DEEP LEARNING FOR NEUROLOGICAL DISORDERS IN CHILDREN

EDITED BY: Saman Sargolzaei, Ali Fatemi, Fahimeh Mamashli, Anna Farzindar, Yalda Mohsenzadeh and Arman Sargolzaei

PUBLISHED IN: Frontiers in Computational Neuroscience, Frontiers in Neuroscience and Frontiers in Neuroinformatics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-848-0
DOI 10.3389/978-2-83250-848-0

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

DEEP LEARNING FOR NEUROLOGICAL DISORDERS IN CHILDREN

Topic Editors:
Saman Sargolzaei, University of Tennessee at Martin, United States
Ali Fatemi, Kennedy Krieger Institute, United States
Fahimeh Mamashli, Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Harvard Medical School, United States
Anna Farzindar, University of Southern California, United States
Yalda Mohsenzadeh, Western University, Canada
Arman Sargolzaei, Tennessee Technological University, United States

Citation: Sargolzaei, S., Fatemi, A., Mamashli, F., Farzindar, A., Mohsenzadeh, Y., Sargolzaei, A., eds. (2022). Deep Learning for Neurological Disorders in Children. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-848-0

Table of Contents

TYPE Editorial
PUBLISHED 04 August 2022
DOI 10.3389/fncom.2022.984882

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Si Wu,
Peking University, China

*CORRESPONDENCE
Saman Sargolzaei
Saman@utm.edu

RECEIVED 02 July 2022
ACCEPTED 25 July 2022
PUBLISHED 04 August 2022

CITATION
Sargolzaei S (2022) Editorial: Deep learning for neurological disorders in children.
Front. Comput. Neurosci. 16:984882.
doi: 10.3389/fncom.2022.984882

Editorial: Deep learning for neurological disorders in children

Saman Sargolzaei*

Department of Engineering, University of Tennessee at Martin, Martin, TN, United States

KEYWORDS

children, deep learning, neurological disorder, artificial intelligence, machine learning, fMRI, EEG, neuroimaging

Editorial on the Research Topic
Deep learning for neurological disorders in children

Aims and objectives

Children are among the vulnerable age population to the incidents and consequences of various neurological disorders. The vulnerability extends beyond the long-lasting health and socioeconomic impact on the patients and families by introducing unique challenges in studying pediatric neurological disorders. Among the distinct challenges are age-specific technological requirements design considerations and the inter-woven nature of such studies' findings with downstream mechanisms and manifestations related to child growth and development.

Deep learning (DL), a subdiscipline of machine learning and artificial intelligence (AI) in a broader context, has provided a unique prospect to unravel complexities in diverse applications, including medicine and biology. The emergence of DL-based approaches in studying pediatric neurological disorders while relying on high-dimensional data and computing capacity could lead to a better understanding of downstream pathways, more accurate diagnosis, and more efficient management plans.

The Research Topic, entitled *Deep learning for neurological disorders in children*, aimed at conveying scientific studies and discoveries related to the role of DL in studying pediatric neurological disorders. Among the main themes called for the topic were (1) prospects of DL frameworks in diagnosis and prognosis of pediatric neurological disorders and delineating downstream pathways of such disorders and (2) DL-driven opportunities to overcome the challenges associated with studying neurological disorders in children. The Research Topic concluded with eight original research articles (in the order of their publication date, Nemzer et al.; McNorgan; Da Silva et al.; Abdelhameed and Bayoumi; Almuqhim and Saeed; Attallah; Hosseini et al.; Molina-Maza et al.), one brief research report article (Laria et al.) and one review article (Sargolzaei et al.). The following section briefly summarizes these contributions and corresponding developments. However, readers are encouraged to consider the summary as a starting point and refer to each contribution to learn more.

Contributions and developments

Most of the contributions were centered around classifying data between healthy controls and disorders and identifying within disorders classes and disorders-driven events. Targeted disorders included attention deficit hyperactivity disorders (Laria et al.), autism spectrum disorders (Almuqhim and Saeed; Hosseini et al.), dyslexia (McNorgan; Da Silva et al.), epilepsy (Abdelhameed and Bayoumi), and childhood medulloblastoma (Attallah). The developed DL-based classifiers utilized various data modalities, including behavioral assessment (Laria et al.), electroencephalography (EEG) (Abdelhameed and Bayoumi), functional magnetic resonance imaging (fMRI) (McNorgan; Da Silva et al.; (Almuqhim and Saeed), histopathological imaging (Attallah), and facial imaging (Hosseini et al.).

Another interesting main facet of the contributions was the DL utilization for developing simulated neuronal models to aid a better understanding of epileptic seizures (Nemzer et al.) and automatic resolution enhancement of pediatric brain magnetic resonance imaging acquired in downsized time (Molina-Maza et al.). These developments can be viewed as potential pathways DL can overcome technical challenges entangled with studying pediatric neurological disorders (Sargolzaei et al.).

As for data sources, contributions made use of a variety of dataset repositories and initiatives, such as OpenNeuro (reported in McNorgan), BraIns initiative (reported in Da Silva et al.), PhysioNet (reported in Abdelhameed and Bayoumi), ABIDE (reported in Almuqhim and Saeed), IEEE DataPort (reported in Attallah), and Kaggle (reported in Hosseini et al.).

Perspectives and insights

Considering the contributions, Sargolzaei et al. can be regarded as a possible entry gateway supplying the reader with history and a scoping review of DL's role in the current field of focus before the presented topic. Furthermore, the contributions exemplify a wide range of DL solutions in terms of theoretical and methodological advancements and enabling decision aids in diagnosing pediatric neurological disorders.

Nemzer et al. discussed Hodgkin-Huxley simulations on small-world network topology to comprehend neuronal behavior linked to epileptic seizures. Such developments will ease fine-tuned research for specific neurological conditions and expand access to the collection of synthetic data for training purposes. Molina-Maza et al. developed a DL-based scheme for automatic resolution enhancement of pediatric brain MRI acquired in downsized time, aiming to advance free-sedation pediatric brain imaging protocols.

Tying the DL realm with classical statistical models is an exciting domain of investigation. Laria et al. introduced dlasso, a neural network version of the lasso algorithm for studying ADHD symptom severity. Regarding probing neural mechanisms and decision aids in learning abilities, McNorgan employed DL joined with brain functional connectivity network to examine reading-skill-dependent mechanisms and their impact on cognitive domains.

Access to domain-specific tools that interpret DL models can lead to a more informed understanding. Da Silva et al. developed a tool called VisualExpplanations, to classify developmental dyslexia while unraveling brain mechanisms insights facilitating the DL models findings interpretation in conjunction with neuroscientific knowledge.

The power of DL solutions is optimized when the structure and pipeline are carefully planned. Abdelhameed and Bayoumi proposed a DL approach for the task of automatic EEG-based seizure detection in pediatric patients while emphasizing the significance of structure fine-tuning and attention to data elements. Attallah developed the CoMB-Deep pipeline to classify childhood medulloblastoma and identify tumor sub-classes based on histopathological images.

Timely diagnosis of pediatric neurological disorders necessitates using all existing data modalities. Almuqhim and Saeed developed ASD-SAENet for the task of ASD classification using fMRI data. Autism classification was also the focus of Hosseini et al., where a MobileNet-based DL model was developed using facial image analysis.

Remarks and recommendations

While the route forward is exhilarating for applying DL models in studying pediatric neurological disorders, it shouldn't be conceived as a magic crystal ball. Instead, the more sharing practices we implement in collected data and developed models, the higher the chances are for realizing DL as one clinically approved solution that outperforms traditional diagnostic and prognostic solutions.

Among the relatively untouched areas of investigation are DL frameworks enabling the fusion of multi-modalities and taking advantage of basic science experimentations to guide the structure and implementation of DL models in studying neurological disorders in children.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

We wish to acknowledge the tremendous effort that authors, reviewers, and editors have put together to make the Research Topic a step forward in advancing scientific discoveries in studying pediatric neurological disorders.

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

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers
in Computational Neuroscience

ORIGINAL RESEARCH
published: 08 January 2021
doi: 10.3389/fncom.2020.583350

Critical and Ictal Phases in Simulated EEG Signals on a Small-World Network

Louis R. Nemzer[1]*, Gary D. Cravens[2], Robert M. Worth[3], Francis Motta[4], Andon Placzek[5], Victor Castro[1] and Jennie Q. Lou[6]

[1] Department of Chemistry and Physics, Halmos College of Arts and Sciences, Nova Southeastern University, Fort Lauderdale, FL, United States, [2] Department of Health Informatics, Dr. Kiran C. Patel College of Osteopathic Medicine, Nova Southeastern University, Fort Lauderdale, FL, United States, [3] Department of Mathematical Sciences, Indiana University – Purdue University Indianapolis, Indianapolis, IN, United States, [4] Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, FL, United States, [5] Department of Medical Education, Dr. Kiran C. Patel College of Allopathic Medicine, Nova Southeastern University, Fort Lauderdale, FL, United States, [6] College of Medicine and Health Sciences, Khalifa University, Abu Dhabi, United Arab Emirates

Healthy brain function is marked by neuronal network dynamics at or near the critical phase, which separates regimes of instability and stasis. A failure to remain at this critical point can lead to neurological disorders such as epilepsy, which is associated with pathological synchronization of neuronal oscillations. Using full Hodgkin-Huxley (HH) simulations on a Small-World Network, we are able to generate synthetic electroencephalogram (EEG) signals with intervals corresponding to seizure (ictal) or non-seizure (interictal) states that can occur based on the hyperexcitability of the artificial neurons and the strength and topology of the synaptic connections between them. These interictal simulations can be further classified into scale-free critical phases and disjoint subcritical exponential phases. By changing the HH parameters, we can model seizures due to a variety of causes, including traumatic brain injury (TBI), congenital channelopathies, and idiopathic etiologies, as well as the effects of anticonvulsant drugs. The results of this work may be used to help identify parameters from actual patient EEG or electrocorticographic (ECoG) data associated with ictogenesis, as well as generating simulated data for training machine-learning seizure prediction algorithms.

Keywords: epilepsy, epileptic seizures, epileptogensis, small-world networks, simulation—computers, neuron, criticality, phase transition

OPEN ACCESS

Edited by:
Ali Fatemi,
Kennedy Krieger Institute,
United States

Reviewed by:
Qing Yun Wang,
Beihang University, China
Bela Turk,
Kennedy Krieger Institute,
United States

****Correspondence:***
Louis R. Nemzer
louisiqn@gmail.com

Received: *14 July 2020*
Accepted: *04 December 2020*
Published: *08 January 2021*

Citation:
Nemzer LR, Cravens GD, Worth RM, Motta F, Placzek A, Castro V and Lou JQ (2021) Critical and Ictal Phases in Simulated EEG Signals on a Small-World Network. Front. Comput. Neurosci. 14:583350. doi: 10.3389/fncom.2020.583350

INTRODUCTION

The human brain must remain sensitive to new stimuli and coordinate spatially distant information processing modules to function optimally in a continuously changing environment. To accomplish this, the brain needs to dynamically operate at or near a critical state (Beggs and Timme, 2012). A failure to remain at this "edge of chaos" (Waldrop, 1992), a point poised between insensitivity and hyper-synchronization, can lead to neurological disorders, including epilepsy (Meisel et al., 2012).

A biological system, such as the human brain, in the critical state near a phase transition, is maximally sensitive to external influences (Larremore et al., 2011), and is thus most efficient in amplifying small perturbations. Such a state also allows for long-range coordination between brain regions, with a theoretical correlation length that diverges to infinity. At this intermediate critical phase, neuronal bursts exhibit power law statistics (Chialvo, 2010), with no typical spatial or temporal scale, unlike the seizure or subcritical phases. The properties of a system that are poised at the critical state of a phase transition may be very different than those of states on either side. In particular, the sensitivity of the system to external perturbations and information processing power is often maximized at this point. For example, in the case of magnetic susceptibility, at a critical temperature, flipping a single spin may lead to a cascade of magnetization changes that is not possible at either very low or very high temperatures. This "critical brain hypothesis" that links the physics of phase transitions with neuroscience was first introduced by Alan Turning seven decades ago, and has grown in acceptance due to increasing empirical and theoretical support (Turing, 1950). The theory has been questioned because of the apparent difficulty of satisfying the requirement for the parameters to be fine-tuned within a very tiny region of parameter space corresponding to the critical phase, as this would be very unlikely to occur by chance alone. However, recent research has demonstrated that the brain uses active homeostatic mechanisms (Beggs, 2019), including synaptic rewiring based on spike-timing dependent plasticity, to remain at the critical state (Shin and Kim, 2006; Ma et al., 2019).

BACKGROUND

Epilepsy is one of the most common central nervous system (CNS) diseases (Zack and Kobau, 2017), affecting ∼50 million individuals worldwide, including both men and women of varying ages. It is a chronic neurological disorder characterized by a persisting predisposition to generate epileptic seizures and by the resultant neurobiological, cognitive, psychological, and social consequences. An epileptic seizure is a sudden, transient, and uncontrolled electrical disturbance in the brain. The signs and symptoms are caused by abnormal excessive or synchronous neuronal activity. Seizures can cause sudden changes in patient behavior, movements, feelings, or in levels of consciousness. Most patients have little or no warning before a seizure occurs, and this unpredictability can have profound impacts on in their lifestyle, including restrictions on driving, or constraints on employment opportunities.

Epilepsy research has enabled remarkable progress in broadening our understanding of the etiologies and mechanisms leading to epilepsy and its associated comorbidities. It has also brought interventions and treatments to improve the management of seizures and their comorbid conditions and consequences. The prognosis for medical seizure control is good, with over 70% of patients achieving remission. Meanwhile, 30% of individuals with epilepsy remain uncontrolled with increased risk of adverse events and lifestyle disruptions. As a result, over the last three decades, researchers have aimed to gain insight into the underlying mechanisms of epilepsy in these patients with uncontrolled symptoms, hoping to identify biomarkers of disease activity that would indicate an impending seizure and allow them to take protective action or perhaps initiate some therapeutic intervention (Mormann et al., 2007).

The critical brain hypothesis is one approach to understanding the onset and lack of control in these epilepsy patients. Epilepsy is believed to occur when the human brain is unable to dynamically operate in or near this critical state. According to this hypothesis, the intermediate critical phase, as opposed to the ictal or subcritical phases, exhibits scale-free phenomena—with no typical spatial or temporal scale. This behavior can be quantified using a power law functional form of the number of simultaneously firing neurons. This differs from the ictal state, which consists of "all-or-nothing" pathological synchronization, and the subcritical state that has only locally disjoint firing with no long-range coordination. Power-law dynamics are a necessary, but not sufficient, observation to conclude that a system is in its critical state. More stringent tests for the relationship between the scaling exponents for the spatial and temporal sizes of bursts have been developed to distinguish actual critical behavior from other phenomena that can also give rise to power laws (Friedman et al., 2012).

To better understand this phenomenon, we employed the Hodgkin-Huxley equations (Hodgkin and Andrew, 1952) for modeling the dynamics of neurons. The model is computationally intensive, and as such, has usually been limited to simulating small networks of neurons for short time periods. Even with these constraints, distinct phases can be distinguished by plotting a histogram of the number of simultaneously firing neurons for each simulation time step, with the ictal phase associated with significant deviations from power-law behavior. This is particularly true when the network topology is chosen to be Small World (SW) (Humphries and Gurney, 2008). SW networks have many local connections between nearby neurons, and a few long-range bridges. It is widely believed that the human brain possesses many aspects of SW architecture (Bassett and Bullmore, 2006). This allows for the modularity of high clustering coefficients to coexist with the rapid and efficient communication of short path lengths. In the case of focal epilepsy, it is possible that these adaptations become harmful. A seizure focus is thought to recruit connected neurons into a growing synchronized cluster. This process nucleates uncontrolled growth that can rapidly spread on SW networks (Hong et al., 2002). The natural refractory period inherent in neural spiking normally protects the brain from reaching this synchronized ictal phase. However, when connectome plasticity increases the synaptic weights between neurons—either because of trauma, the reinforcement of neural pathways from previous seizures, or channelopathies that lengthen the action potential—these protections can fail.

Therefore, our research efforts focus on access to high-resolution data, in this case, the simulated voltages of individual neurons, in contrast with patient measurements that only capture averaged field potentials at best. This allows us to easily classify the phase—silent, exponential, power, or ictal—to better understand the process of ictogenesis. Using full HH simulations on an SW network, we have generated synthetic EEG signals with intervals corresponding to seizure (ictal) or non-seizure (interictal or subcritical) states that can occur based on the hyperexcitability of the artificial neurons and the strength and topology of the synaptic connections between them. By

TABLE 1 | Model parameters for the THINKER 1.0 Simulation.

