. The system utilizes an enhanced progressive learning approach that dynamically modifies regularization approaches, including data augmentation and dropout algorithms, based on the input image size. In order to utilize EfficientNet v2 and ViT v2 models for brain tumor recognition, the classification layers (final layers) that were initially designed for distinct classification tasks are eliminated. As shown in Figure 1, the initial steps of preprocessing involve scaling the incoming RMI photos to dimensions of 224 × 224 pixels. Subsequently, a variety of data augmentation methods, such as rotation, shearing, shifting, and zooming, are employed to increase the diversity of the learning data. This enhances the capability of our model to accurately classify brain tumors into 44 distinct categories and prevents overfitting. Subsequently, the MRI images are inputted into the EfficientNet v2 and ViT v2 models to extract intricate features, thereby generating a comprehensive representation of different brain tumor classifications. The fusion model applies techniques such as Weighted Average Ensembling and Simple Average Ensembling after concatenating the feature maps generated by the EfficientNet v2 and ViT v2 models. This ensemble learning model will compute a probability classification for each image. Weighted ensembling is a form of model averaging ensembling, which falls under the domain of ensemble methods that aim to enhance prediction accuracy by aggregating the predictions of numerous models. Weighted ensembling involves assigning a specific weight or factor to each model's prediction, which represents the model's relative importance or performance. The weights allocated to the prediction of each model can be calculated using several strategies, including cross-validation, grid search, or meta-learning. Ultimately, we employ the genetic algorithm to determine the most ideal and superior weight for each deep learning model utilized.

3.1 EfficientNet model

EfficientNet utilizes a compound coefficient to scale the size of a CNN (İncir and Bozkurt, 2024c). The EfficientNet scaling approach employs a standardized set of scaling coefficients to evenly change these values, hence enhancing the standard procedure. By increasing the network depth by a factor of α, β, and γ, we are able to utilize a total of 2N times the quantity of processors that are currently accessible. By conducting a rapid grid search on the initial, simplified model, we obtained these fixed coefficients. If the input image is larger, it is logical to conclude that a network requires extra layers and channels to expand the receptive field and capture more detailed patterns. The EfficientNet-B0 network was built by incorporating MobileNetV2's inverted bottleneck residual blocks and incorporating additional squeeze-and-excitation blocks.

3.2 Vision transformer model

The Transformer architecture, presented in Vaswani et al. (2017), is currently at the forefront of new research in natural language processing (NLP). Dosovitskiy et al. (2021) was inspired by the success of self-attention-based deep neural networks in natural language processing to design the Vision Transformer

(ViT) architecture for picture categorization in NLP. Training these models often requires decomposing the input image into its individual components and subsequently considering each embedded component as if it were a word in a natural language processing (NLP) system. These models utilize self-observation modules to establish the connection between the concealed patches. Due to their exceptional efficacy, numerous scientists have investigated ViT models for diverse visual tasks (Yu et al., 2021). Carion et al. (2020) introduced a novel architecture for object recognition systems. This architecture utilizes an aset-based global loss and a transformer-encoder-decoder technique. They obtained equivalent results to the popular R-CNN approach on the challenging COCO dataset.

Steiner et al. (2021) initially presented his established architecture, which closely resembled the original ViT design by Dovosviky, but with a linear classifier instead of the MLP header. To summarize, the initial stage of training a ViT model involves dividing the input image into smaller segments. The transformer encoder takes a sequence of 1D patch embeddings as input. It employs self-attention mechanisms to compute a weighted sum of the outputs from each hidden layer, considering their interdependencies. This is accomplished by feeding the sequence into an encoder. The transformers utilize this mechanism to decode the global dependencies of the input photos. Refer to Figure 2 for a simple vision 348 transformer architecture.

3.3 Ensemble learning for brain tumor classification

At this point, we were concentrating on the last step to classify the input image. Ensemble learning models strive to combine the outcomes of several algorithms to improve overall performance and facilitate interpretation. This section highlights the hybridization strategy commonly employed to enhance predictive accuracy and resilience. Numerous ensemble learning techniques are documented in the literature, and we will use two approach:

Ensemble averaging is a technique used to enhance the accuracy and stability of a predictive model by merging the predictions of numerous base models. In this methodology, every foundational model is trained using the identical dataset, but with distinct hyperparameters, algorithms, or subsets of features. This is done to capture various facets of the data and mitigate the likelihood of overfitting. After training the base models, their forecasts are consolidated into a single prediction by an averaging strategy, such as simple averaging or weighted averaging. The arithmetic mean is determined by summing the forecasts of each base model and thereafter dividing by the total number of models. Simple average ensembling is a direct and efficient method for merging the predictions of many models, particularly when the separate models exhibit comparable performance and reliability.

Simple average ensembling is a versatile technique that may be used in various machine learning applications and algorithms, including as regression, classification, and clustering, among others. In addition, it is possible to integrate this strategy with other ensemble methods, such as weighted averaging or bagging, in order to enhance the model's performance and accuracy. Weighted ensembling involves assigning a specific weight or factor to each model's prediction, which indicates the model's relative importance or performance. The weights allocated to the prediction of each model can be established by many strategies, including cross-validation, grid search, or meta-learning.

Weighted ensembling is a versatile and potent strategy that can merge the advantages of several models and alleviate their limitations, resulting in enhanced prediction accuracy and resilience. In our model, we use an optimal algorithm to select the optimal weight for each model.