Parameter	Symbol	Value
Normalized membrane capacitance	C	1
Maximum Na^+ conductivity	g_{Na}	20
Maximum K^+ conductivity	g_K	12
Maximum leak conductivity	g_{Leak}	0.05
Na^+ reversal voltage	V_{Na}	50 mV
K^+ reversal voltage	V_K	−90 mV
Leak reversal voltage	V_{Leak}	−60 mV
Na^+ deactivation "stickiness"	S	*Varies*
Minimum connectivity weight	g_{min}	0
Maximum connectivity weight	g_{max}	*Varies*
Initial voltage	V_0	−64 mV
Time step	dt	0.1 ms

generating this synthetic data, we aim to train machine learning (ML) algorithms for predicting seizures in patients with epilepsy. Due to the difficulty in acquiring high-quality patient data, and the relative rarity of seizures events compared with interictal intervals, the ability to generate synthetic data may alleviate an important bottleneck in seizure prediction methods (Aznan et al., 2019). Rapid advances in computational power permit more realistic full HH simulations that were considered not practical only a few years ago. Overall, the goals here are twofold. First, to better understand the causal biophysical abnormalities of the ictal transition. Second, to generate a set of surrogate EEG data for use as input to ML algorithms which can then leverage their power to identify this interval and rigorously characterize the spatiotemporal characteristics. Since the biophysical changes generating the surrogate data are known, it should be helpful in "reverse engineering" the ML results to better understand possible seizure prediction strategies.

METHOD

HH neurons in a SW network have intrinsic regulatory features, including threshold-and-fire activity and refractory periods. As a result, they are comparable with the archetypal self-organized criticality (SOC) (Rubinov et al., 2011) situations of sandpile avalanches and earthquakes. In each of these cases, SOC is a natural consequence of the underlying balance of slow and fast variables corresponding to loading and release, respectively (Dickman et al., 1998, 2000; Gal and Marom, 2013).

Computational models of isolated HH neurons show that criticality requires exquisite fine-tuning (Chua, 2013) of processes, like the slow inactivation of sodium channels, in order for the cell to remain excitable (Ori et al., 2018) but not oscillatory. Based on the theory of percolation phase transitions (Breskin et al., 2006; Zhou et al., 2015), in which adjacent nodes of a network are connected with probability *p*, the number distribution, *n,* of clusters of size *s*, obeys the proportion:

$$n(s,p)\ \alpha\ s^{-\tau}e^{-s/s_0}$$

Where s_0 is a function of p. For values of p below the critical percolation threshold p_c, there is a typical cluster size s_0. The formation of larger clusters is strongly suppressed, since the exponential function dominates the power law except for very low values of *s*. However, at the critical threshold, the value of s_0 diverges to infinity. This means that there is no longer a "typical" cluster size—the relation has become scale free—resulting in a power-law distribution with many small clusters, fewer medium clusters, and a small number of large clusters:

$$n(s,p=p_c)\ \alpha\ s^{-\tau}$$

Extracted burst sizes show a peak around the physiological value of the Na^+ inactivation ion-channel gating parameter. Coordination both within and between specific brain modules is thought to be maximized at or near the critical state, since activity at all length scales becomes important. This is not possible in the hypercritical (Netoff et al., 2004) ictal state seen in epilepsy, when global synchronization overwhelms everything else.

To simulate the dynamics of the small-world neuron networks, we introduce the "Theoretical HH Ion-Gated Network Connectome Electroencephalographic Replicator" (THINKER) 1.0 and 2.0 mathematical models. THINKER version 1.0 simulations were instantiated in *Mathematica*. First, the Watts-Strogatz algorithm (Watts and Strogatz, 1998) was used to generate the SW directed-network architecture. The connectivity matrix and directionality were fixed for all runs, but the weights could vary with uniform probability from 0 to g_{max}. The HH coupled differential equations we implemented for each neuron using Euler's method in 0.1 ms time steps to find the voltage *V* are:

$$-C\frac{dV}{dt}=m^3hg_{Na}\left(V-V_{Na}\right)+n^4g_K\left(V-V_K\right)+g_{leak}\left(V-V_{leak}\right)+I_{inject}$$

Where m, n, and h, are the voltage-dependent gating parameters for Na^+ activation, K^+ activation, and Na^+ inactivation, respectively. The term I_{inject} represents incoming synaptic stimuli from connected neurons, and C is the membrane capacitance (**Table 1**). The response of these parameters to changes in voltage are controlled by the opening (α) and closing (β) rate-constants for individual gates in the neuron's membrane:

$$\frac{dm}{dt}=\alpha_m\left(1-m\right)-\beta_m m$$
$$\frac{dn}{dt}=\alpha_n\left(1-n\right)-\beta_n n$$
$$\frac{dh}{dt}=\alpha_h\left(1-h\right)-\beta_h h$$

This implies the steady state values for z = {m,n,h} are:

$$z_\infty=\frac{\alpha_z}{\alpha_z+\beta_z}$$

with corresponding rate constants, in inverse seconds, assuming constant voltage:

$$k_z = \frac{1}{\alpha_z + \beta_z} = 1/\tau_z$$

The "stickiness" of sodium inactivation gates can be modeled by multiplying α_h and β_h by the same constant s:

$$\alpha_h^{'} = s\alpha_h$$
$$\beta_h^{'} = s\beta_h$$

This decreases the rate constant for h, which is the variable describing the gates that inactivate the sodium channels when the neuron is supposed to return to its resting state after firing. The result is an elongation of the characteristic time for h to equilibrate to changes in voltage, from τ_h to $s\tau_h$, while keeping the steady state value of h unchanged.

The explicit forms of the α and β parameters for neuron voltages in millivolts are given by:

$$\alpha_m(V) = 0.1\frac{V+40}{1-Exp[-0.1(V+40)]}$$
$$\beta_m(V) = 4\,Exp[-0.05(V+65)]$$
$$\alpha_n(V) = -0.01\frac{V+55}{Exp[-0.1(V+55)]-1}$$
$$\beta_n(V) = 0.25\,Exp[-0.0125(V+65)]$$
$$\alpha_h(V) = s\,0.07\,Exp[-0.05(V+65)]$$
$$\beta_h(V) = \frac{s}{Exp[-0.1(V+35)]+1}$$

A hyperbolic tangent is used for the transfer function for synaptic connections. One neuron is "voltage clamped" to a white-noise source that drives the system without imparting a characteristic frequency. THINKER 2.0 is written in *Python* and follows Ermentrout and Terman (Ermentrout and Terman, 2010) with the same structure but slightly different parameters. The simulation outputs the voltages of each simulated neuron at every time step. The data are converted into a raster array that records the firing of each neuron when its voltage exceeds a preset threshold. The phases can be labeled by fitting a histogram of simultaneously firing neurons, to either an exponential or power-law function, following the established methods for identifying critical behavior. These neuronal-level, voltage-resolved data are generally not available to the ML algorithm, which would as a matter of practice only have access to EEG or electrocorticographic (ECoG) data.

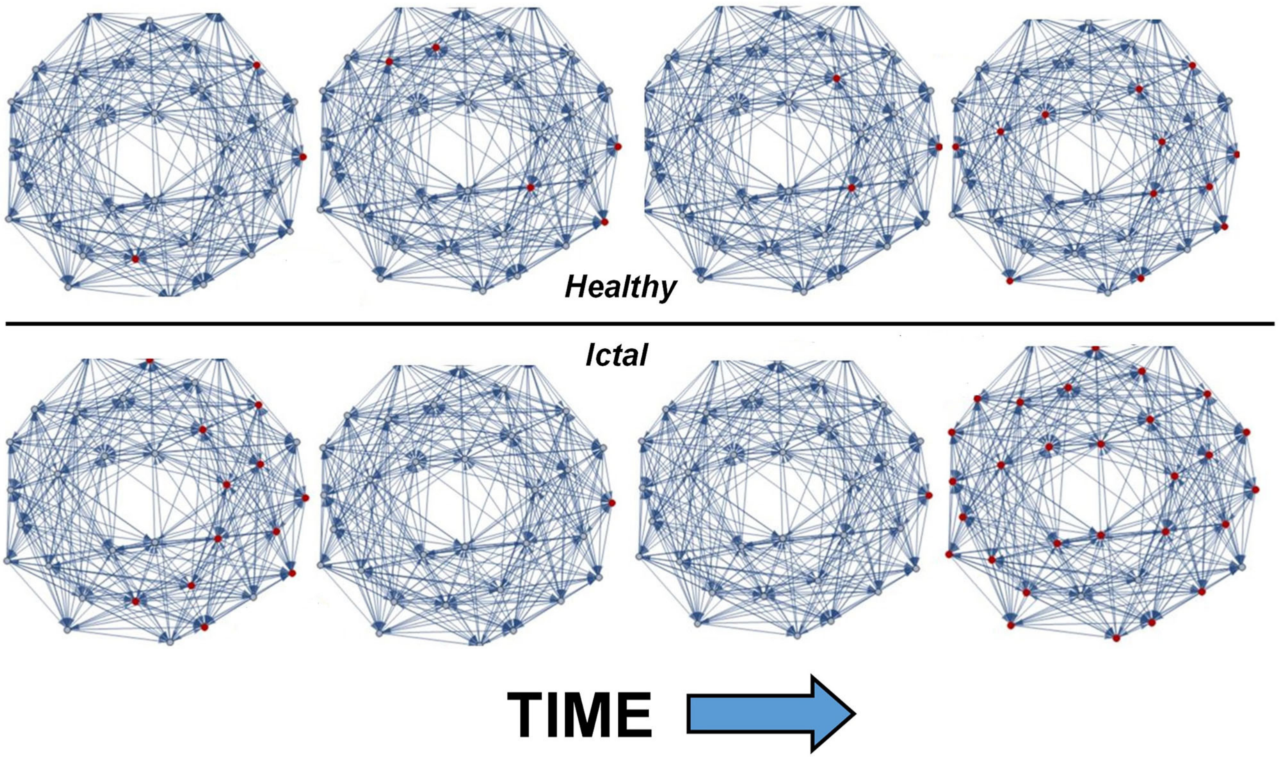

FIGURE 1 | Visual representation of THINKER 1.0 simulation time evolution showing healthy (top) and ictal (bottom) phases. The synaptic connections between neurons are shown in blue, and the firing neurons are marked red. Time increases to the right, and the rightmost neuron is "voltage-clamped" to a white noise source. In the healthy/critical phase, the activity follows a power-law distribution, so "bursts" of all sizes are possible. By contrast, in the ictal/supercritical phase, "all-or-nothing" pathological synchronization is observed.

RESULTS

Figure 1 shows the output of THINKER 1.0 simulations on small-world networks with the firing neurons labeled in red. In the healthy phase (top), the activity has bursts with a distribution of sizes, with coordination on local and extended length scales. In contrast, the supercritical ictal state (bottom) has globally synchronized firing with timesteps that have almost all the nodes either on or off simultaneously. In **Figure 2**, the difference between these phases can also be demonstrated using the simulated EEG. The left panels for the healthy and seizure state show the individual neuron voltages, and the middle graphs display the overall mean voltage at each time step. These are called "simulated" or "synthetic" EEGs, although they can be compared with either conventional scalp electrocochleographic (EEG) or intracranial electrocorticographic (ECoG) patient data. The right panels show the calculated histograms of the number of simultaneously firing neurons. The upper subfigure is linear, while the lower subfigure is log-log scale so that power laws will be represented by straight lines. In the healthy critical phase, the histogram follows a power-law form, and the occurrence of more than 10 simultaneously firing neurons is suppressed. By contrast, in the ictal seizure state, the histogram has a spike around 12 simultaneous neurons, corresponding to near complete synchronization.

To further visualize the different brain phases, an image of a 3D-printiable file is shown in **Figure 3**. Here, the voltages at each time times for all neurons are represented by the height of the model. The difference between the complexity of the critical healthy state and the repetitive synchronization of the ictal state can be observed.

Figure 4 (top) shows a THINKER 2.0 small world network near the critical state at different time points. An animated movie version is available as part of the **Supplementary Material**. Below are the corresponding synthetic EEG, wavelet transform scalogram, and power spectral density.

The wavelet transform is accomplished by convolving a set of orthogonal wavelet "chirps" that have identical shape, but different scaling factors, with the data (Akansu et al., 2010). The scalograms produced are similar to the spectrograms created using short-time Fourier transforms. The primary difference is the dynamic way in which the inherent tradeoff between temporal and frequency resolution is handled. The wavelet transform uses basis functions that are localized in both time and frequency. As a result, at low frequencies scalograms possess good frequency resolution at the expense of poor temporal resolution. Conversely, high frequencies enjoy good time resolution but reduced frequency resolution. Wavelet transforms reveal the complexity of activity present in actual human brain function, particularly in the high frequency regions. These can be interpreted as non-periodic neuronal bursts of

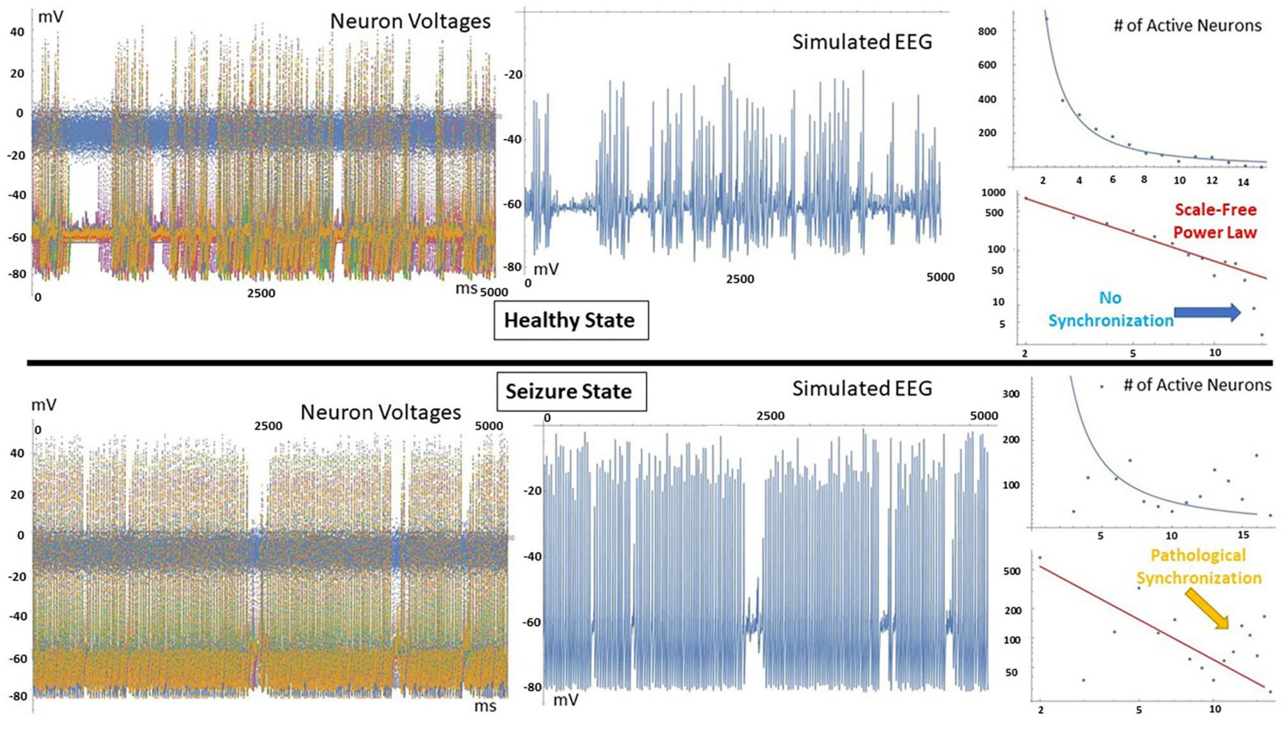

FIGURE 2 | (Left) All simulated neuron voltages. The blue trace is the voltage-clamped neuron driven by white noise. (Middle) Resulting synthetic EEG, calculated at each time step as the mean value of all neurons. (Right) Corresponding histogram showing the number of time steps that have that number of simultaneously firing neurons. In the healthy state, a power-law relationship is observed, and time steps that show near complete synchronization are very rare. In contrast, the seizure state has a spike at high node numbers, meaning that pathological synchronization occurs.

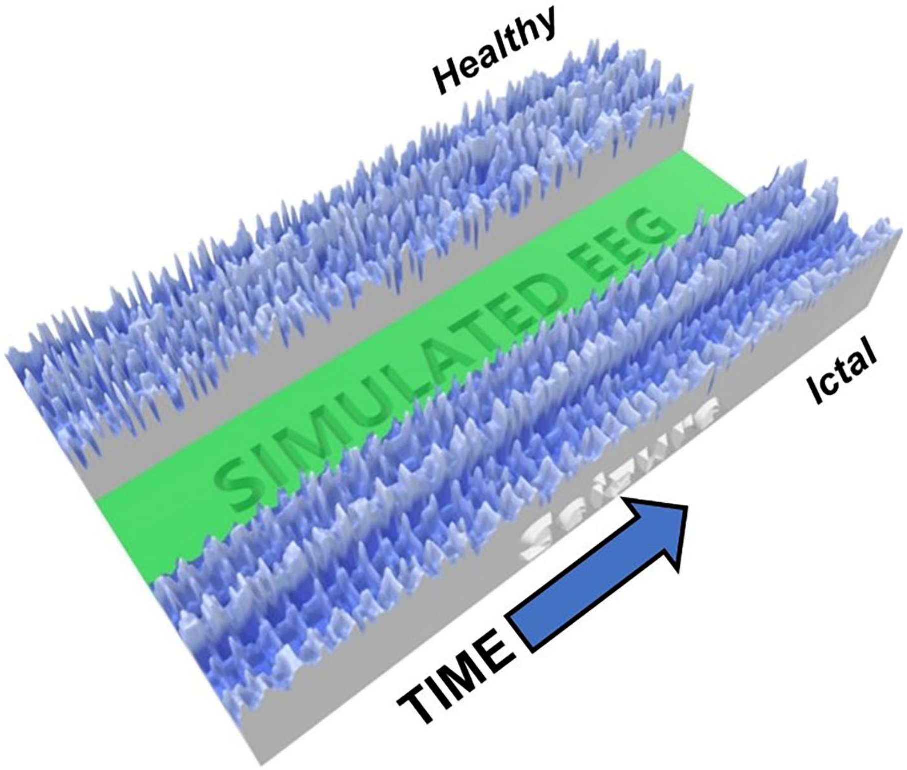

FIGURE 3 | Image of 3D-Printed model showing an example of critical (healthy, top) and ictal (seizure, bottom) phases. The heights represent the neuron voltage for each time step. The synchronization in the seizure state is readily visible, while the signals in the healthy state are much more varied.

varying sizes within these frequency bands, as observed with normal brain function. In the high frequency region of the power spectral density, the data closely follows a power law with exponent -3.6, as marked by the orange line. This agrees with previous measurements of human brain activity (Miller et al., 2009).

A comparison of the different phases possible from THINKER 2.0 simulations is shown in **Figure 5**, where the synthetic EEG, corresponding wavelet transform, and power spectral density are shown for the subcritical, critical, and ictal regimes. Again, the critical phase shows the most complex activity in the wavelet transform, and most closely follows a power law (green line) in the power spectral density plot. The ictal state has a prominent spike (black arrow) representing synchronization.

The differences between the subcritical/interictal, critical, and ictal phases are easy to see when the scalograms are converted into a 3D-printed representation (**Figure 6**). The subcritical phase ("EXP" for exponential) has too little activity overall, while the ictal phase is highly locked into a single pattern. Only the critical phase ("Power" for power law) has the complexity, in both the low and high frequency bands, to capture the neurocorrelates of healthy cognition.

The uniqueness of the critical state also appears in the histograms of simultaneously firing neurons as seen in **Figure 7**. Here, the network topology was frozen, and only the stickiness was varied. The extracted values of the mean cluster sizes are shown in **Figure 7A**. As predicted by percolation theory, subcritical states will have large clusters exponentially suppressed, while the mean cluster size diverges in the critical state, leaving a scale-free power-law relationship. Here, the g_{max} represents the maximum synaptic weight of connected neurons, and the "sticky" parameter again controls the Na^+ channel inactivation after each neuron fires. The finding that the critical state occurs with a "sticky" value of 1.05, with 1.0 corresponding to the physiological value, agrees with the concept that the brain is regulated to be at or slightly below the critical threshold (Beggs, 2018). In the histograms of simultaneously firing neurons shown in **Figure 7B**, a power law will produce a straight line, while an exponential function will curve downward. Only the histogram for the critical state exhibits a power law tail.

Figure 8 shows actual ECoG patient data collected from intracranial electrodes. The readings were taken as part of preoperative testing prior to epilepsy surgery. A clinician marked time intervals as either interictal (not seizure) or ictal (seizure). Here, a peak in the power spectral density around 8 Hz is seen in the ictal but not interictal data. The shift in peak frequency compared with the simulations may reflect differences in the natural oscillation periods for the modeled network using the chosen parameters.

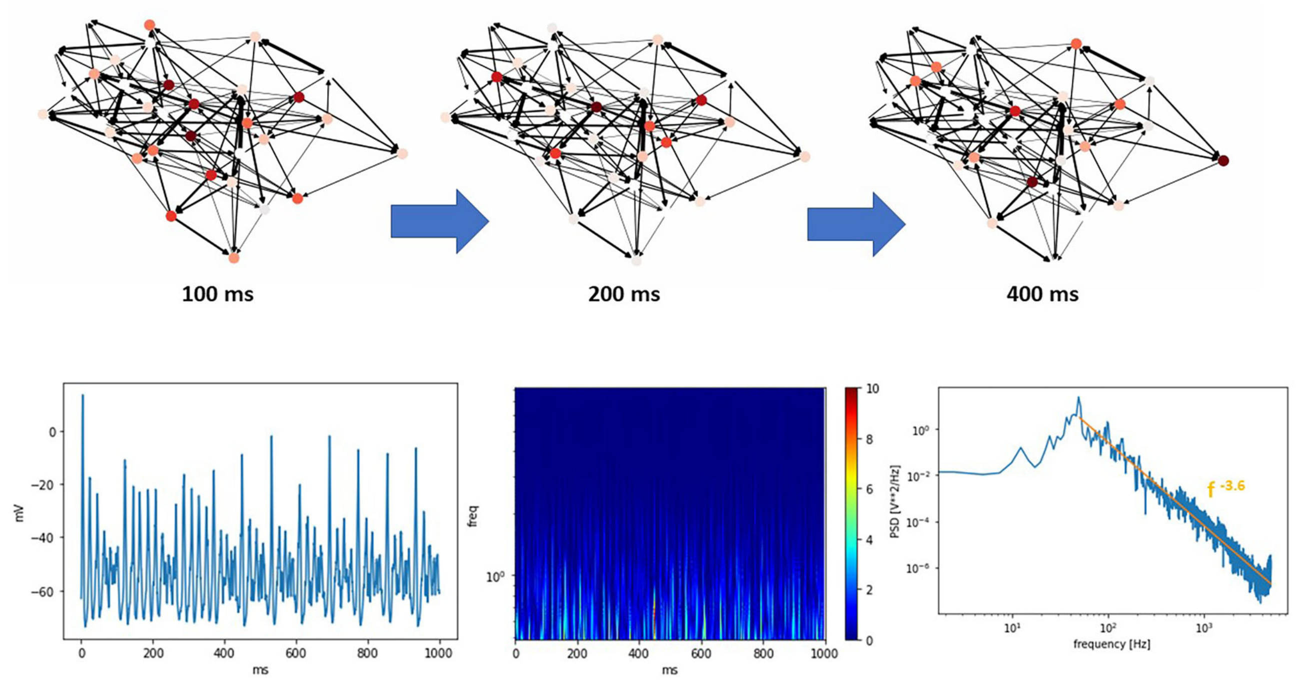

FIGURE 4 | (Top) A THINKER 2.0 Small World network showing critical phase behavior over time. Local and global coordination between bursts is seen, as observed in other systems with SW topologies. An animated movie version is available in the **Supplementary Material**. (Bottom) The corresponding simulated EEG, scalogram, and power spectral density plots. In the power spectral density plot, the data at high frequencies closely follow a power law with exponent −3.6 (orange line).

DISCUSSION

Understanding ictogenesis, the cascade of biophysical events that culminates in a clinically expressed epileptic seizure, has remained an elusive goal, of special concern for those epilepsy patients whose seizures remain uncontrolled with currently available drugs and approaches. While mechanisms of disease activity have been proposed and most seizures remain well-controlled, a significant subset of patients remain at high risk for adverse events and require lifestyle adjustment that negatively impacts their ability to perform activities of daily living. It is thought that epilepsy occurs when the human brain is unable to dynamically operate in or near this critical state, where certain biological neuronal networks work near phases of pathological synchronization and insensitivity. A failure to remain at this point poised between instability and homeostasis can lead to epilepsy. One of the most prominent approaches to understanding the underlying mechanisms of epileptic seizure activity has been mathematical modeling of the relevant physiological processes. The formalism described by Hodgkin and Huxley in 1952 provides an elegant mathematical description of neuronal behavior and is considered the gold standard for describing neuronal physiology. However, the computational complexity involved in modeling systems of neurons large enough to exhibit meaningful behavior have up until now rendered models of this type computationally intractable. As a result, previous such approaches have generally worked with some sort of approximated forms, the assumptions of which may vitiate some of the conclusions drawn from such models. Meanwhile, advances in computational power have now made it feasible to perform simulations using full HH models describing a biophysically relevant number of neurons. In this paper, we described THINKER 1.0 and THINKER 2.0, which implement full HH model simulations with the neurons configured in a SW topology. It goes without saying that no simulation, especially with orders of magnitude fewer nodes, can reflect all of the complexity of the human brain, and we have focused on aspects more relevant for future machine learning algorithms for predicting seizures in patients with epilepsy. Here, the simulated networks reproduce both the power-law tail at high frequencies seen in the power spectral density of patient data at times of normal cognition, as well as the peaks corresponding to pathological synchronization that occur during seizures.

The work to date has yielded some interesting insights into the process of ictogenesis. The results of this modeling work identified three brain states—subcritical/interictal (non-seizure state, with little or no activity), critical (non-seizure state, with the number of active neurons characterized by a power law distribution), and supercritical (ictal seizure state, displaying pathological synchronization of large numbers of neurons). The model biophysical parameter whose variation determines in which state the system exists is h in the HH formalism, which describes the process of Na^+ inactivation. On an individual neuronal basis, alteration of this parameter will affect the rate of neuronal repolarization, which in turn alters the spike frequency relative to other neurons, which can affect neuronal

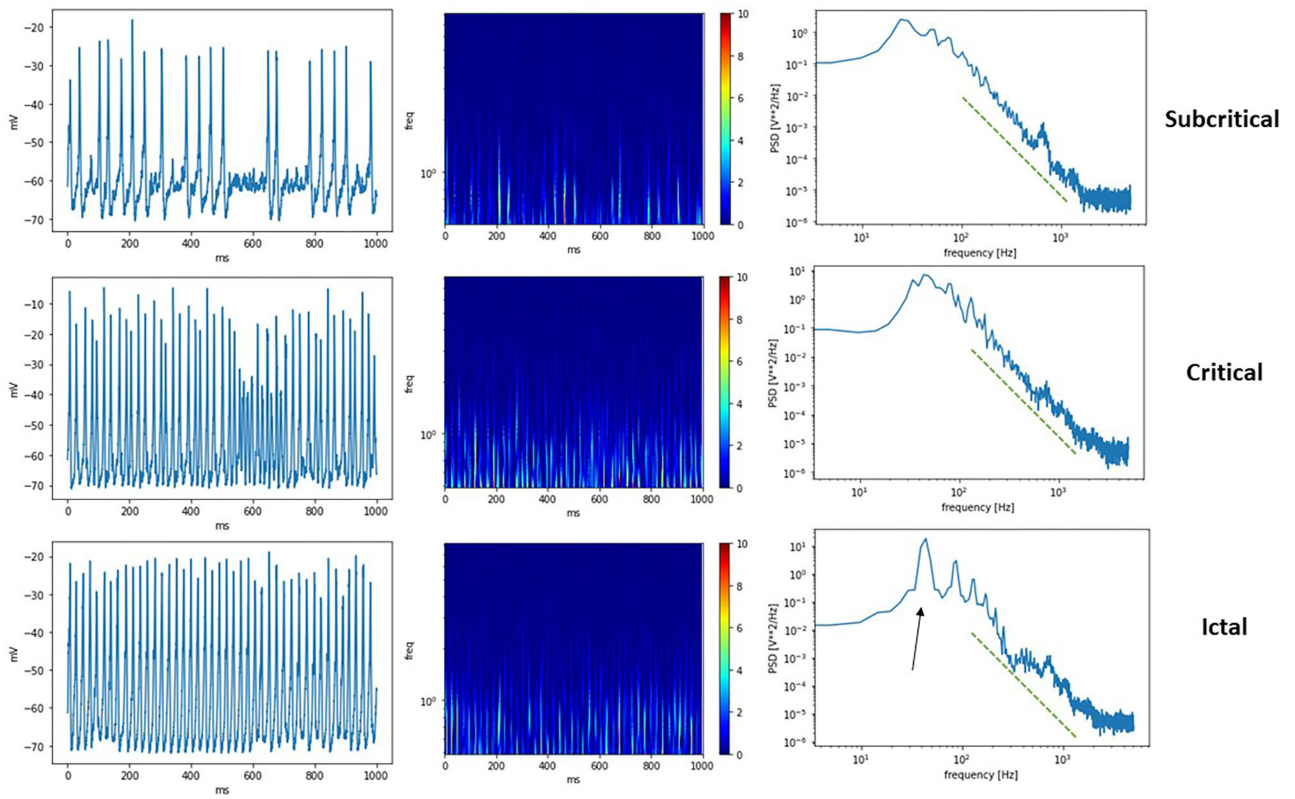

FIGURE 5 | (Left) Synthetic EEG from a subcritical, critical, a supercritical ictal state. (Middle) Corresponding wavelet transform scalograms. (Right) Calculated power spectral density plots. The green dashed line shows the same power law as a guide for the eye. The critical state follows the line most closely, while the ictal state has peaks representing synchronization at that frequency (black arow).

synchronization. In the model, we see an example where a single parameter can determine whether the system is in a normal critical state or transitions into a pathologic ictal regime. Future work will involve increasing the number of neurons, extending the simulation time, and looking at the effect of the other gating parameters on neuronal state.

One of the traditional criticisms of ML, including deep learning neural networks (DLNNs), has been that the trained systems appear to be "black boxes" for which it is not easy to understand how they are making decisions. The recent development of topological data analysis (TDA) (Carlsson, 2009) has led to attempts to treat the parameters of a trained DLNN system as just another data set. Then TDA methods can be applied to the data set of system weights to gain insights into how the DLNN system is making decisions. Such insights may ultimately lead to neurophysiologic insights into ictogenesis. ECoG data from epilepsy patients can serve as the foundation for refining and validating the THINKER models described in this article. This model-generated data can then be used as input to ML routines to attempt to identify biomarkers of impending seizures, and also to attempt to gain insight into the underlying neurobiology of ictogenesis. For ethical reasons, human data acquisition must be driven by clinical parameters and hence is available in extremely limited amounts. Once the simulated ECoG generated by the THINKER models has been validated against actual patient data, it can then serve as input to ML techniques to attempt further characterize electrophysiologic biomarkers of impending seizures. Although there are numerous ML methods, we anticipate using DLNNs in our initial work. DLNN has been found to generate increasingly abstract concepts as it progresses deeper into the hierarchy of hidden layers (Schmidhuber, 2015). This should increase the likelihood that the DLNN model can be useful for developing a system for predicting the real-time risk of impending seizures.

CONCLUSION

The present work utilized full Hodgkin-Huxley simulations with artificial neurons connected in a small-world topology to demonstrate that alteration in a single neuronal parameter can catalyze the transition to the pathological synchronization characteristic of epilepsy, a change from non-seizure (interictal or subcritical) to seizure (ictal) states. These simulations help provide a better understanding of the ways network topology, synaptic strength, and neuron excitability influence both healthy brain function and pathological states such as epilepsy. A primary biomarker of the seizure state is a spike in either the

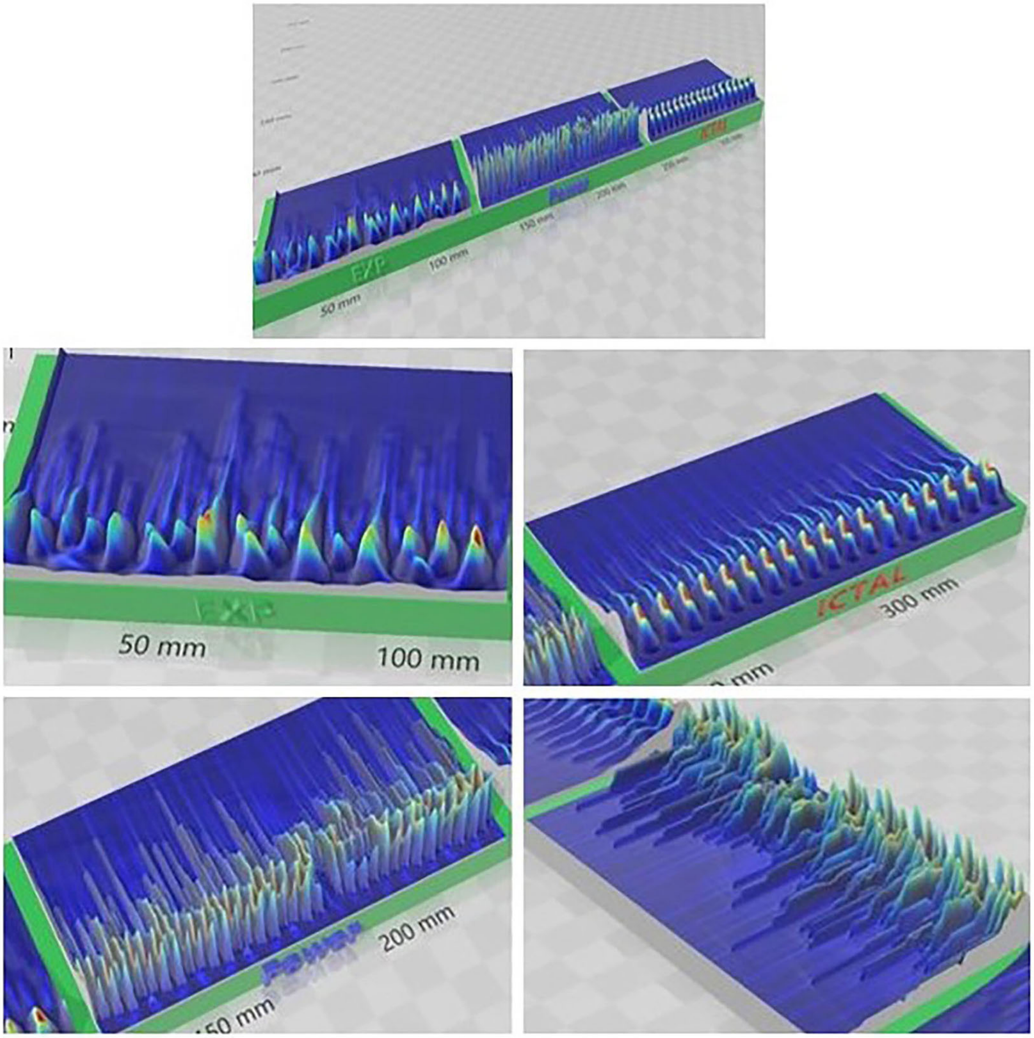

FIGURE 6 | 3D-Printed representation of subcritical (EXP) critical (Power) and supercritical (Ictal) phases. Complex activity is evident at higher frequencies most in the critical phase. Compare this with the repetitive nature of the ictal phase and the reduced activity in the subcritical phase.