3.4 Optimal weight selection

In the domain of medical image classification, achieving high accuracy is paramount and ensemble learning, which combines the predictions from multiple models, often yields superior performance compared to individual models; however, determining the optimal weights for each model in the ensemble is a challenging task and to address this, we propose the use of a Genetic Algorithm (GA) for weight selection, aiming to maximize the accuracy of our brain tumor classification system; a Genetic Algorithm is an optimization technique inspired by the principles of natural selection and genetics, operating through a process of selection, crossover, and mutation to evolve solutions toward optimality, with the main components including a population of potential solutions (chromosomes) to the optimization problem, each chromosome representing a pair of weights alpha and beta for the deep learning models in the ensemble, and a fitness function evaluating how well each chromosome solves the problem, which in our case is the accuracy of the ensemble model; the selection process chooses the best chromosomes for reproduction based on their fitness scores, with crossover combining pairs of selected chromosomes to produce offspring, and mutation introducing random variations to maintain genetic diversity, all iterating through generations until convergence criteria are met, such as a maximum number of generations or a satisfactory fitness score; the implementation of the GA involves generating an initial population with random values for alpha and beta, evaluating the fitness function for each chromosome based on the accuracy of the ensemble model, selecting the top-performing chromosomes, performing crossover and mutation to create a new generation, and repeating the process until the optimal weights are found, thus leveraging the complementary strengths of the Vision Transformer and other deep learning models, with results demonstrating that the GA-enhanced weighted average ensemble outperforms individual models and simple averaging methods, achieving superior accuracy, precision, recall, and F1-score, ultimately contributing to more reliable and accurate diagnostic tools in healthcare, and highlighting the potential of evolutionary algorithms in fine-tuning ensemble learning models for complex medical image classification tasks.

To adapt the Genetic Algorithm for selecting the optimal weights (alpha and beta) in our ensemble learning model, we follow these steps, Figure 3: 1) Chromosome Representation: Each chromosome represents a pair of weights α and β for the models in the ensemble. The weights must satisfy the condition : α and $\beta = 1$

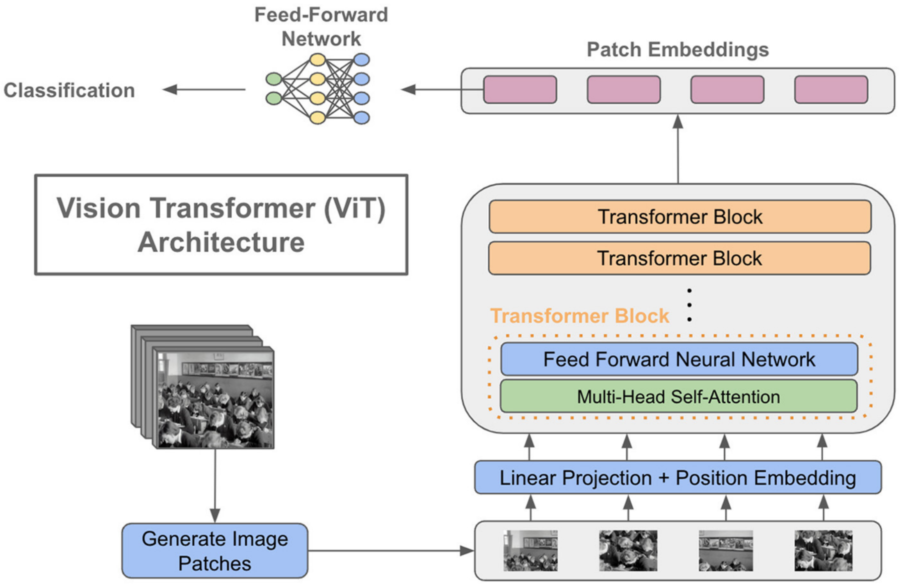


FIGURE 2
A basic depiction of a vision transformer architecture. [a]https://cameronrwolfe.substack.com/p/vision-transformers.