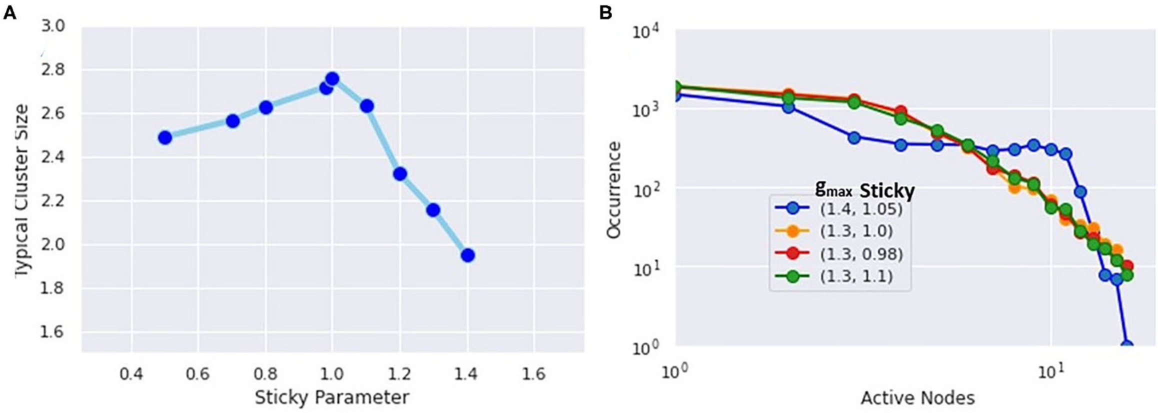

FIGURE 7 | (A) Extracted mean cluster size using the formula from percolation theory. The largest value occurs near the physiological value of the Na$^+$ inactivation parameter called "sticky." **(B)** Histogram of simultaneously firing neurons. The critical state most closely follows a power-law distribution, while the subcritical states show exponential decay.

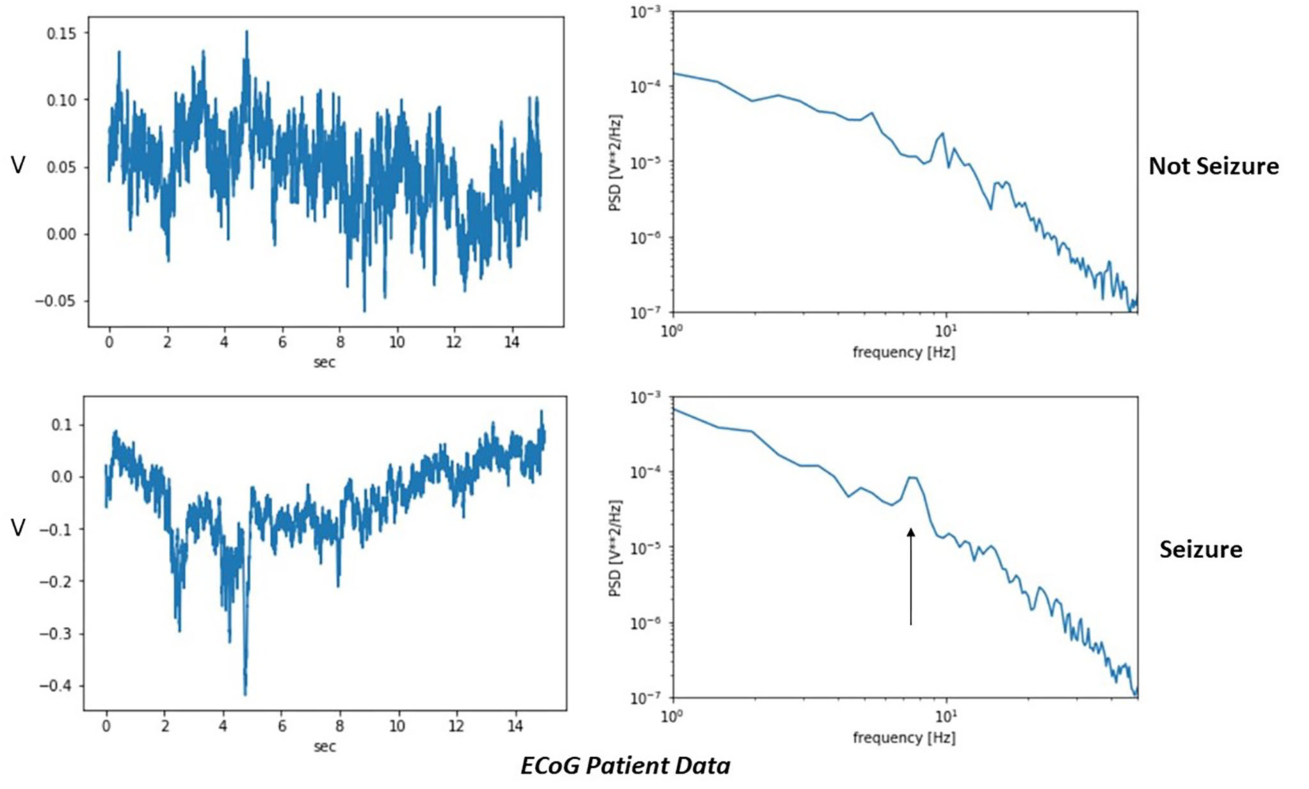

FIGURE 8 | Example high-resolution patient data using electrocorticography (ECoG) intracranial electrodes. A time interval marked by the clinician as a seizure has a peak in the power spectral density (black arrow) that is absent during the non-seizure sample.

power spectral density of an averaged signal like patient ECoG measurements, or the histogram of the number of simultaneously firing neurons in the highly spatially resolved simulation data. In addition, during the healthy intervals marked by activity at or near the critical point, the simulations reproduce power-law behavior seen in patient data. This project also makes more visible the distinct brain phases observed in actual patients. The next steps are obtaining ECoG recordings from patients with temporal lobe epilepsy who underwent presurgical evaluation for temporal lobectomy. This will allow us to refine and validate the insights from the models, especially important given the scarcity of actual patient ECoG data. The synthetic data generated by the model can also be used to train future machine-learning algorithms to assess the real-time risk of seizure onset.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

LN created and performed the simulations and wrote portions of the manuscript. GC conceived the project and wrote portions of the manuscript. RW worked on the neuroscience aspects and wrote portions of the manuscript. FM and VC worked on the mathematical theory. AP worked on the physiological aspects and edited the manuscript. JL cowrote the introduction and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Nova Southeastern University President's Faculty Research and Development Grant #335472. The publication fee was funded by the College of Medicine and Health Sciences, Khalifa University, Abu Dhabi, UAE.

ACKNOWLEDGMENTS

The authors thank Travis Craddock for helpful discussions and Beth Gilbert for her invaluable assistance in editing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncom.2020.583350/full#supplementary-material

REFERENCES

Akansu, A. N., Serdijn, W. A., and Selesnick, I. W. (2010). Emerging applications of wavelets: a review. *Phys. Commun.* 3, 1–18. doi: 10.1016/j.phycom.2009.07.001

Aznan, N. K. N., Atapour-Abarghouei, A., Bonner, S., Connolly, J. D., Al Moubayed, N., and Breckon, T. P. (2019). “Simulating brain signals: creating synthetic eeg data via neural-based generative models for improved ssvep classification,” in *2019 International Joint Conference on Neural Networks (IJCNN)*. (Budapest: IEEE). doi: 10.1109/IJCNN.2019.8852227

Bassett, D. S., and Bullmore, E. (2006). Small-world brain networks. *Neuroscientist* 12, 512–523. doi: 10.1177/1073858406293182

Beggs, J. M. (2018). “Session C42: emergent dynamics in neural systems,” in *APS March Meeting 2018*. Los Angeles, CA: Bulletin of the American Physical Society. Available online at: http://meetings.aps.org/link/BAPS.2018.MAR.\hboxC42.4

Beggs, J. M. (2019). The critically tuned cortex. *Neuron* 104, 623–624. doi: 10.1016/j.neuron.2019.10.039

Beggs, J. M., and Timme, N. (2012). Being critical of criticality in the brain. *Front. Physiol.* 3:163. doi: 10.3389/fphys.2012.00163

Breskin, I., Soriano, J., Moses, E., and Tlusty, T. (2006). Percolation in living neural networks. *Phys. Rev. Lett.* 97:188102. doi: 10.1103/PhysRevLett.97.188102

Carlsson, G. (2009). Topology and data. *AMS Bull.* 46, 255–308. doi: 10.1090/S0273-0979-09-01249-X

Chialvo, D. R. (2010). *Emergent Complex Neural Dynamics: the Brain at the Edge*. Available online at: https://arxiv.org/pdf/1010.2530.pdf (accessed February 20, 2020).

Chua, L. (2013). Memristor, Hodgkin-huxley and edge of chaos. *Nanotechnology* 24:383001. doi: 10.1088/0957-4484/24/38/383001

Dickman, R., Muñoz, M. A., Vespignani, A., and Zapperi, S. (2000). Paths to self-organized criticalit. *Braz. J. Phys.* 30, 27–41. doi: 10.1590/S0103-97332000000100004

Dickman, R., Vespignani, A., and Zapperi, S. (1998). Self-organized criticality as an absorbing-state phase transition. *Phys. Rev. E* 57, 5095–5105. doi: 10.1103/PhysRevE.57.5095

Ermentrout, G. B., and Terman, D. H. (2010). “The hodgkin–huxley equations,” in *Mathematical Foundations of Neuroscience. Interdisciplinary Applied Mathematics, (vol. 35).*, eds S. S. Antman, J. E. Marsden, L. Sirovich, and S. Wiggins (New York, NY: Springer). doi: 10.1007/978-0-387-87708-2_1

Friedman, N., Ito, S., Brinkman, B. A., Shimono, M., deVille, R. E., Dahmen, K. A., et al. (2012). Universal critical dynamics in high resolution neuronal avalanche data. *Phys. Rev. Lett.* 108:208102. doi: 10.1103/PhysRevLett.108.208102

Gal, A., and Marom, S. (2013). Self-organized criticality in single-neuron excitability. *Phys. Rev. E* 88:062717. doi: 10.1103/PhysRevE.88.062717

Hodgkin, A. L., and Andrew, F. H. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544. doi: 10.1113/jphysiol.1952.sp004764

Hong, H., Choi, M. Y., and Kim, B. J. (2002).Synchronization on small-world networks. *Phys. Rev. E* 65:026139. doi: 10.1103/PhysRevE.65.026139

Humphries, M. D., and Gurney, K. (2008). Network “small-world-ness:” a quantitative method for determining canonical network equivalence. *PLoS ONE* 3:e0002051. doi: 10.1371/journal.pone.0002051

Larremore, D. B., Shew, W. L., and Restrepo, J. G. (2011). Predicting criticality and dynamic range in complex networks: effects of topology. *Phys. Rev. Lett.* 106:58101. doi: 10.1103/PhysRevLett.106.058101

Ma, Z., Turrigiano, G. G., Wessel, R., and Hengen, K. B. (2019). Cortical circuit dynamics are homeostatically tuned to criticality *in vivo. Neuron* 104, 655–664 doi: 10.1016/j.neuron.2019.08.031

Meisel, C., Storch, A., Hallmeyer-Elgner, S., Bullmore, E., and Gross, T. (2012). Failure of adaptive self-organized criticality during epileptic seizure attacks. *PLoS Comput. Biol.* 8:e1002312. doi: 10.1371/journal.pcbi.1002312

Miller, K. J., Sorensen, L. B., Ojemann, J. G., and Den Nijs, M. (2009).Power-law scaling in the brain surface electric potential. *PLoS Comput. Biol.* 5:e1000609. doi: 10.1371/journal.pcbi.1000609

Mormann, F., Andrzejak, R. G., Elger, C. E., and Lehnertz, K. (2007). Seizure prediction: the long and winding road. *Brain* 130, 314–333. doi: 10.1093/brain/awl241

Netoff, T. I., Clewley, R., Arno, S., Keck, T., and White, J. A. (2004). Epilepsy in small-world networks. *J. Neurosci.* 24, 8075–8083. doi: 10.1523/JNEUROSCI.1509-04.2004

Ori, H., Marder, E., and Marom, S. (2018). Cellular function given parametric variation in the Hodgkin and Huxley model of excitability. *Proc. Natl. Acad. Sci. U. S. A.* 115, E8211–E8218. doi: 10.1073/pnas.1808552115

Rubinov, M., Sporns, O., Thivierge, J. P., and Breakspear, M. (2011). Neurobiologically realistic determinants of self-organized criticality in networks of spiking neurons. *PLoS Comput. Biol.* 6:e1002038. doi: 10.1371/journal.pcbi.1002038

Schmidhuber, J. (2015). Deep learning in neural networks:an overview. *Neural Netw.* 61, 85–117. doi: 10.1016/j.neunet.2014.09.003

Shin, C. W., and Kim, S. (2006). Self-organized criticality and scale-free properties in emergent functional neural networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 74:045101. doi: 10.1103/PhysRevE.74.045101

Turing, A. M. (1950). Computing machinery and intelligence. *Mind* 59, 433–460. doi: 10.1093/mind/LIX.236.433

Waldrop, M. M. (1992). *Complexity: The Emerging Science at the Edge of Order and Chaos*. New York City, NY: Simon & Schuster.

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of ‘small world, networks. *Nature* 393, 440–442. doi: 10.1038/30918

Zack, M. M., and Kobau, R. (2017). National and state estimates of the numbers of adults and children with active epilepsy—United States, 2015. *MMWR Morb. Mortal Wkly Rep*. 66, 821–825. doi: 10.15585/mmwr.mm6631a1

Zhou, D. W., Mowrey, D. D., Tang, P., and Xu, Y. (2015). Percolation model of sensory transmission and loss of consciousness under general anesthesia. *Phys. Rev. Lett.* 115:108103. doi: 10.1103/PhysRevLett.115.108103

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.

frontiers
in Computational Neuroscience

ORIGINAL RESEARCH
published: 12 February 2021
doi: 10.3389/fncom.2021.590093

The Connectivity Fingerprints of Highly-Skilled and Disordered Reading Persist Across Cognitive Domains

*Chris McNorgan**

Department of Psychology, University at Buffalo, Buffalo, NY, United States

The capacity to produce and understand written language is a uniquely human skill that exists on a continuum, and foundational to other facets of human cognition. Multivariate classifiers based on support vector machines (SVM) have provided much insight into the networks underlying reading skill beyond what traditional univariate methods can tell us. Shallow models like SVM require large amounts of data, and this problem is compounded when functional connections, which increase exponentially with network size, are predictors of interest. Data reduction using independent component analyses (ICA) mitigates this problem, but conventionally assumes linear relationships. Multilayer feedforward networks, in contrast, readily find optimal low-dimensional encodings of complex patterns that include complex nonlinear or conditional relationships. Samples of poor and highly-skilled young readers were selected from two open access data sets using rhyming and mental multiplication tasks, respectively. Functional connectivity was computed for the rhyming task within a functionally-defined reading network and used to train multilayer feedforward classifier models to simultaneously associate functional connectivity patterns with lexicality (word vs. pseudoword) and reading skill (poor vs. highly-skilled). Classifiers identified validation set lexicality with significantly better than chance accuracy, and reading skill with near-ceiling accuracy. Critically, a series of replications used pre-trained rhyming-task models to classify reading skill from mental multiplication task participants' connectivity with near-ceiling accuracy. The novel deep learning approach presented here provides the clearest demonstration to date that reading-skill dependent functional connectivity within the reading network influences brain processing dynamics across cognitive domains.

Keywords: functional connectivity, dyslexia, cognition, machine learning, math cognition, learning disabilities, fMRI

OPEN ACCESS

Edited by:
Saman Sargolzaei,
University of Tennessee at Martin,
United States

Reviewed by:
Daniela Traficante,
Catholic University of the Sacred Heart, Italy
Tuo Zhang,
Northwestern Polytechnical University, China
Fangfei Ge,
University of Georgia, United States

****Correspondence:***
Chris McNorgan
cpmcnorg@buffalo.edu

Received: *31 July 2020*
Accepted: *21 January 2021*
Published: *12 February 2021*

Citation:
McNorgan C (2021) The Connectivity Fingerprints of Highly-Skilled and Disordered Reading Persist Across Cognitive Domains.
Front. Comput. Neurosci. 15:590093.
doi: 10.3389/fncom.2021.590093

THE CONNECTIVITY FINGERPRINTS OF HIGHLY-SKILLED AND DISORDERED READING PERSIST ACROSS COGNITIVE DOMAINS

Proficient reading is a prerequisite for formal instruction and independently navigating the world, but is a skill that exists on a continuum. Developmental dyslexia is a learning disorder characterized by a difficulty in reading in the absence of any other pronounced cognitive difficulty, and is the most commonly diagnosed learning disorder (Shaywitz and Shaywitz, 2005). By definition, children with specific reading disability possess normal-range intelligence, though reading difficulty is often comorbid with dyscalculia. The factors underlying learning difficulty in both domains is not well understood, but it has been proposed that it may be attributable to a shared reliance on core cognitive processes (Archibald et al., 2013). At the other end of the continuum of reading skill, precocious reading is associated with a modest advantage in other language abilities in later childhood (Mills and Jackson, 1990), but it has not been shown to confer a general cognitive advantage in other domains.

Reading maps visual to phonological representations, and is thus fundamentally an audiovisual process. An extensive literature has explored neural processing in dyslexics and typically developing readers, and points to a model in which reading difficulty is attributable to disordered audiovisual integration of orthographic and phonological representations (Richlan, 2019). Relative to controls, dyslexics under-activate the left temporo-parietal cortex (Temple et al., 2001), and show delayed specialization in the ventral visual object processing stream for visual word processing (van der Mark et al., 2009). Under-activation in this network of regions has been argued to reflect failure of audiovisual integration processing (Blau et al., 2010; Randazzo et al., 2019) and failure to engage lexical-semantic representations (Richlan et al., 2009). Relative to controls, dyslexics over-activate the right occipitotemporal cortex and anterior inferior temporal gyrus, which has been suggested to reflect compensatory activation (Shaywitz and Shaywitz, 2005). This set of anatomically distributed brain regions supporting orthographic, phonological and semantic processing of written language is referred to as the reading network (Perfetti et al., 2007; Perfetti and Tan, 2013; Morken et al., 2017). All contemporary brain-based models of fluent and disordered reading assume that reading entails interactions within this network of more-or-less functionally-specialized brain regions (e.g., Dehaene et al., 2010; Price and Devlin, 2011).

If patterns of regional activation within this network are the dynamic product of connectivity among these regions, connectivity differences between skilled and poor readers must underlie the group differences described above. Dyslexic readers show weaker reading task-based functional connectivity between the visual word form area and other regions within the left hemisphere reading network, but greater connectivity between the visual word form area and left middle temporal and middle occipital gyrus (van der Mark et al., 2011). Left angular gyrus has also been implicated as a critical hub, with reduced task-based functional connectivity with other reading network nodes for dyslexics, but increased connectivity to posterior right hemisphere, possibly attributable to compensatory recruitment during phonological tasks (Pugh et al., 2000).

Multivariate Studies of Normal and Disordered Reading

All brain-based reading models agree that fluent reading entails cooperation among regions within the reading network that may be only conditionally involved (e.g., when the task involves phonological assembly, as in Pugh et al., 2000). Nonetheless, models are largely informed by a literature that relies on univariate general linear model analyses (GLMA), which are limited in two important respects: First, they assume linear relationships between observed and modeled values, requiring multiple independent hypothesis tests to identify conditionally-involved regions or connections. Second, because univariate analyses examine only local relationships, they cannot incorporate informative contextual information from other sources. To address the second concern, multivariate analyses have been increasingly important for informing the literature.

Multivariate pattern analyses (MVPA) commonly use machine learning classifiers to decode patterns of activity among voxel populations, revealing regional categorical sensitivity that may not manifest as category-dependent mean differences in response amplitude in a conventional GLMA. Models of normal and disordered reading have been refined by MVPA in studies, showing, for example, that impaired access to phonological representations, rather than distorted representations, may underlie reading difficulty (Boets et al., 2013; Norton et al., 2015). MVPA have also shown that regional gray matter volume patterns in posterior occipito-temporal and temporal-parietal cortices differ between dyslexics, typical readers and those with specific reading comprehension deficit (Bailey et al., 2016). The enhanced sensitivity of MVPA was leveraged in a whole-brain fMRI analysis of longitudinal data using recursive feature elimination to find that dyslexics who made gains in reading skill over a 2.5 year period could be discriminated from those who did not on the basis of activity among voxels in right IFG, left prefrontal cortex and left temporoparietal cortex (Hoeft et al., 2011).

Multivariate approaches have also been applied at the network-level. Wolf et al. (2010) used a multivariate independent components analysis (ICA) to examine network-level functional connectivity differences in older adolescents using a working memory task, arguing ICA-based methods are better-able to detect distributed network components (Esposito et al., 2006). Their analysis found that the working memory delay period was associated with increased connectivity for dyslexics within a left-lateralized frontoparietal network, and mixed differences in a second bilateral frontoparietal network, both including regions implicated in phonological processing, which they argue may reflect increased reliance on working memory in dyslexics during phonological processing. A later ICA network analysis found that improvements in word reading and comprehension following reading remediation were correlated with connectivity changes in functional networks supporting visual attention and executive control (Horowitz-Kraus et al., 2015). Multivariate studies have thus refined a model of skilled and disordered reading as dependent on interactions among multiple cognitive systems, and not reliant on a single system.

The univariate and multivariate approaches described above rely on the general linear model to test relationships between observed and predicted values for one or more variables. However, linear models may be good approximations only within certain ranges, as all biological systems exhibit nonlinearity (e.g., saturation) along some range of inputs (Korenberg and Hunter, 1990). Conditional relationships are an important class of nonlinear relationships, and though they can often be linearly modeled—for example by partitioning a continuum into groups and demonstrating a nonremovable interaction—doing so requires an experiment designed around testing the interaction.

Multilayer Feedforward Classifiers

Support vector machines are the most commonly applied machine learning application in MVPA (Mokhtari and Hossein-Zadeh, 2013). These kernel-based approaches can use nonlinear kernels to form classification boundaries, but require advanced knowledge of a suitable nonlinear kernel function. The multilayer feedforward classifier is an alternative machine learning approach that uses the backpropagation of error algorithm (Rumelhart et al., 1985) to discover an optimal nonlinear classification function. Though these models are arguably more difficult to interpret than are their simpler counterparts (Norman et al., 2006), they have several advantages: First, they are extremely powerful, and with the development of training algorithms and architectures that mitigate concerns associated with high-dimensional data (Poggio et al., 2017), variations of these networks have been foundational to the recent *Deep Learning* renaissance in machine learning. Second, they are arbitrarily flexible, with the capacity to accommodate multiple outputs. This feature allows these networks to find a solution space that best fits the training data with respect to multiple problem domains. McNorgan et al. (2020b) applied such a network to the classification of imagined objects from fMRI data, simultaneously learning to distinguish among object categories and identifying the functional networks supporting category processing. Because the model use shared parameters to solve each problem, the solution in one problem domain (e.g., object categorization) is explicitly tied to the other problem domain (e.g., functional connectivity), making alternative explanations more unlikely.

The Present Study

Because there are n^2 connections among n regions, connectivity studies of large networks face practical computational, interpretation and statistical challenges, and seed-based approaches are thus often used to restrict analyses to the n-1 connections to a seed region. Larger networks are not well suited for exploration using conventional parametric methods because univariate methods must avoid inflating the Type I error rate, and even multivariate methods like the MANOVA mathematically require a sample size that exceeds the number of variables. The present study uses a multilayer feedforward classifier network to explore large-network connectivity, and as a nonparametric multivariate analysis, suffers neither of these drawbacks. The analysis takes advantage of the extensibility of feedforward neural networks to simultaneously identify task-related functional connections that distinguish between poor- and highly-skilled readers and between word and pseudoword processing.

MATERIALS AND METHODS

Archival Data

Neuroimaging and phenotypic data were obtained from two open access longitudinal datasets hosted on the OpenNeuro.org data repository, and described in detail in Lytle et al. (2019) and Suárez-Pellicioni et al. (2019). The first ("Reading Set") comes from Lytle et al. (2019), and was collected from 188 children between the ages of 8 and 13 years spanning a range of reading ability. The Reading Set data were collected while participants engaged in rhyming judgments of pairs of sequentially-presented lexical items (monosyllabic words or pseudowords). Within each run of the task were 6 trial types: Four types of lexical items (rhyming vs. non-rhyming items that had either similar or dissimilar spelling), a fixation cross response baseline, and a nonlinguistic symbol matching judgment. This experiment was blocked by run, using multiple presentation modality (auditory, visual, audiovisual) and lexicality (words, pseudowords) conditions, and the present study analyzed only data from the unimodal visual condition.

The second ("Math Set") comes from Suárez-Pellicioni et al. (2019), and was collected from 132 children between the ages of 8 and 14 years spanning a range of math and reading ability. These data were selected for two reasons: First, the Math Set study drew from the same population as the Reading Set study, and was carried out concurrently with the Reading Set study by the same research staff, using the same equipment and the same standardized testing procedures. Second, because mental arithmetic arguably bears little similarity to rhyming judgments of written words, validation of the classifier models against these data provides an extremely challenging test of generalizability. The Math Set experiment was blocked by run, during which participants made relative numerosity judgments, performed mental subtraction or multiplication. The present study analyzed only data from the single-digit mental multiplication runs, which was selected because single-digit multiplication is commonly taught by rote memorization and was assumed to be the mental arithmetic task most likely to involve lexical processing. Aside from domain-specific difficulties in reading or math fluency, participants in both studies were cognitively normal.

The Reading Set and Math Set studies were carried out in accordance with recommendations of the Institutional Review Board at Northwestern University and the anonymized data were released for reanalysis. The protocols were approved by the Institutional Review Board of Northwestern University. Parents of all participants gave written informed consent in accordance with the Declaration of Helsinki.

Participant Selection

Participants were selected from among those who had completed both longitudinal time points in either dataset. This initial selection criterion was motivated by two important considerations: First, participants with poor-quality data resulting from, e.g., excessive movement, failure to satisfactorily perform the task etc., were not invited to return for the second longitudinal time point, and were thus heuristically excluded from the study. Second, the longitudinal dataset permits analytic flexibility to accommodate follow-up investigations of developmental changes. From among those participants, subsets of participants were selected from each of the datasets on the basis of MRI data quality and standardized test scores. Fourteen poor readers with low scores ($<$ 85 scaled score) across multiple standardized measures of reading skill at the first longitudinal time point were selected from among the Reading Set participants. These participants were matched against 14 highly-skilled readers with high scores across the same standardized measures of reading skill. Aside from the poor scores in measures specific to reading ability, the poor readers had otherwise normal language ability, as indicated by their spoken word recognition and picture vocabulary scores, which were within the normal range, as were their WASI subtest and full-scale IQ scores. This selection method followed the approach applied to this dataset in McNorgan et al. (2013) which identified matched samples of 13 typically-developing readers and 13 children reaching a clinical criterion for reading-specific impairment using similar selection criteria. Standardized testing scores for the Reading Set participants are presented in **Table 1**. Five Math Set participants with the lowest Woodcock-Johnson III reading subtest scores ($\leq$85) at the first longitudinal time point were matched against five participants with high scores at the first longitudinal time point on the same tests. Standardized testing scores for these participants are presented in **Table 2**.

Neuroimaging Data Processing

MRI acquisition details can be found in the dataset descriptors provided in Lytle et al. (2019) and Suárez-Pellicioni et al. (2019). FreeSurfer/FS-FAST (version 6.0, http://surfer.nmr.mgh.harvard.edu) was used for all fMRI data preprocessing and GLM analyses. Reading Set and Math Set data were collected from the same MRI scanner with the same acquisition parameters, and identical processing pipelines were applied to both data sets to obtain functional connectivity estimates and generate machine learning classifier patterns.

Anatomical Data Processing

After segmenting into white and gray matter volumes, the mean of the timepoint 1 and timepoint 2 T1-weighted images were rendered in 3D surface space and normalized to the FreeSurfer standard template space. Cortical surface space was labeled using an automated atlas-based parcellation of the gyri and sulci (Destrieux et al., 2010).

Functional Data Processing

We applied here the data processing pipeline used in a recent application of a multilayer machine learning classifier to functional connectivity and coarse-scale cortical pattern analysis (McNorgan et al., 2020a). Functional images for both timepoint 1 and timepoint 2 were co-registered with the 3D anatomical surface generated from the mean anatomical value for each subject by FreeSurfer (Version 6.0) for each participant and mapped onto a common structural template for group analysis using isomorphic 2 mm voxels. Functional data were preprocessed using the standard FS-FAST BOLD processing pipeline interoperating with FSL (Version 5.0) to apply motion-correction, slice-time correction and spatial smoothing using a 2 mm Gaussian kernel. For the functional connectivity estimation, an additional voxel-wise detrending step removed linear trends and regressed out white matter and CSF signal from the data.

GLM Analysis and Region of Interest Construction

Subject-level GLM analyses were performed for each participant in the Reading Set, combining the functional data from both longitudinal timepoints. The event-related GLM analyses used the SPM canonical hemodynamic response function to model blood oxygen dependent (BOLD) responses for the lexical, fixation cross baseline, and symbol matching trials. Subject-level contrasts between lexical trials and fixation cross baseline were carried out separately for word and pseudoword runs. Group-level random effects analyses concatenated the first level analyses for poor- and highly-skilled readers separately. These single-sample group-level contrast maps allowed unbiased selection of regions involved in either word or pseudoword processing for either poor or highly-skilled readers.

Large cortical patches are unlikely to be homogenously organized, and so custom Python scripts, written by the author, computed the union of the group-level cluster map cortical surface annotation files that was then intersected with the FreeSurfer surface annotation of the Destrieux et al. (2010) atlas. This procedure subdivided functional clusters along anatomical boundaries into multiple regions of interest (ROI). Large ROIs that remained after this subdivision were manually subdivided into smaller regions of visually similar size to other ROIs. The union of clusters from the four significant contrast maps were thus subdivided into 115 ROIs of comparable size ($M = 145$ mm^2) to the Lausanne parcellation ROIs used in previous studies of functional connectivity in surface space (Hagmann et al., 2008, 2010; Honey et al., 2009; McNorgan and Joanisse, 2014; McNorgan et al., 2020a).

Classifier Training

Pattern Generation

Classifier input patterns were generated from task-related functional connectivity among ROIs within residualized BOLD time series data for each functional run (4 pseudoword, 4 word). An initial regression removed linear trends and white matter and CSF signal, and the mean BOLD time series was computed across the voxels in each of the ROIs. The weights in a machine learning classifier are free parameters that are iteratively adjusted to fit the training data. An overabundance of free parameters can allow the classifiers to *overfit* the training data, "memorizing" the

TABLE 1 | Mean (SD) standardized measure scores for Reading Set participants.

Measure	Poor	Skilled	*t*(26)	
WJ-III WordID	89 (6)	119 (8)	11.04	
WJ-III Word Attack	91 (11)	117 (8)	9.84	
WJ-III Passage Comprehension	90 (11)	110 (12)	5.01	
WJ-III Oral Comprehension	106 (12)	112 (11)	1.29	*n.s.*
WJ-III Picture Vocabulary	112 (12)	110 (13)	−0.3	*n.s.*
TOWRE SW	92 (13)	115 (10)	5.02	
TOWRE PDE	90 (17)	118 (12)	4.76	
WASI Verbal IQ	110 (16)	123 (15)	1.77	
WASI Performance IQ	107 (12)	120 (11)	2.64	
WASI FSIQ	110 (14)	124 (11)	2.59	

WJ-III, Woodcock-Johnson III; TOWRE, Test of Word Reading Efficiency; WASI, Wechsler Abbreviated Scale of Intelligence. All t statistics are significant at p < 0.05 unless otherwise noted (n.s.).

TABLE 2 | Mean (SD) standardized measure scores for Math Set participants.

Measure	Poor	Skilled	*t*(8)
WJ-III WordID	92 (8)	125 (5)	4.53
WJ-III Word Attack	96 (8)	114 (11)	2.07
WJ-III Passage Comprehension	91 (6)	116 (4)	4.61
WJ-III Passage Comprehension	N/A	N/A	
WJ-III Oral Comprehension	N/A	N/A	
TOWRE SW	90 (6)	124 (7)	7.58
TOWRE PDE	86 (8)	113 (18)	2.65
WASI Verbal IQ	83 (1)	143 (3)	37.4
WASI Performance IQ	89 (5)	119 (14)	4.05
WASI FSIQ	84 (3)	135 (6)	14.39

WJ-III, Woodcock-Johnson III; TOWRE, Test of Word Reading Efficiency; WASI, Wechsler Abbreviated Scale of Intelligence. All t statistics are significant at p < 0.05.

patterns rather than learning rules that generalize to novel cases. Overfitting is measured by the degree of discrepancy between training set fit and validation set fit: A model has overfit the data if it demonstrates high training set accuracy but poor validation set accuracy. Machine learning approaches commonly mitigate overfitting by increasing the number of *distinct* patterns in the training set through data augmentation (Lemley et al., 2017). This increases the ratio of unique patterns to the dimensionality of the feature encodings, introducing additional noise in the process, but providing additional contexts in which to identify reliably predictive features (see Koistinen and Holmström, 1991, for a discussion of the utility of noise in classifier training).