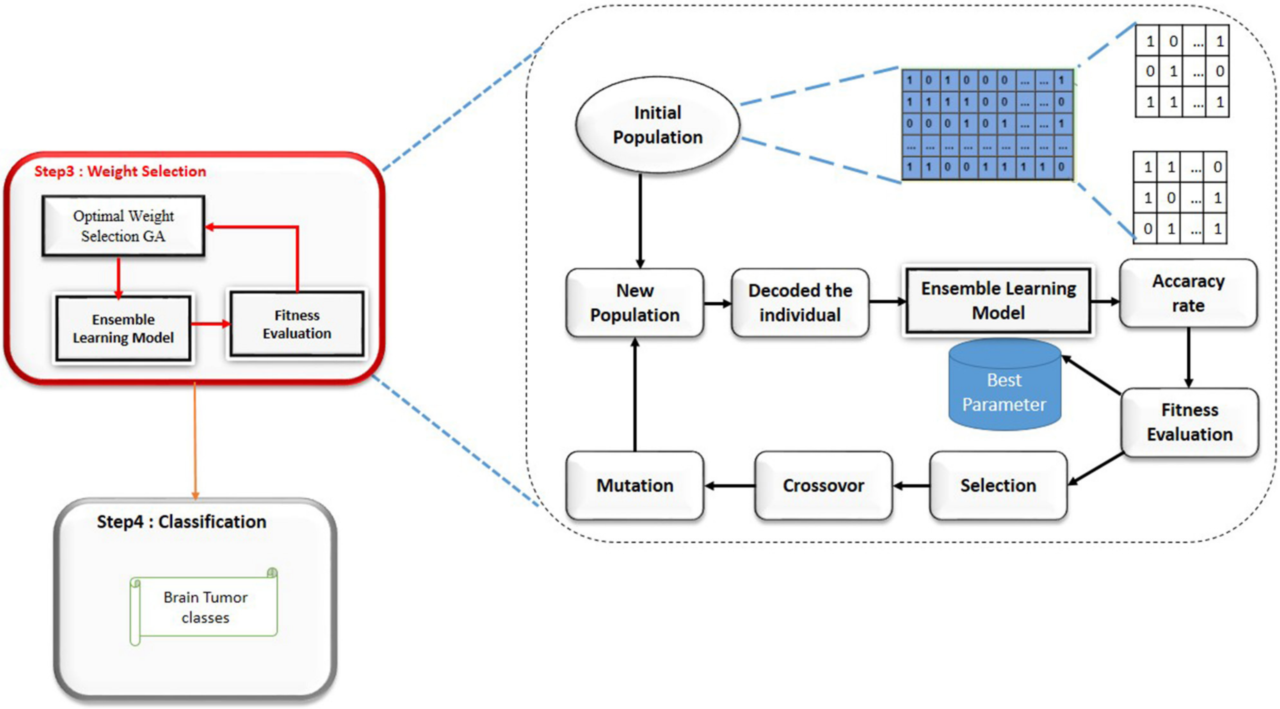


FIGURE 3
Genetic algorithm steps (Gasmi et al., 2024).

2) Initial Population: Generate an initial population of chromosomes with random values for α and β.
3) Fitness Function: The fitness function evaluates the accuracy of the ensemble model for each chromosome. Given a chromosome with weights α and β, the fitness function is defined as: Fitness(α, β)=Accuracy;
4) Selection: Select the top-performing chromosomes based on their fitness scores. Techniques such as roulette wheel selection or tournament selection can be used.
5) Crossover: Combine pairs of selected chromosomes to produce

offspring. For example, a simple crossover method could be:

Offspring1 = (α_1, β_2) and Offspring2 = (α_2, β_1)
where (α_1, β_2) and (α_2, β_1) are parent chromosomes
6) Mutation: Introduce random changes to the offspring to maintain genetic diversity. For instance, a small random value could be added or subtracted from (α, β), ensuring that the sum remains 1.
7) New Generation: Replace the old population with the new generation of chromosomes.
8) Termination: Repeat the selection, crossover, and mutation steps until convergence criteria are met, such as a maximum number of generations or a satisfactory fitness score.

3.5 Results and discussions

This section provides the findings of the experiments conducted on the model described in this study. In addition, we evaluate the outcomes achieved through various classification techniques using deep learning models. The Python implementation of the suggested model utilized a Rtx 2,060 graphics card and 16 GB of RAM.

In order to implement our suggested model, we conducted multiple sets of tests utilizing deep learning models. The structure of our model encompasses three distinct scenarios:

1. Scenario 1: Brain Tumor classification based on the deep learning models.
2. Scenario 2: Brain Tumor classification based on the hybrid deep learning models.
3. Scenario 3: Brain tumor classification based in the weight selection method.

3.6 Data set description and evaluation metrics

To assess our improved deep learning model, we used a dataset with 44 classes,[1] which is available online. Figure 4 provides an example of each classes in the data set.

We employed typical measures, such as accuracy, precision, recall, and F-measure, to test our classification approach.

$$Accuracy = (TP + TN)/(TP + TN + FP + fN) \tag{1}$$

$$Precision = TP/(TP + FP) \tag{2}$$

$$Recall = TP/(TP + FN) \tag{3}$$

$$F1 - score = TP/(TP + 0.5 * (FN + FP)) \tag{4}$$

Where TP stands for true positive, FP stands for false positive, P stands for precision, R stands for recall, TPR stands for true positive rate, and FPR stands for false positive rate.

3.7 Brain tumor classification based on deep learning models

The main objective of this study was to utilize hybrid deep learning models for the categorization of brain tumors. However, in this section, we aimed to classify MRI images into 44 distinct classes of brain tumors by leveraging various state-of-the-art deep learning models. The models evaluated in this study include Inception-ResNet-v2, EfficientNet V2, MobileNet V2, and the Vision Transformer (ViT). These models were chosen because of their strong track record in image classification, particularly in medical imaging. They offer a good balance between high accuracy and computational efficiency, making them ideal for complex tasks like brain tumor classification. The decision to use these models was based on results from similar studies in the same field, where they consistently performed well. Each model brings unique strengths: Inception-ResNet-v2 is excellent at capturing detailed features in images, EfficientNet V2 is highly efficient and scalable, MobileNet V2 is lightweight and suited for mobile applications, and ViT uses transformer-based attention mechanisms to extract deep features. Together, these models form a strong foundation for evaluating and improving tumor classification using medical image datasets. Our results, as detailed in Table 1, demonstrate significant variation in performance metrics such as accuracy, precision, recall, and F1-score across the different models.

From the results, EfficientNet V2 emerges as the top-performing model with an accuracy of 93.95%, precision of 95.02%, recall of 93.95%, and F1-score of 94.01%. This superior performance can be attributed to EfficientNet V2's optimized architecture, which balances depth, width, and resolution for better accuracy and efficiency.

The Vision Transformer (ViT B16) also performed remarkably well, achieving an accuracy of 87.90%, precision of 89.64%, recall of 87.90%, and F1-score of 88.11%. This indicates that the ViT model, with its ability to capture global dependencies in the image data, is well-suited for complex image classification tasks such as brain tumor classification.

In contrast, the Inception-ResNet-v2 and MobileNet V2 models demonstrated lower performance metrics. The Inception-ResNet-v2 model achieved an accuracy of 70.82%, precision of 73.37%, recall of 70.82%, and F1-score of 71.04%. MobileNet V2, known for its lightweight architecture, attained an accuracy of 74.02%, precision of 77.36%, recall of 74.02%, and F1-score of 73.90%. While these models are efficient, their performance is not as robust as EfficientNet V2 and ViT for this specific task.