Data were augmented by splitting the 6-min time series in half, and computing weighted connectivity between time series vectors separately for the first and second half of each run using two methods: Pearson correlation, which measures a linear dependency between time series vectors, and cross-mutual information (XMI), which is sensitive to the general dependency between two variables, which may or may not be linear (Li, 1990), may be more sensitive to synchronization in noisy systems (Paluš, 1997), and has been shown to be a useful connectivity-based predictor in classifier-based studies of learning difficulty (McNorgan et al., 2020b). The main diagonal of the 115 × 115 symmetric matrix was eliminated, as was the redundant lower triangle of the symmetric matrix. The remaining 6,555 values were normalized, and rescaled to fall between 0 and 1. It must be noted that the correlation-based and XMI-based connectivity patterns were only moderately correlated with each other ($r = 0.42$). This is important because it ensures that an accurate classification of a validation-set pattern, e.g., computed using XMI, is not attributable to training exposure to a highly-similar pattern computed over the same time series using the Pearson correlation. Rather, the classifiers were forced to identify both linear and nonlinear coactivation dynamics that are predictive of reading skill and lexicality.

The distributions of the positively-skewed functional connectivity values were made normal by application of a square-root transformation. Each connectivity pattern vector was tagged with a value indicating lexicality (0 = pseudoword; 1 = word), and a value indicating group (0 = poor reader; 1 = highly-skilled reader). The source and destination nodes of each functional connection was also noted, and the indices of the 253 functional connections between subregions of a single cluster within the functional mask were recorded for subsequent filtering, as these connections might reflect trivial correlations.

Classifier Model Architecture

Classifier models were implemented in *TensorFlow* (Version 2.2, https://www.tensorflow.org), using hyperparameters and a model architecture informed by a previous study performing orthogonal ADHD diagnosis and behavioral profile classification in an unrelated dataset (McNorgan et al., 2020a). Input values fed forward through a Gaussian noise ($SD = 0.05$) and a dropout layer (rate = 0.2), which further distorted the training patterns as a standard data augmentation method (Srivastava et al., 2014). The standard deviation of the Gaussian noise was selected to roughly match the standard deviation of input values ($SD = 0.07$). The perturbed input patterns fed forward through three densely connected rectified linear unit hidden layers, with the size of each layer determined by a logarithmic function (base 2) of the number of input features: The size of the first hidden layer was determined using formula (1), where i is the number of input features:

$$\mathit{max}(16, 2 \times \mathit{ceil}(\log_2 i)) \tag{1}$$

The second and third hidden layers were always half the size of the first hidden layer. Batch normalization was applied at each hidden layer (Ioffe and Szegedy, 2015). The first hidden layer additionally used an $\ell 1$ (LASSO) regularizer of 0.0005 to promote sparsity among the weights and act as a pruning mechanism (Allen, 2013) by eliminating redundant or non-predictive (i.e., noisy) features. These activations fed forward to two single-unit logistic classifier layers using the sigmoid activation function to simultaneously classify patterns with respect to condition lexicality and group. Models of this architecture are also known as multilayer perceptrons (MLP), and compute a Bayesian posterior probability of category membership, given a set of input features (Ruck et al., 1990). An important feature of multilayer networks is that the imposition of small hidden layers between the input and classifier layers requires the network to compute a low-dimensional transformation of the high-dimensional input patterns. This denoises and computes a nonlinear stochastic independent components analysis (ICA) on the input patterns (Hyvärinen and Bingham, 2003).

In the context of this network, classification decisions were thus made from configurations of functional connections encoded in the final hidden layer, rather than on individual functional connections encoded in the input layer as would be the case in a standard SVM classifier. This is important to bear in mind because the ratio of training examples to the dimensionality of the hidden layer ICA transformation was over 30:1. A comparison between performance of SVM and MLP single-class classification of high-dimensional connectivity vectors in McNorgan et al. (2020a) found that the SVM classifier had a propensity to overfit the training patterns, demonstrating poor validation set accuracy (58% for group classification). In contrast, the MLP classifier showed high validation set accuracy (91%) for group classification of the same data, demonstrating the utility of hidden layer ICA dimensionality reduction afforded by MLP models. The training set was balanced with respect to both classifications and the categories were orthogonal. The classifier model architecture is illustrated in **Figure 1**.

Training Parameters and Procedure

Classifier models were trained over 216 epochs using the standard gradient descent optimizer with a learning rate of $\varepsilon = 0.01$, decay $= 0.05$ and momentum $= 0.9$, and a batch size of 16. Early-stopping was used, monitoring lexical classifier error (Caruana et al., 2001). Stratified K-folds cross-validation (Diamantidis et al., 2000) pseudo-randomly partitioned the patterns into $K = 5$ balanced sets of training and withheld validation set data, such that each pattern appeared among the validation set data exactly once. One model was trained for each of the folds, and performance was evaluated on the withheld validation set, allowing performance metrics to reflect the model's ability to generalize to novel data. Note that the machine learning literature may distinguish between validation and test sets, with validation sets partitioned from the training data for initial model hyperparameter tuning, and test sets removed from the training data for measuring the model's ability to generalize to new data (Larsen et al., 1996), however this study refers to withheld cross-validation set accuracy as validation set performance, as hyperparameter tuning was not performed.

Iterative Feature Reduction

Classifier training often applies feature set reduction to eliminate uninformative features and improve accuracy (Chu et al., 2012). In addition to the nonlinear ICA imposed by the model's hidden layer architecture, feature reduction at the input layer was used as an analog to backwards stepwise linear regression in this study as a means of identifying the small proportion of functional connections with the greatest influence on classification decisions.

An iterative feature selection algorithm was applied over a series of model generations. Beginning with the full set of functional connectivity patterns, the 253 functional connections between nodes within clusters from the unified GLM contrast maps were eliminated from the input patterns prior to the first model generation. After training, cross-validation set performance was assessed for each fold, and the summed path weights from each remaining input feature unit to each of the classifier output units was computed. Each input feature unit contributes toward each classification decision by driving the classifier unit toward either 0.0 (if the summed path weights are negative) or 1.0 (if the summed path weights are positive). After all training folds had completed, the input feature set was reduced through *decimation*, by which the summed absolute value of input feature weights to both classifier units were ranked-ordered, and features that were in the bottom tenth percentile for *either* classification were eliminated for the subsequent generation. Training proceeded in this way until the full feature set had been reduced to 0.05 of its original size (i.e., finding the 95th percentile of functional connections predictive of *both* lexicality and group) while still demonstrating above-chance classification accuracy, requiring either 15 or 16 generations. This was repeated twenty times to produce a sample of $n = 20$ model families. Note that, because the hidden layer size was a logarithmic function of the number of input features, hidden layers were also reduced across generations. The final generation of models contained 0.01 of the number of trainable weights as the first generation models, despite containing 0.05 the number of input features. The disproportionately large reduction in trainable weights across model generations greatly reduced the representational capacity of later models.

The feature selection procedure leaks information about the most informative features between generations of a single family of models, however this is not problematic for several reasons: First, the relative diagnostic utility of an input feature is not fixed as the models become more constrained with the elimination of hidden units across generations. Second, feature reduction was intended to facilitate interpretation, rather than improve accuracy (it will be shown that removing 0.95 of the input data and 0.99 of the trainable model weights decreased classification accuracy); the survival and subsequent inclusion of predictor x_i in the $n{+}1^{\text{th}}$ model generation is analogous to the survival of predictor x_i into the $n{+}1^{\text{th}}$ step in a backwards stepwise multiple regression. Finally, each of the 20 model families are independent, precluding information leakage *between* model families. The analyses that follow aggregate results across all model families, permitting measures of predictive reliability for each functional connection, and more importantly, the evaluation of a composite model comprising the most informative features independently identified by each of the model families.

Model Evaluation

Each generation of models was evaluated with respect to validation set classification accuracy for lexical condition and group, and d′ was computed, collapsing across all cross-validation folds. The frequency with which each functional

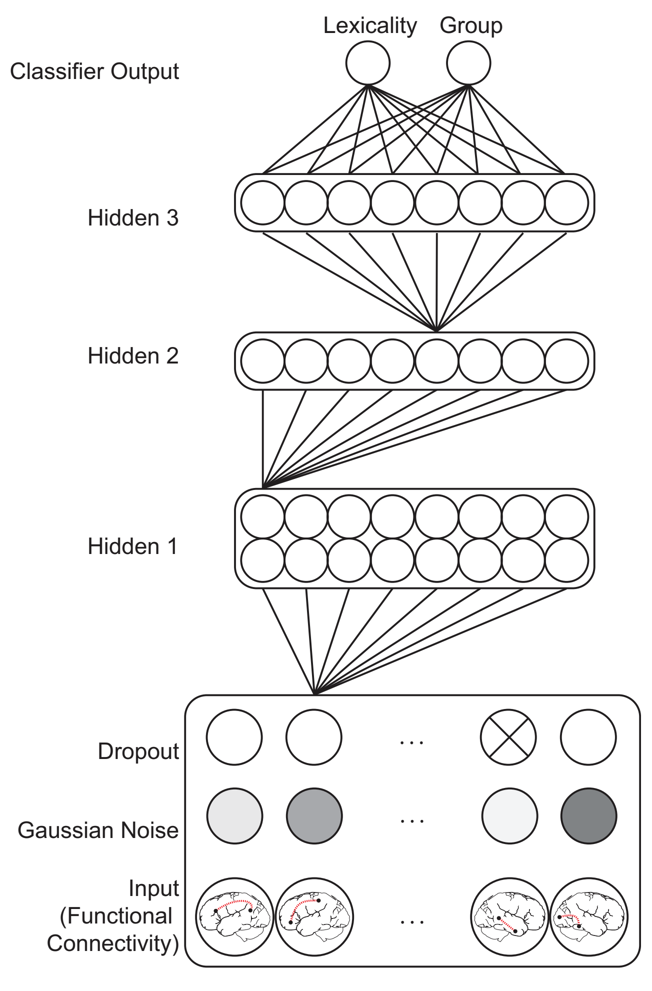

FIGURE 1 | Feedforward classifier model architecture. Gaussian noise and probabilistic dropout perturbs functional connectivity input patterns, which feed forward through a series of densely connected hidden layers before reaching two classifier output units. Most model weights are omitted for clarity. The influence of a particular functional connection on a classification decision is indexed by summing the weights of all paths between the corresponding input unit and a classifier unit. Multiplication of the series of the matrices encoding the weights between each layer computes these values in parallel for all input features.

connection was retained in the final generation of models was summed across the twenty model families, and used to inform the construction of a composite model comprising the most informative features across all model families. Input features appeared with decreasing frequency among the final generation of models, with more than a third appearing in none of the models, and most of the remaining appearing in at most one model. The most informative features were defined as those 327 functional connections that appeared in at least 4 of 20 final-generation models, because they represented 0.049 of the full set of functional connections, and thus corresponded to the 95th percentile of most predictive features. Validation set lexicality and group classification performance was assessed for a composite model trained using this set of functional connections using 10-fold cross-validation to permit relative comparisons among the most reliably predictive features. This composite model served as a direct evaluation of these putatively predictive functional connections in isolation.

Math Set Evaluation

Five sets of first-generation classifier models were trained using all Reading Set patterns as training data. Classification accuracy and d′ scores were computed for classification decisions on functional connectivity patterns from the Math Set mental multiplication task as a replication to bolster claims of classifier generalizability and test whether lexically-related functional connectivity within the reading network distinguishes poor from highly-skilled readers, even when engaged in a non-reading task. This evaluation followed the procedure described above for evaluating classifier models performance at the first generation, with two differences: First, only 3 cross-validation folds were used in each set—each network was trained using a random .66 partitioning of the Reading Data. The second critical difference was that, in place of the withheld Reading Data, classification performance was evaluated using patterns from the Math Set.

RESULTS

GLM Lexicality Contrasts

GLM contrast maps were thresholded using a voxel-wise significance level of $p = 0.001$ and a cluster-size threshold of $P = 0.05$ based on the FreeSurfer random permutation cluster simulation. **Figure 2** and **Tables 3**, **4**, **5** shows retained clusters showing contrasts for highly-skilled readers between words and fixation (A) and pseudowords and fixation (B), and for poor readers between words and fixation (C). Poor readers had no above-threshold regions showing significant differences between pseudowords and fixation. The union of the GLM contrast maps depicted in **Figure 2** served as a functionally-defined mask that include regions with high signal-to-noise and associated with preferential responses to either words or pseudowords for either highly-skilled readers or poor readers, and functional connectivity among these regions was computed for the analyses that follow. Contrasts between lexical conditions or groups would select regions with *a priori* bias toward one condition or group and thus not performed. Similar analyses have been reported elsewhere on supersets of these data (McNorgan et al., 2014; McNorgan and Booth, 2015; Edwards et al., 2018; Smith et al., 2018), and so the GLM results are not discussed further.

Classifier Performance

Reading Set validation set classification accuracy and d′ was computed separately for lexicality and group classifications among the twenty families of models. These values were computed separately for each lexicality and group, however because the 95% confidence intervals overlapped between words and pseudowords and between poor and very skilled readers in each analysis, both lexicality conditions and both group conditions were collapsed in the presentation of results that follow.

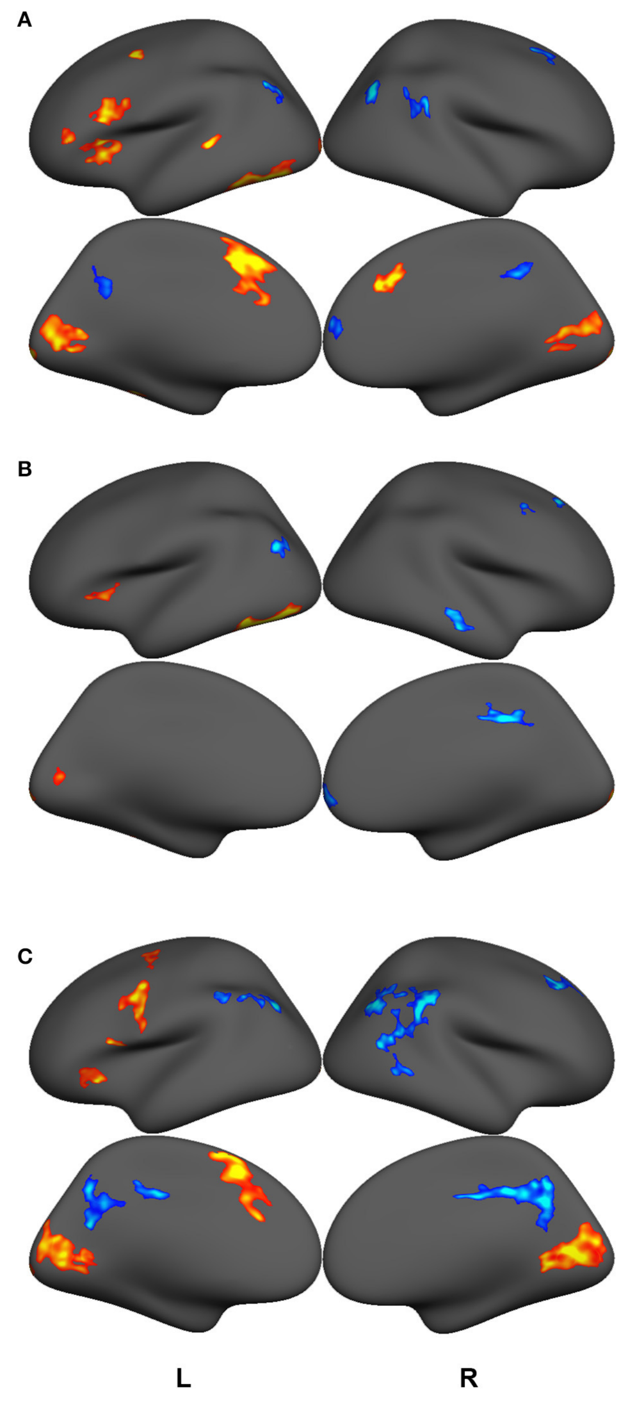

FIGURE 2 | GLM contrast maps showing regions with significant activation (hot) or deactivation (cold) for lexical trials vs. fixation cross baseline for good readers making rhyming judgments on words **(A)** and pseudowords **(B)** and for poor readers making rhyming judgments on words **(C)**. All maps thresholded with a voxel-wise $p = 0.001$ and a cluster-size corrected threshold of $P = 0.05$. There were no regions showing above-threshold differences between pseudoword and fixation cross baseline among poor readers. The union of these clusters was used as a functional mask to define the Reading Network. Functional connectivity was computed over anatomically-delineated subdivisions of the Reading Network.

TABLE 3 | Coordinates of peak surface activity within clusters reaching $P < 0.05$ (cluster-level corrected) significance in the words vs. fixation contrast for highly-skilled readers.

Contrast	Region	*t*	Size	X	Y	Z	*p*
W > Fix	r. superior frontal	6.46	315	12	28	33	0.0001
	l. superior frontal	6.34	888	−11	12	43	0.0001
	l. superior temporal sulcus	5.99	155	−58	−35	3	0.0116
	l. fusiform	5.89	1383	−41	−64	−14	0.0001
	l. lateral occipital	5.83	443	−17	−99	−11	0.0001
	l. caudal middle frontal	5.68	172	−36	1	48	0.0068
	l. pars opercularis	5.61	458	−36	24	10	0.0001
	r. lateral occipital	5.50	295	21	−92	−8	0.0001
	l. pars opercularis	5.49	556	−51	17	22	0.0001
	l. pericalcarine	5.25	1118	−13	−83	4	0.0001
	r. pericalcarine	4.74	959	7	−78	12	0.0001
	l. pars triangularis	4.63	220	−46	34	4	0.0018
W < Fix	r. inferior parietal	−5.55	333	40	−71	39	0.0001
	l. inferior parietal	−5.07	274	−41	−67	39	0.0005
	r. supramarginal	−4.82	353	60	−41	22	0.0001
	r. superior frontal	−4.73	179	22	12	47	0.0049
	r. precuneus	−4.63	188	9	−43	39	0.0030
	r. superior frontal	−4.27	182	9	54	15	0.0043
	l. precuneus	−4.02	198	−13	−55	32	0.0034

L, left; R, right; t, peak t-statistic; Size, cluster extent in mm^2. Coordinates reflect standard MNI space. Region labels derived from the FreeSurfer Desikan-Killiany atlas surface.

TABLE 4 | Coordinates of peak surface activity within clusters reaching $P < 0.05$ (cluster-level corrected) significance in the pseudowords vs. fixation contrast for highly-skilled readers.

Contrast	Region	*t*	Size	X	Y	Z	*p*
PW > Fix	l. fusiform	8.081	936	−40	−71	−12	0.0001
	l. lateral occipital	6.268	339	−17	100	−7	0.0001
	l. lateral occipital	5.937	171	−17	−99	−8	0.0033
	l. insula	4.255	154	−28	23	5	0.0063
	l. pericalcarine	4.152	137	−12	−79	12	0.0113
PW < Fix	r. precuneus	−7.245	304	6	−37	42	0.0001
	l. inferior parietal	−5.427	272	−45	−69	25	0.0002
	r. superior frontal	−5.355	112	21	26	46	0.0138
	r. middle temporal	−5.314	287	65	−19	−15	0.0001
	r. medial orbitofrontal	−4.256	301	9	55	−4	0.0001
	r. caudal middle frontal	−4.243	113	40	10	51	0.0138

L, left; R, right; t, peak t-statistic; Size, cluster extent in mm^2. Coordinates reflect standard MNI space. Region labels derived from the FreeSurfer Desikan-Killiany atlas surface.

First Generation

Among all first-generation models, which used all but the 253 short-distance within-cluster functional connections as predictors, lexical classification accuracy ($M = 0.64$, $SD =$

TABLE 5 | Coordinates of peak surface activity within clusters reaching $P < 0.05$ (cluster-level corrected) significance in the words vs. fixation contrast for poor readers.

Contrast	Region	*t*	Size	X	Y	Z	*p*
W > Fix	l. superior frontal	7.08	798	−10	13	52	0.0001
	l. lateral occipital	6.591	229	−17	−101	−6	0.0007
	r. pericalcarine	6.326	1863	18	−72	11	0.0001
	l. precentral	5.69	600	−52	−2	45	0.0001
	l. pericalcarine	5.312	1623	−9	−81	3	0.0001
	l. lateral orbitofrontal	5.28	308	−27	28	−3	0.0001
	l. pars opercularis	5.148	144	−53	14	14	0.0093
	l. precentral	3.867	206	−25	−14	56	0.0010
W< Fix	r. precuneus	−6.447	1064	6	−38	42	0.0001
	r. supramarginal	−6.141	1197	53	−45	28	0.0001
	l. inferior parietal	−5.713	266	−37	−74	38	0.0002
	r. rostral middle frontal	−5.323	316	27	25	38	0.0001
	l. precuneus	−5.29	616	−14	−62	23	0.0001
	r. inferior parietal	−5.289	715	41	−69	39	0.0001
	l. posterior cingulate	−5.007	176	−6	−27	39	0.0026
	l. inferior parietal	−4.665	195	−49	−55	38	0.0015
	r. middle temporal	−4.35	188	53	−55	−2	0.0013
	l. supramarginal	−4.3	180	−56	−45	40	0.0023

L, left; R, right; t, peak t-statistic; Size, cluster extent in mm^2. Coordinates reflect standard MNI space. Region labels derived from the FreeSurfer Desikan-Killiany atlas.

0.04,0.95 CI [0.63, 0.66]) was well above chance, with all chi-squared tests of the ratios of classification decisions to expected chance performance significant at $p < 10^{-21}$, and mean d′ of 0.74, 0.95 CI [0.65, 0.83]. The mean phi coefficients of the contingency table for each model family was 0.29, $SD = 0.09$,0.95 CI [0.25, 0.32], indicating that middle- to long-distance task-dependent functional connections within the reading network bear a weak-to moderate explanatory relationship with lexical task.

Group classification accuracy ($M = 0.94$, $SD = 0.06.95$ CI [0.92,0.96]) was high for all first-generation model families, with all chi-squared tests of the ratios of classification decisions to expected chance performance significant at $p < 10^{-256}$, and mean d′ of 3.58, 0.95 CI [3.28, 3.88]. The mean phi coefficient of the contingency table for each model family was 0.88, $SD = 0.11$, 0.95 CI = [0.84, 0.93], indicating that middle- to long-distance task-dependent functional connections within the reading network bear a very strong explanatory relationship with reading skill.

Final Generation

Among all final-generation models, lexical classification accuracy ($M = 0.53$, $SD = 0.01.95$ CI [0.52, 0.53]) was above chance for most models, with 18 of 20 chi-squared tests of classification accuracy significant at $p < 0.05$, and a mean d′ of 0.13,0.95 CI [0.11, 0.15]. The mean phi coefficients of the contingency table for each model family was 0.05, $SD = 0.02$,0.95 CI = [0.04, 0.06], indicating that the top 5% most predictive middle- to long-distance task-dependent functional connections within the reading network bear a negligible explanatory relationship with lexical task.

Group classification accuracy ($M = 0.71$, $SD = 0.03.95$ CI [0.70, 0.72]) remained well above chance for all final-generation model families, with all chi-squared tests of the ratios of classification decisions to expected chance performance significant at $p < 10^{-110}$, and mean d′ of 1.22,0.95 CI [1.15, 1.29]. The mean phi coefficient of the contingency table for each model family was 0.42, SD = 0.06,0.95 CI = [0.40, 0.44], indicating that the 0.05 most informative middle- to long-distance task-dependent functional connections among the reading network bear a strong explanatory relationship with reading skill.

Figure 3 plots the d′ scores for lexical and group classification performance for all models as a function of the proportion of the total number of functional connections used as input features. Trendlines were fit with a third-order polynomial. The performance trajectory shows a nearly linear relationship between lexicality classification performance and the number of functional connections used in the models, suggesting that most functional connections within the task-defined network are roughly equally predictive of lexicality condition, and thus that classification accuracy is proportional to the number of connections in the model. The performance trajectory for group classification shows a nearly logarithmic relationship, with a large increase in classification performance between models using 0.05 and 0.15 of all functional connections, and declining gains in accuracy as models become larger. This indicates that a small number of functional connections within the task-defined network are disproportionately predictive of reading skill. Finally, d′ measures of lexicality and group classification performance were positively correlated across all models, $r_{(315)} = 0.71$, $p < 10^{-48}$. As would be expected by the equal weighting given to the error signal for the two classifier units, this indicates that optimizing accuracy in one classification decision was not at the expense of the other.

Evaluation of Predictive Functional Connections Using a Composite Model

Of the 6,555 functional connections, 2,770 never appeared in a final model, with most of the remaining features appearing in at most one of the 20 final-generation models. Validation set lexicality and group classification performance was assessed for a composite model trained using the set of 327 functional connections appearing in at least 4 final generation models using 10-fold cross-validation to permit relative comparisons among the most reliably predictive features, which were otherwise found in independent model families. Collapsed across all 10 model folds, the d' measure was 0.03 for lexical classification accuracy, but 1.02 for group classification accuracy, which was significantly better than chance, $X^2_{(1)} = 58.58$, $p < 10^{-13}$, and indicates that the 95th percentile of most predictive functional connections within the task-defined network are significant predictors of reading skill.

The weight structures of the composite models generated by each fold were used to construct functional networks for further analysis and visualization: The magnitude of a

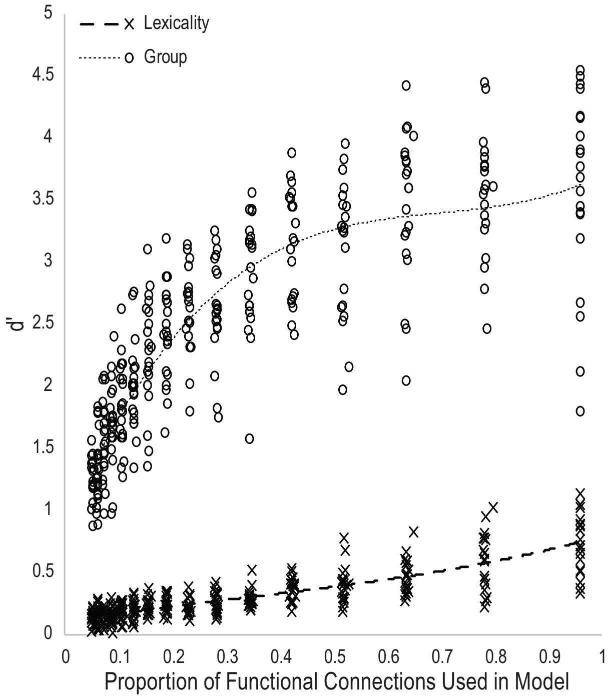

FIGURE 3 | Scatterplot of d-prime measures of classifier performance for each generation of twenty model families as a function of the proportion of the total number of functional connections used to predict reading skill group or lexicality.

path weight indexes the relevance of that feature on the classification decision, and the valence of the weight indicates the classification associated with high functional connectivity (negative: pseudoword, poor reader; positive: word, highly-skilled reader). To limit the scope of analysis, the weight magnitudes were normalized, and those functional connections with $|Z| > 1.0$ were selected as *highly-relevant*. To enable visualization of networks comprising only highly relevant connections, adjacency matrices were constructed for functional networks predicting group and lexicality classification by adding to an empty adjacency matrix the corresponding mean functional connectivity score for all highly-relevant functional connections identified in the previous step. For example, high connectivity between left inferior frontal sulcus and right cingulate ROIs was a highly-relevant predictor of the highly-skilled reading group. The mean connectivity between these regions for all run splits for all highly-skilled readers was thus added to the Group adjacency matrix. These adjacency matrices were used to generate **Figures 4**, **5**.

Group Classification

Functional connectivity networks including only those functional connections with strong classification weights for reading skill are rendered in **Figure 4** and **Table 6**. From this figure are highlighted regions that participate in multiple predictive functional connections, and therefore may be important network hubs. Three adjacent nodes belonging to a functional cluster spanning the right occipital pole and right calcarine sulcus were terminal points of multiple functional connections within the right hemisphere that are predictive of poor reading skill. Similarly, two adjacent nodes belonging to a functional cluster within the left occipitotemporal cortex were terminal points of multiple functional connections both within and between hemispheres that are predictive of high reading skill. Finally, two regions of interest, one within the left precuneus and the other within the anterior intermediate parietal sulcus of Jensen, were terminal points of the functional connections that most reliably predict *both* poor and highly-skilled reading, suggesting that these regions may be critical hubs supporting reading.

Lexicality Classification

The functional connectivity network including only those functional connections with strong classification weights for lexical condition are rendered in **Figure 5** and **Table 7**. It is notable that that the set of highly-relevant functional connections was dominated by multiply-connected hubs within the left hemisphere, many along the left occipitotemporal sulcus, consistent with the privileged role of this region in orthographic processing. However, because lexicality classification among composite models was at chance accuracy, this network should be interpreted with caution: Though the identified connections are the *most* individually predictive of lexicality, they are only a subset of the functional connections that accurately predict lexicality. This follows from the observation that iterative pruning produced a linear decrease in classification accuracy. Thus, the core functional networks supporting word and pseudoword processing appear to be very similar, and become reliably differentiable only when examining connectivity among the reading network as a whole.

Math Dataset

Across all five sets of models, classification accuracy was comparable to classification of the Reading Dataset ($M = 0.96$, $SD = 0.04$, 0.95 CI = [0.92, 0.98]), with all chi-squared tests of the ratios of classification decisions to expected chance performance significant at $p < 10^{-18}$, and mean d′ score of 5.60 ($SD = 1.42$;0.95 CI = [4.25, 6.95]). The mean phi coefficient for the 5 model families was .91 ($SD = 0.07$;0.95 CI = [0.84, 0.98]), indicating that functional connectivity during mental multiplication within the reading network has a very strong explanatory relationship with reading skill.

There was no clear correct lexicality classification for mental multiplication connectivity patterns, and so it was expected that the lexicality classifications would be random and equiprobable. It was thus noteworthy that all models classified 100% of mental multiplication patterns as consistent with pseudoword processing.

DISCUSSION

This study used a classifier model in a novel analytic approach that identified functional connectivity patterns that simultaneously distinguished word from pseudoword processing

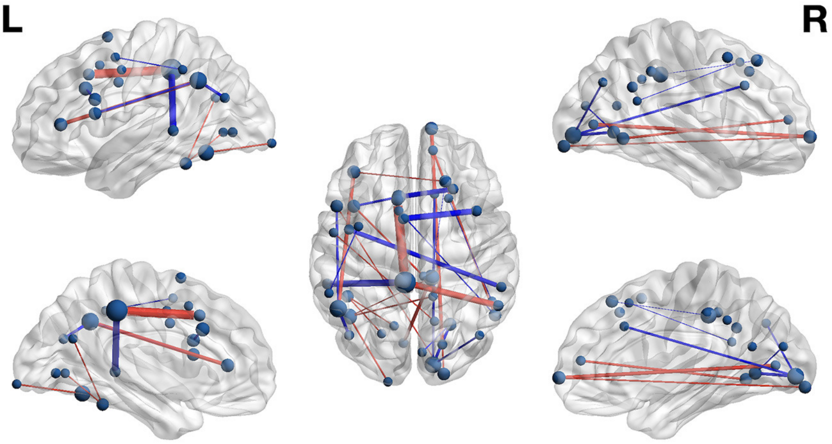

FIGURE 4 | Functional networks including only those highly-relevant functional connections where high connectivity was predictive of poor (blue) or highly-skilled (red) reading in the set of composite models. Edge thickness indexes connectivity strength, and ROI node diameter indexes summed connectivity with other regions.

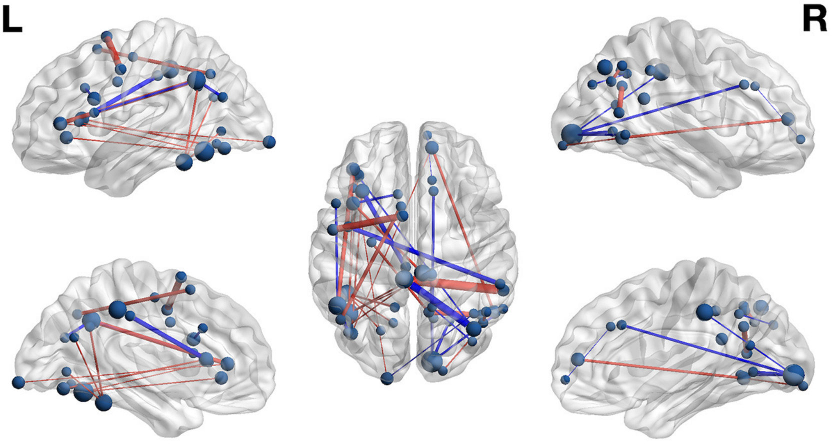

FIGURE 5 | Functional networks including only those highly-relevant functional connections where high connectivity was predictive of pseudoword (blue) and word (red) lexicality condition in the set of composite models. Edge thickness indexes connectivity strength, and ROI node diameter indexes summed connectivity with other regions.

with above chance accuracy and distinguished between patterns from poor and highly-skilled readers with near-ceiling accuracy, sensitivity, and specificity. Indeed, because each participant contributed multiple patterns, if the modal classification decision was used for each participant, the probability of misclassifying a participant on 5 of 8 task runs is $<10^{-5}$. The phi coefficient is a non-parametric relative of the Pearson correlation, measuring the strength of association between two variables. To provide context for the phi coefficients computed from the classification of reading skill in both the Reading and Math sets, they are well above the correlations reported between reading skill and predictors such as working memory ($r = 0.29$; Peng et al., 2018), or phonological processing, which has been argued to play a causal role in reading acquisition ($r = 0.09$ to 0.73 depending on measure; Wagner, 1988). The results indicate that a functional connectivity fingerprint of reading skill can be found within the network supporting word and pseudoword reading, and produced two unexpected findings not discoverable by traditional univariate parametric approaches.