The discrepancy in performance can be attributed to several factors. EfficientNet V2's scaling strategy allows it to adaptively scale network dimensions, leading to better feature extraction and classification accuracy. ViT's attention mechanisms enable it to handle the spatial hierarchies and complex patterns in MRI images more effectively. On the other hand, while Inception-ResNet-v2 combines the strengths of Inception and ResNet architectures, it might not be as finely tuned for this specific application. Similarly, MobileNet V2's design for mobile and edge devices might limit its capacity to capture intricate details necessary for distinguishing between 44 brain tumor classes.

---

[1] https://www.kaggle.com/datasets/fernando2rad/brain-tumor-mri-images-44c

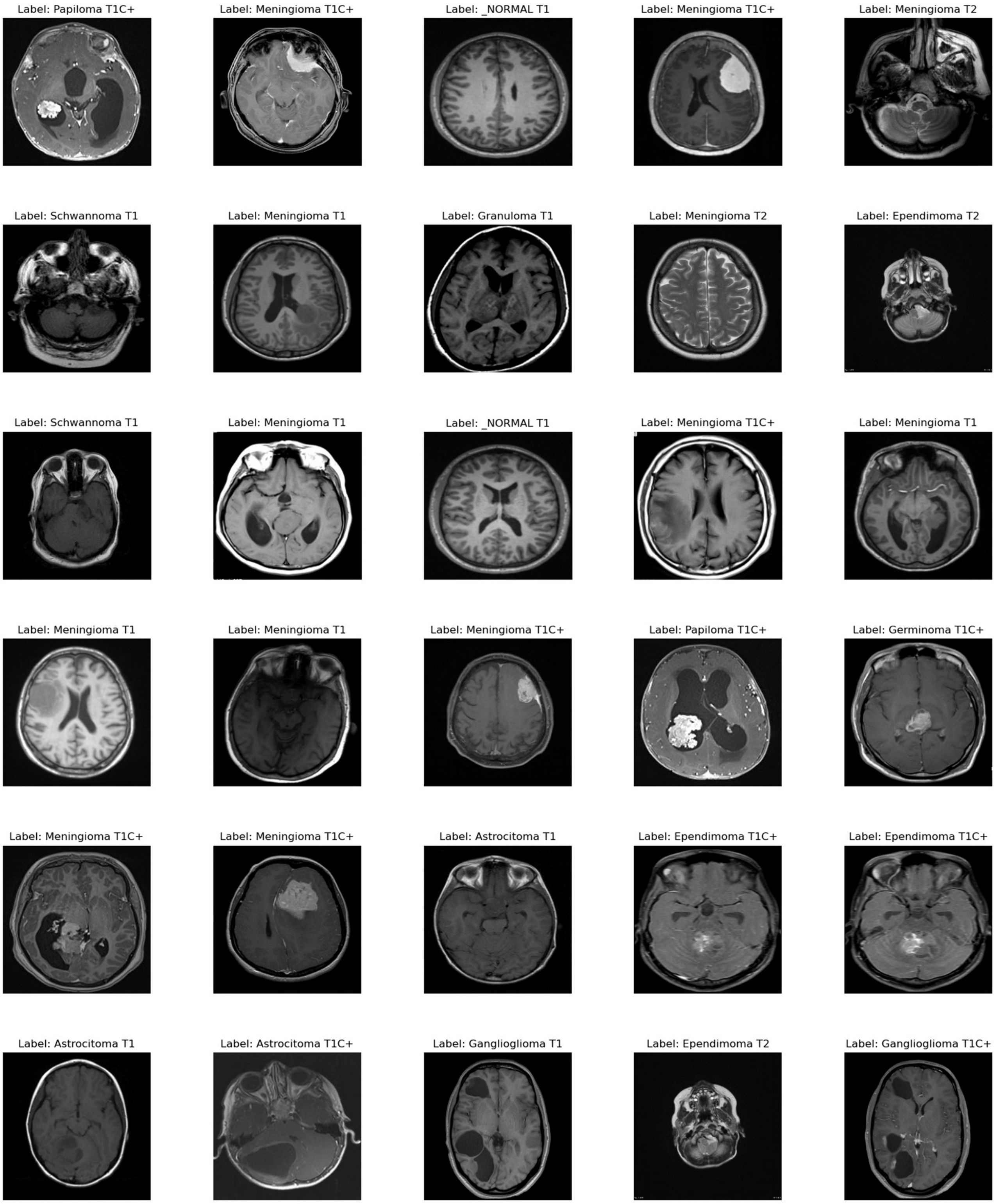


FIGURE 4
Examples of images taken from the dataset.

In conclusion, our findings underscore the importance of model selection in medical image classification tasks. EfficientNet V2 and Vision Transformer (ViT) stand out as particularly effective for classifying brain tumors from MRI images. Future work could focus on further optimizing these models and exploring ensemble methods to enhance classification performance even further.

TABLE 1 Performance metrics for different deep learning models.

	Accuracy	Precision	Recall	f1_score
Inception-resnet-v2	0.708185	0.733747	0.708185	0.710366
Efficientnet-v2	**0.939502**	**0.950218**	**0.939502**	**0.940100**
MobileNet-V2	0.740214	0.773633	0.740214	0.738958
Vision transformer	0.879004	0.896360	0.879004	0.881115

Additionally, incorporating domain-specific data augmentation and preprocessing techniques could potentially improve the accuracy and reliability of the classification results.

3.8 Brain tumor classification based on an hybrid deep learning model

In our study, we aimed to classify MRI images into 44 distinct classes of brain tumors using various state-of-the-art deep learning models. To further enhance classification performance, we employed ensemble learning techniques, combining the Vision Transformer (ViT) model with other models such as Inception-ResNet-V2, EfficientNet V2, and MobileNet V2. The ensemble methods included both average and weighted average techniques, with the latter varying the weights (alpha for the first model and beta for ViT) to optimize performance. Table 2 below summarizes the results.

From the results, the combination of ViT and EfficientNet-V2 using average ensemble learning achieved the highest performance with an accuracy of 95.37%, precision of 96.15%, recall of 95.37%, and F1-score of 95.41%. This demonstrates the robustness of ensemble methods in leveraging the strengths of different models to enhance classification performance.

The weighted average ensemble also showed promising results, particularly with a weight distribution of $\alpha = 0.2$ and $\beta = 0.8$ for the ViT and EfficientNet-V2 combination, achieving an accuracy of 93.95%. This suggests that placing more emphasis on the ViT model, which effectively captures global dependencies, can significantly boost performance.

Interestingly, the ViT and Inception-ResNet-V2 combination also performed well, especially with the weighted average ensemble method ($\alpha = 0.2$, $\beta = 0.8$), achieving an accuracy of 90.04%. However, the performance dropped when the weight distribution favored the Inception-ResNet-V2 model ($\alpha = 0.8$, $\beta = 0.2$), indicating the importance of optimal weight selection in ensemble methods.

Similarly, the combination of ViT and MobileNet-V2 showed improved performance with the weighted average method ($\alpha = 0.2$, $\beta = 0.8$), achieving an accuracy of 88.25%, while favoring MobileNet-V2 ($\alpha = 0.8$, $\beta = 0.2$) resulted in lower performance.

In summary, our findings highlight the efficacy of ensemble learning in brain tumor classification. By integrating different models and optimizing weight distributions, we can significantly improve classification accuracy and other performance metrics. The success of the ViT and EfficientNet-V2 combination, in particular, underscores the potential of using advanced deep learning models and ensemble techniques for complex medical image classification tasks. Future work could explore further optimizations and the incorporation of additional models to continue enhancing classification performance.

3.9 Evaluation of the significance improvement of ensemble learning by weighted selection method for brain tumor classification

In this section, we evaluate the significance of our optimization method in improving the performance of ensemble learning for brain tumor classification. By applying a genetic algorithm to determine the optimal weights for combining deep learning models, we aimed to enhance classification accuracy, precision, recall, and F1-score. The results in Table 3 demonstrate the effectiveness of our approach, highlighting the potential of weighted ensemble learning in advancing medical image classification tasks.

The application of a genetic algorithm to optimize the weights (alpha and beta) for combining deep learning models demonstrated significant improvements in performance metrics for brain tumor classification. The results shown in the Table 3 illustrate the efficacy of the weighted average ensemble learning approach across multiple generations, where each generation represents a different combination of alpha (α) and beta (β) values. The optimized weighted average model with $\alpha = 0.44$ and $\beta = 0.56$ achieved the highest accuracy score of 0.9609, coupled with a precision score of 0.9691, indicating that the genetic algorithm successfully identified a near-optimal balance between the two models, enhancing the overall accuracy and precision of the classification. Across different generations, the results consistently show high values for accuracy, precision, recall, and F1-score, underlining the robustness of the genetic algorithm in optimizing the ensemble weights. For instance, models with $\alpha = 0.44$, $\beta = 0.56$ and $\alpha = 0.73$, $\beta = 0.27$ both achieved an accuracy of 0.9609, reflecting reliable performance improvements. The impact of varying the weights is evident, as seen in the model with $\alpha = 0.17$, $\beta = 0.83$, which achieved a slightly lower accuracy of 0.9395, emphasizing the importance of fine-tuning the weights to leverage the strengths of each model optimally. The top-3 accuracy metric remains consistently high (0.9822) across different weight combinations, indicating the model's robustness and reliability in clinical scenarios. The enhanced performance metrics obtained through the genetic algorithm optimization have significant implications for clinical practice, ensuring more reliable diagnoses critical for treatment planning and patient outcomes, thus providing confidence in the model's predictions and aiding clinicians in making more informed decisions.

3.10 Comparative study

The Table 4 compares our work to several other studies on brain tumor classification, highlighting differences in the methods used, the datasets, the number of classes, and the resulting accuracy.

TABLE 2 Performance metrics for hybrid models.

Hybrid model	Ensemble learning technique	Accuracy	Precision	Recall	f1_score
ViT + InceptionResnet V2	Average Ensemble	0.886121	0.903569	0.886121	0.888955
	Weighted average: ($\alpha = 0.2$, $\beta = 0.8$)	0.900356	0.908926	0.900356	0.900590
	Weighted average: ($\alpha = 0.8$, $\beta = 0.2$	0.779359	0.802382	0.779359	0.781597
ViT + MobileNet V2	Weighted average: ($\alpha = 0.2$, $\beta = 0.8$)	0.882562	0.892685	0.882562	0.882396
	Weighted Average: ($\alpha = 0.8$, $\beta = 0.2$)	0.786477	0.805635	0.786477	0.781143
ViT + EfficientNet V2	Average Ensemble	***0.953737***	***0.961463***	***0.953737***	***0.954080***
	Weighted Average: ($\alpha = 0.8$, $\beta = 0.2$)	***0.950178***	***0.961736***	***0.950178***	***0.950381***
	Weighted Average: ($\alpha = 0.2$, $\beta = 0.8$)	0.939502	0.952808	0.939502	0.941447

TABLE 3 Performance metrics of optimized weighted average models for brain tumor classification using genetic algorithm-selected alpha (α) and beta (β) values.