Reading Skill Shapes Functional Organization in Other Cognitive Domains

First, task-based functional connectivity fingerprints of poor- and highly-skilled reading are present in task-based functional connectivity in other cognitive domains. Functional networks are dynamic, reflecting the demands of the cognitive task (Gonzalez-Castillo and Bandettini, 2018), and so it is remarkable that the manner in which the brain organizes during mental multiplication reflects an individual's reading skill. As noted earlier, reading difficulty is often comorbid with math difficulty, and the co-occurrence of math difficulty with reading difficulty is attributed to reliance on shared components (Fletcher, 2005). The mental multiplication task presented problems symbolically, rather than as word problems. However, the classifiers were

TABLE 6 | Highly relevant functional connections predictive of reading skill in composite model.

Group	Source region	X	Y	Z	Destination region	X	Y	Z	r	n
Poor Skill	l Cingulate Gyrus	−10	20	30	l Inferior Frontal Sulcus	−39	14	23	0.52	9
	l Cingulate Gyrus	−10	20	30	r Superior Frontal Sulcus	25	25	41	0.64	8
	l Jensen Sulcus	−49	−52	35	l Pars Opercularis	−50	14	14	0.52	10
	l Jensen Sulcus	−49	−52	35	l Superior Temporal Sulcus	−42	−69	24	0.55	4
	l Precentral Sulcus	−42	−1	34	r Supramarginal Gyrus	56	−38	41	0.62	9
	l Precuneus	−6	−35	42	l Superior Temporal Sulcus	−56	−36	3	0.72	8
	l Superior Frontal Gyrus	−10	11	45	r Inferior Temporal Sulcus	52	−56	−2	0.47	10
	l Supplementary Motor Area	−7	6	63	r Middle Frontal Gyrus	40	11	50	0.69	5
	l Supramarginal Gyrus	−55	−42	42	l Middle Frontal Gyrus	−36	1	50	0.47	7
	r Cuneus	5	−80	19	r Calcarine Sulcus	23	−62	2	0.48	5
	r Occipital Pole	12	−88	0	r Angular Gyrus	42	−69	34	0.50	7
	r Occipital Pole	12	−88	0	r Anterior Cingulate Gyrus	13	22	32	0.52	11
	r Occipital Pole	12	−88	0	r Calcarine Sulcus	23	−62	2	0.54	4
	r Superior Frontal Gyrus	20	30	48	r Marginal Sulcus	11	−32	39	0.43	6
	r Superior Frontal Gyrus	22	19	47	r Superior Temporal Sulcus	50	−47	22	0.44	4
High Skill	l Inferior Frontal Sulcus	−39	14	23	r Cingulate Gyrus	6	−31	40	0.50	6
	l Inferior Frontal Sulcus	−39	36	7	r Superior Frontal Gyrus	20	30	48	0.47	5
	l Jensen Sulcus	−49	−52	35	l Inferior Frontal Sulcus	−39	36	7	0.59	7
	l Lateral Occipitotemporal Gyrus	−42	−57	−11	l Occipital Pole	−17	−99	−5	0.49	6
	l Lateral Occipitotemporal Gyrus	−42	−57	−11	r Subparietal Sulcus	12	−44	36	0.48	6
	l Lateral Occipitotemporal Gyrus	−43	−44	−18	l Calcarine Sulcus	−21	−68	2	0.47	4
	l Lateral Occipitotemporal Gyrus	−43	−44	−18	l Parieto–occipital Sulcus	−13	−63	24	0.47	6
	l Lateral Occipitotemporal Gyrus	−42	−57	−11	r Marginal Sulcus	11	−32	39	0.46	5
	l Medial Lingual Gyrus	−7	−74	2	r Marginal Sulcus	11	−32	39	0.57	9
	l Pars Opercularis	−50	14	14	r Occipital Pole	19	−94	−7	0.47	7
	l Precentral Gyrus	−52	−3	42	r Parieto–occipital Sulcus	13	−60	19	0.47	7
	l Precuneus	−6	−35	42	l Cingulate Gyrus	−11	18	39	0.84	6
	l Precuneus	−6	−35	42	r Angular Gyrus	54	−50	31	0.71	8
	r Calcarine Sulcus	18	−75	7	r Transverse Frontopolar Gyrus	11	64	−1	0.59	8
	r Inferior Temporal Sulcus	52	−56	−2	r Anterior Cingulate Gyrus	12	50	10	0.52	10
	r Occipital Pole	19	−94	−7	r Transverse Frontopolar Gyrus	11	64	−1	0.49	7

L, left; R, right; n, number of models (/20) containing connection; Coordinates reflect standard MNI space. Region labels derived from the FreeSurfer, Destrieux et al. (2010) atlas.

required to simultaneously classify reading skill and lexicality, and the shared weight structure for these classifications ensured that the functional connections that discriminate reading skill are prima facie relevant to lexical processing (and vice versa). This makes it difficult to explain this finding through appeals to non-lexical processes. The single-digit multiplication task was selected because it is often taught by rote memorization, so that these problems are solved by retrieving verbalized facts, rather than algorithmically. This result may thus reflect reading-skill dependent differences in the networks that are engaged whenever accessing lexicalized knowledge.

Lexicality Sensitivity Differentiates Poor- and Highly-Skilled Readers

Second, iterative pruning on the basis of diagnosticity for both classification decisions provided novel insight into lexical processing in poor and skilled readers. There were equal numbers of poor and highly-skilled readers, and all participants performed both word and pseudoword reading, and it was thus expected that characteristic functional connections associated with both word and pseudoword reading would be found for both groups. Instead, among the most categorically-diagnostic functional connections in the composite model, there were no cases where strong functional connectivity was strongly predictive of both poor reading skill and of word reading. Conversely, there were no cases where strong functional connectivity was strongly predictive of both high-reading skill and of pseudoword reading. This suggests that poor- and highly-skilled readers might be best differentiated by how they process pseudowords and familiar words, respectively. The most transparent explanation is that there are functional networks that are used almost exclusively by children with reading difficulty when encountering unfamiliar letter strings, and other functional networks used almost exclusively by highly-skilled readers for decoding known

TABLE 7 | Highly relevant functional connections predictive of lexicality in composite model.

Lexicality	Source region	X	Y	Z	Destination region	X	Y	Z	*r*	*n*
Pseudoword	l Anterior Lateral Fissure	−36	32	−2	r Angular Gyrus	39	−67	43	0.46	7
	l Cingulate Gyrus	−10	20	30	l Inferior Frontal Sulcus	−39	14	23	0.49	9
	l Jensen Sulcus	−49	−52	35	l Superior Temporal Sulcus	−42	−69	24	0.52	4
	l Occipital Pole	−17	−99	−5	r Intraparietal Sulcus	39	−57	41	0.44	6
	l Pars Opercularis	−50	14	14	l Jensen Sulcus	−49	−52	35	0.49	10
	l Posterior Cingulate Gyrus	−5	−27	38	l Insula	−33	22	10	0.59	7
	l Precentral Sulcus	−42	−1	34	r Supramarginal Gyrus	56	−38	41	0.59	9
	l Precuneus	−6	−35	42	r Angular Gyrus	39	−67	43	0.72	7
	r Anterior Cingulate Gyrus	12	29	31	r Superior Frontal Gyrus	9	58	−3	0.41	5
	r Occipital Pole	12	−88	0	r Calcarine Sulcus	23	−62	2	0.51	4
	r Occipital Pole	12	−88	0	r Anterior Cingulate Gyrus	13	22	32	0.5	11
	r Occipital Pole	12	−88	0	r Middle Temporal Gyrus	58	−54	0	0.48	6
	r Occipital Pole	12	−88	0	r Cingulate Marginalis	11	−32	39	0.45	4
	r Subparietal Sulcus	12	−53	39	r Middle Occipital Gyrus	38	−75	32	0.45	5
Word	l Inferior Frontal Sulcus	−39	36	7	l Jensen Sulcus	−49	−52	35	0.57	7
	l Inferior Frontal Sulcus	−39	14	23	r Posterior Cingulate Gyrus	6	−31	40	0.49	6
	l Inferior Frontal Sulcus	−39	36	7	l Occipital Pole	−17	−99	−5	0.43	4
	l Inferior Temporal Sulcus	−42	−65	−6	r Posterior Cingulate Gyrus	6	−31	40	0.45	4
	l Occipital Sulcus	−40	−70	−9	l Insula	−33	22	10	0.46	5
	l Occipital Sulcus	−40	−70	−9	r Anterior Cingulate Gyrus	12	50	10	0.43	5
	l Occipitotemporal Sulcus	−43	−44	−18	l Subparietal Sulcus	−12	−53	31	0.48	4
	l Occipitotemporal Sulcus	−42	−57	−11	r Posterior Cingulate Gyrus	6	−31	40	0.46	6
	l Occipitotemporal Sulcus	−43	−44	−18	l Calcarine Sulcus	−21	−68	2	0.45	4
	l Occipitotemporal Sulcus	−43	−44	−18	l Parieto–occipito Sulcus	−13	−63	24	0.45	6
	l Occipitotemporal Sulcus	−42	−57	−11	l Lateral Fissure	−36	32	−2	0.44	5
	l Occipitotemporal Sulcus	−42	−57	−11	l Lateral Fissure	−39	24	8	0.44	5
	l Occipitotemporal Sulcus	−42	−57	−11	r Cingulate Marginalis	11	−32	39	0.44	5
	l Precentral Sulcus	−27	−11	50	r Parieto–Occipital Sulcus	13	−60	19	0.5	7
	l Precuneus	−6	−35	42	r Supramarginal Gyrus	58	−42	23	0.78	7
	l Superior Frontal Gyrus	−8	12	55	l Angular Gyrus	−47	−61	39	0.54	5
	l Supplementary Motor Area	−7	6	63	l Precentral Gyrus	−52	−3	42	0.68	8
	r Angular Gyrus	47	−56	44	r Angular Gyrus	45	−63	34	0.52	6
	r Inferior Temporal Sulcus	52	−56	−2	r Anterior Cingulate Gyrus	12	50	10	0.5	10
	r Occipital Pole	19	−94	−7	r Inferior Temporal Sulcus	52	−56	−2	0.45	6
	r Subparietal Sulcus	10	−56	31	r Superior Temporal Sulcus	44	−58	14	0.62	7

L, left; R, right; n, number of models (/20) containing connection; Coordinates reflect standard MNI space. Region labels derived from the FreeSurfer, Destrieux et al. (2010) atlas.

words. This informs the interpretation of previous network studies of normal and impaired reading (Fraga González et al., 2016; Edwards et al., 2018), suggesting that reading-skill related differences in graph-theoretic metrics of functional networks reflect different network configurations, rather than the same networks used to different extents. The absence of characteristic connectivity predicting poor reading and word processing suggests that when dyslexics encounter highly-practiced words, they use the same networks as do their non-impaired counterparts.

Poor reading skill was predicted by high connectivity between right occipital pole and right anterior cingulate and right calcarine sulcus that was also predictive of pseudoword reading. This highlights a right-lateralized network preferentially recruited by poor readers when encountering unfamiliar lexical strings, and may indicate that the posterior right hemisphere compensatory mechanism proposed by Shaywitz and Shaywitz (2005) reflects ongoing processing of lexical items after initial failed lexical retrieval. High-skilled reading was predicted by strong connectivity from two hubs in the left lateral occipitotemporal gyrus that were also predictive of word reading. This is consistent with the privileged role this region plays in orthographic processing and provides a connectivity-based explanation for the under activation of this region in dyslexics (Richlan et al., 2009). The pattern indicates that activity in this region is less coordinated with other regions in the reading network for poor readers, and also suggests precocious specialization of the region is accompanied by strong coherence with upstream left visual cortex in strong readers.

This pattern is interesting in the context of both the Price and Devlin (2011) Interactive Account and the hierarchical organization model proposed by Dehaene et al. (2005). The Interactive Account assumes that connectivity between occipitotemporal and higher-level processing regions is experience-dependent, but that connectivity with earlier visual cortex is not. Dehaene's hierarchical account argues that this region represents complex conjunctions of visual features represented in earlier visual cortex. In both of these models, strong connectivity with earlier visual cortex should support a robust representation of orthographically-relevant visual information and promote reliable orthography-phonology mapping for higher-skilled reading processing known words, consistent with the pattern-separation mechanism described by McNorgan et al. (2011).

The left Jensen sulcus, which bounds the angular and supramarginal gyri at the posterior end of the superior temporal sulcus, was the terminus of functional connections predictive of both poor reading and highly-skilled reading, and of word and pseudoword reading, depending on the connected region. Shahin et al. (2009) proposed that a circuit between left angular gyrus and superior temporal sulcus was part of a phonological repair network. Moreover, Del Tufo and Myers (2014) found that dyslexics with greater reading difficulty were more likely to engage the phonological repair processes when listening to distorted speech sounds. This connection was predictive of pseudoword reading and poor reading skill, and in this light, suggests that a dyslexic reader's difficulty mapping from orthography to phonology initiates two parallel processes: the engagement of a left hemisphere circuit to clean up a phonological representation to find a most likely match, and the engagement of a posterior right hemisphere visual circuit that facilitates processing unrecognized orthographic forms. The inferior frontal sulcus has been argued to be involved in both semantic and phonological processing (Poldrack et al., 1999; Turkeltaub et al., 2003), and high connectivity between Jensen sulcus and this posterior frontal region was predictive of highly-skilled reading of familiar words. This suggests that the rapid transmission of phonological information between left temporoparietal cortex and posterior frontal cortex for semantic and higher-order phonological analysis is a characteristic feature of highly-skilled reading.

Other Theoretical Implications

The Primacy of Left Occipitotemporal Cortex in Visual Word Identification

Though brain-based theories of reading may disagree on the mechanisms underlying its privileged role in lexical processing, it is widely accepted that the left occipitotemporal cortex plays a critical role in decoding written words. Though connectivity within this region was excluded from the classifiers, multiple functional connections for which strong connectivity was predictive of word reading terminated in the left occipitotemporal sulcus, suggesting that strong functional connectivity with this region is associated with lexical decoding of familiar words. It follows that functional connectivity to and from the left occipitotemporal sulcus must be weaker when decoding unfamiliar letter strings, perhaps as a consequence of incoherent inter-regional activation patterns when pattern matching fails. The predictive reciprocal backwards connectivity to left calcarine sulcus and forwards connectivity to the inferior frontal gyrus maps well on to the circuit proposed in the Interactive Account (Price and Devlin, 2011) and those proposed to be responsible for the specialization of the visual word form area (Bouhali et al., 2014), and suggests that lexical decoding of familiar words relies on multiple intact pathways involving this region (Richardson et al., 2011).

Reading on a Continuum

By focusing on the extremes of the reading skill continuum, this study avoided ambiguous cases falling close to the category boundary. The multilayer classifier architecture used in this study supports multiclass categorization, allowing for classification of poor, typical and highly-skilled readers. However, though the poor readers had either diagnosed reading difficulty or received reading intervention, the highly-skilled readers did not come from an identified special population, and the extent to which this group deviates from typical readers is unknown. If reading skill is on a continuum, the inclusion of typical readers, as a midpoint between the extremes, would provide additional context for interpreting predictive connections as characteristic of a particular level of reading skill, as opposed to merely absent in one group or another. However, model performance should decrease with increasing similarity between adjacent groups.

It was hypothesized that the classifier would distinguish between opposite ends of a continuum of reading skill, however it was surprising that between-group classification of reading skill was far more accurate than was within-subject classification of lexicality. Classification in these models depends on overall pattern similarity, rather than on a mean difference threshold as in conventional parametric null hypothesis testing, and accuracy decreases when there is no clearly-matching category prototype. The original studies used an event-related design, and functional connectivity was thus computed over time series including both lexical and non-lexical trials. Non-lexical trials were common to both lexicality conditions and comprised roughly half the time series. Thus, functional connectivity was likely more similar between word and pseudoword runs than would be the case if computed over homogeneous blocks of lexical trials. This similarity should contribute toward confusability between word and pseudoword patterns. That said, task overlap alone cannot account for poorer lexicality classification because poor and highly-skilled readers completed identical tasks, yet the classifiers distinguished groups with near-ceiling accuracy. This implies that the functional connectivity patterns characteristic of reading skill are detectable even when functional connectivity measures are substantially influenced by a non-lexical task, confirmed by the near-ceiling classification accuracy of the mental multiplication task.

The overall pattern of results therefore suggests that word and pseudoword reading entails very similar processes, leading to similar category prototypes; that word and pseudoword processing is highly variable, leading to multiple prototypical patterns for each category; or possibly both. However, the results also indicate that poor and highly-skilled readers engage in characteristic processing that leads to easily-identified functional connectivity patterns. Again, the capacity to make multiclass categorizations open up the possibility of further exploration including nonwords in addition to words and pseudowords to further disentangle the networks involved in lexicality processing.

As noted earlier, there was no obvious lexical category for the mental multiplication trials, which used neither words nor pseudowords. It was thus expected that random lexicality classifications of these patterns would produce approximately equal numbers of word and pseudoword categorizations. Instead, all classifiers categorized all mental multiplication patterns as pseudowords, indicating that the functional networks recruited during mental multiplication are consistent with those used during pseudoword processing. Pseudowords are lexical strings using legal arrangements of known orthographic symbols, but with no associated semantic content. Mental multiplication problems (e.g., "3 × 5") are likewise composed of legal arrangements of orthographic symbols with no semantic content. One explanation