	Accuracy	Precision	Recall	f1_score	top_3_accuracy
Weighted average ($\alpha = 0.8$, $\beta = 0.2$)	0.950178	0.959987	0.950178	0.951458	0.982206
Weighted average ($\alpha = 0.44$, $\beta = 0.56$)	0.9609	0.9691	0.9609	0.9616	0.9822
Weighted average ($\alpha = 0.17$, $\beta = 0.83$)	0.9395	0.9530	0.9395	0.9415	0.9822
Weighted average ($\alpha = 0.73$, $\beta = 0.27$)	0.9609	0.9673	0.9609	0.9610	0.9822

A key point of comparison is the complexity of the datasets, particularly the number of classes, which has a significant impact on the difficulty of the classification task.

In earlier studies, such as those proposed by Avants et al. (2008); Hinton and Sejnowski (1986), the datasets consist of only three classes. These studies achieved high accuracies, ranging from 88.86% to 95.82%, using CNN-based models. However, with fewer classes, the task is relatively simpler, as the models have fewer categories to differentiate between, making high accuracy more achievable.

Authors in Tomasila and Emanuel (2020), are worked with a dataset of four classes, comprising 53 images per class, and reported accuracies between 92% and 98.41%. The highest accuracy in this group, 98.41%, was achieved by reviewer, who used a combination of EfficientNetV2-M and Inception-V3.

In contrast, our study takes on a much more complex task, utilizing a dataset with 44 classes, significantly increasing the difficulty of the classification problem. Despite this challenge, our model–combining Vision Transformers (ViT) with EfficientNetV2–achieved an accuracy of 96.09%. This is a remarkable result, considering the higher number of classes. Our model needed to differentiate between many more categories, each with potentially subtle differences, making the classification process more demanding.

While some studies report slightly higher accuracy, it's important to recognize that they were working with much simpler datasets. Our model's ability to maintain a high accuracy on a dataset with 44 classes demonstrates its robustness and ability to handle more complex tasks. This makes our approach particularly valuable for real-world applications where models need to classify a wide range of tumor types or conditions with high precision.

4 Conclusion

In this study, we have explored the efficacy of combining different deep learning models, including Inception-ResNet-V2, EfficientNet-V2, MobileNet V2, and Vision Transformers (ViT), for the classification of MRI images into 44 brain tumor classes. Our experiments demonstrated that ensemble learning techniques, particularly weighted average ensembles, significantly enhance classification performance, with the ViT + EfficientNet-V2 model achieving the highest accuracy. Furthermore, by adopting a Genetic Algorithm to optimize the weights in the ensemble models, we were able to achieve even greater improvements in accuracy, precision, recall, and F1-score. These results underscore the potential of hybrid deep learning approaches and optimization algorithms in advancing medical image classification. Future work

TABLE 4 Comparative study between proposed approaches for brain tumor classification.

Study	Method	Dataset	Number of class	Accuracy
Abiwinanda et al. (2019)	CNN	[1]	3	88.86
Sajjad et al. (2019)	VGG-16	–	3	94.58
Deepak and Ameer (2021)	CNN+SVM	–	3	95.82
Asif et al. (2023)	Xception	[2]	4	95.87
İncir and Bozkurt (2024a)	EfficientNetV2-M + Inception-V3	–	4	98.41
Sandhiya and Raja (2024)	ELM+PSO	–	4	97.97
Vankdothu et al. (2022)	LSTM+CNN	–	4	92
Our work	**ViT+ EfficientNetV2**	**1**	**44**	**96.09**

[1] https://figshare.com/articles/dataset/brain_tumor_dataset/1512427. [2] https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri.

will focus on further refining these models and exploring their applicability to other complex medical imaging tasks, ultimately contributing to more accurate and reliable diagnostic tools in healthcare.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.kaggle.com/datasets/fernando2rad/brain-tumor-mri-images-44c.

Author contributions

KG: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. NB: Data curation, Formal analysis, Investigation, Software, Writing – original draft, Conceptualization. KA: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. IL: Conceptualization, Formal analysis, Methodology, Resources, Validation, Visualization, Writing – original draft, Investigation. IA: Investigation, Project administration, Resources, Software, Visualization, Writing – original draft. LA: Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. MM: Data curation, Investigation, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. AS: Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by the Deanship of Graduate Studies and Scientific Research at Jouf University under grant No (DGSSR-2023-02-02064).

Acknowledgments

This work was funded by the Deanship of Graduate Studies Scientific Research at Jouf University under grant No (DGSSR-2023-02-02064).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abd El Kader, I., Xu, G., Shuai, Z., Saminu, S., Javaid, I., and Salim Ahmad, I. (2021). Differential deep convolutional neural network model for brain tumor classification. *Brain Sci.* 11:352. doi: 10.3390/brainsci11030352

Abiwinanda, N., Hanif, M., Hesaputra, S. T., Handayani, A., and Mengko, T. R. (2019). "Brain tumor classification using convolutional neural network," in *World Congress on Medical Physics and Biomedical Engineering 2018: June 3–8,* *2018, Prague, Czech Republic* (Springer), 183–189. doi: 10.1007/978-981-10-9035-6_33

Ansari, A. S. (2023). Numerical simulation and development of brain tumor segmentation and classification of brain tumor using improved support vector machine. *Int. J. Intell. Syst. Applic. Eng.* 11, 35–44. Available at: https://www.ijisae.org/index.php/IJISAE/article/view/2505

Asif, S., Zhao, M., Tang, F., and Zhu, Y. (2023). An enhanced deep learning method for multi-class brain tumor classification using deep transfer learning. *Multimed. Tools Appl.* 82, 31709–31736. doi: 10.1007/s11042-023-14828-w

Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41. doi: 10.1016/j.media.2007.06.004

Aziz, M. J., Zade, A. A. T., Farnia, P., Alimohamadi, M., Makkiabadi, B., Ahmadian, A., et al. (2021). "Accurate automatic glioma segmentation in brain mri images based on capsnet," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)* (IEEE), 3882–3885. doi: 10.1109/EMBC46164.2021.9630324

Babu, P. A., Rao, B. S., Reddy, Y. V. B., Kumar, G. R., Rao, J. N., Koduru, S. K. R., et al. (2023). Optimized CNN-based brain tumor segmentation and classification using artificial bee colony and thresholding. *Int. J. Comput. Commun. Control* 18:4577. doi: 10.15837/ijccc.2023.1.4577

Bansal, T., and Jindal, N. (2022). An improved hybrid classification of brain tumor mri images based on conglomeration feature extraction techniques. *Neur. Comput. Appl.* 34, 9069–9086. doi: 10.1007/s00521-022-06929-8

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020). "End-to-end object detection with transformers," in *European Conference on Computer Vision* (Springer), 213–229. doi: 10.1007/978-3-030-58452-8_13

Dai, Y., Gao, Y., and Liu, F. (2021). Transmed: transformers advance multi-modal medical image classification. *Diagnostics* 11:1384. doi: 10.3390/diagnostics11081384

Deepak, S., and Ameer, P. (2021). Automated categorization of brain tumor from mri using cnn features and SVM. *J. Ambient Intell. Humaniz. Comput*. 12, 8357–8369. doi: 10.1007/s12652-020-02568-w

Dev, C., Kumar, K., Palathil, A., Anjali, T., and Panicker, V. (2019). "Machine learning based approach for detection of lung cancer in dicom ct image," in *Ambient Communications and Computer Systems* (Springer), 161–173. doi: 10.1007/978-981-13-5934-7_15

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Gasmi, K., Ammar, L. B., Krichen, M., Alamro, M. A., Mihoub, A., and Mrabet, M. (2024). Optimal ensemble learning model for dyslexia prediction based on an adaptive genetic algorithm. *IEEE Access*. 12, 64754–64764. doi: 10.1109/ACCESS.2024.3395803

Ge, C., Gu, I. Y.-H., Jakola, A. S., and Yang, J. (2020). Enlarged training dataset by pairwise gans for molecular-based brain tumor classification. *IEEE Access* 8, 22560–22570. doi: 10.1109/ACCESS.2020.2969805

Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H. R., and Xu, D. (2021). "Swin unetr: Swin transformers for semantic segmentation of brain tumors in MRI images," in *International MICCAI Brainlesion Workshop* (Springer), 272–284. doi: 10.1007/978-3-031-08999-2_22

Hinton, G. E., and Sejnowski, T. J. (1986). Learning and relearning in boltzmann machines. *Parallel Distr. Proc.* 1:2.

İncir, R., and Bozkurt, F. (2024a). Improving brain tumor classification with combined convolutional neural networks and transfer learning. *Knowl.-Based Syst.* 299:111981. doi: 10.1016/j.knosys.2024.111981

İncir, R., and Bozkurt, F. (2024b). A study on effective data preprocessing and augmentation method in diabetic retinopathy classification using pre-trained deep learning approaches. *Multimed. Tools Appl.* 83, 12185–12208. doi: 10.1007/s11042-023-15754-7

İncir, R., and Bozkurt, F. (2024c). "A study on the segmentation and classification of diabetic retinopathy images using the k-means clustering method," in *2024 32nd Signal Processing and Communications Applications Conference (SIU)* (IEEE), 1–4. doi: 10.1109/SIU61531.2024.10600987

Jena, B., Nayak, G. K., and Saxena, S. (2022). An empirical study of different machine learning techniques for brain tumor classification and subsequent segmentation using hybrid texture feature. *Mach. Vis. Appl.* 33:6. doi: 10.1007/s00138-021-01262-x

Kang, J., Ullah, Z., and Gwak, J. (2021). Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers. *Sensors* 21:2222. doi: 10.3390/s21062222

Lei, B., Huang, S., Li, R., Bian, C., Li, H., Chou, Y.-H., et al. (2018). Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder-decoder network. *Neurocomputing* 321, 178–186. doi: 10.1016/j.neucom.2018.09.043

Machhale, K., Nandpuru, H. B., Kapur, V., and Kosta, L. (2015). "MRI brain cancer classification using hybrid classifier (SVM-KNN)," in *2015 International Conference on Industrial Instrumentation and Control (ICIC)* (IEEE), 60–65. doi: 10.1109/IIC.2015.7150592

Pereira, S., Pinto, A., Alves, V., and Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in mri images. *IEEE Trans. Med. Imaging* 35, 1240–1251. doi: 10.1109/TMI.2016.2538465

Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C., and Dosovitskiy, A. (2021). "Do vision transformers see like convolutional neural networks?" in *Advances in Neural Information Processing Systems*, 12116–12128.

Rodriguez, R., Dokladalova, E., and Dokládal, P. (2019). "Rotation invariant cnn using scattering transform for image classification," in *2019 IEEE International Conference on Image Processing (ICIP)* (IEEE), 654–658. doi: 10.1109/ICIP.2019.8804467

Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A., and Baik, S. W. (2019). Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *J. Comput. Sci*. 30, 174–182. doi: 10.1016/j.jocs.2018.12.003

Sandhiya, B., and Raja, S. K. S. (2024). Deep learning and optimized learning machine for brain tumor classification. *Biomed. Signal Process. Control* 89:105778. doi: 10.1016/j.bspc.2023.105778

Sekhar, A., Biswas, S., Hazra, R., Sunaniya, A. K., Mukherjee, A., and Yang, L. (2021). Brain tumor classification using fine-tuned googlenet features and machine learning algorithms: IOMT enabled cad system. *IEEE J. Biomed. Health Inform*. 26, 983–991. doi: 10.1109/JBHI.2021.3100758

Sharif, M. I., Li, J. P., Amin, J., and Sharif, A. (2021). An improved framework for brain tumor analysis using MRI based on yolov2 and convolutional neural network. *Complex Intell. Syst*. 7, 2023–2036. doi: 10.1007/s40747-021-00310-3

Steiner, A., Kolesnikov, A., Zhai, X., Wightman, R., Uszkoreit, J., and Beyer, L. (2021). How to train your VIT? Data, augmentation, and regularization in vision transformers. *arXiv preprint arXiv:2106.10270*.

Tanvir Rouf Shawon, M., Shahariar Shibli, G., Ahmed, F., and Saha Joy, S. K. (2023). Explainable cost-sensitive deep neural networks for brain tumor detection from brain MRI images considering data imbalance. *arXiv e-prints arXiv-2308*.

Tomasila, G., and Emanuel, A. W. R. (2020). "MRI image processing method on brain tumors: a review," in *AIP Conference Proceedings* (AIP Publishing LLC), 020023. doi: 10.1063/5.0030978

Vankdothu, R., Hameed, M. A., and Fatima, H. (2022). A brain tumor identification and classification using deep learning based on CNN-LSTM method. *Comput. Electr. Eng*. 101:107960. doi: 10.1016/j.compeleceng.2022.107960

Varuna Shree, N., and Kumar, T. (2018). Identification and classification of brain tumor mri images with feature extraction using dwt and probabilistic neural network. *Brain Infor*. 5, 23–30. doi: 10.1007/s40708-017-0075-5

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *Advances in Neural Information Processing Systems*, 30.

Wenxuan, W., Chen, C., Meng, D., Hong, Y., Sen, Z., and Jiangyun, L. (2021). "Transbts: multimodal brain tumor segmentation using transformer," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer), 109–119. doi: 10.1007/978-3-030-87193-2_11

Williams, T., and Li, R. (2018). "Wavelet pooling for convolutional neural networks," in *International Conference on Learning Representations*, 1. doi: 10.1109/ICINPRO43533.2018.9096860

Wisaeng, K., and Sa-Ngiamvibool, W. (2023). Brain tumor segmentation using fuzzy OTSU threshold morphological algorithm. *IAENG Int. J. Appl. Mathem*. 53, 1–12.

Yu, S., Ma, K., Bi, Q., Bian, C., Ning, M., He, N., et al. (2021). "Mil-vt: multiple instance learning enhanced vision transformer for fundus image classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer), 45–54. doi: 10.1007/978-3-030-87237-3_5

Zacharaki, E. I., Wang, S., Chawla, S., Soo Yoo, D., Wolf, R., Melhem, E. R., et al. (2009). Classification of brain tumor type and grade using mri texture and shape in a machine learning scheme. *Magn. Reson. Med.*62, 1609–1618. doi: 10.1002/mrm.22147

Zahid, U., Ashraf, I., Khan, M. A., Alhaisoni, M., Yahya, K. M., Hussein, H. S., et al. (2022). Brainnet: optimal deep learning feature fusion for brain tumor classification. *Comput. Intell. Neurosci*. 2022:1465173. doi: 10.1155/2022/1465173

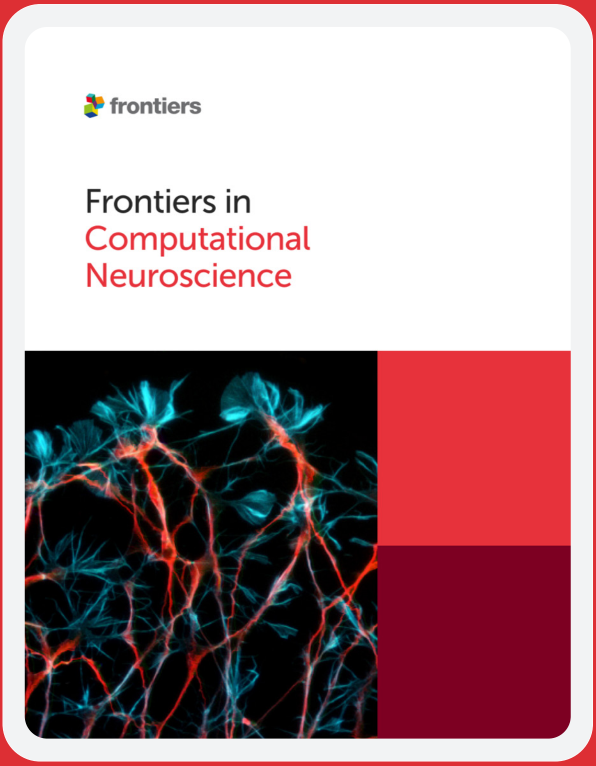
Frontiers in
Computational Neuroscience

Fosters interaction between theoretical and experimental neuroscience

Part of the world's most cited neuroscience series, this journal promotes theoretical modeling of brain function, building key communication between theoretical and experimental neuroscience.

Discover the latest Research Topics

See more →



frontiers

Frontiers in
Computational
Neuroscience

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact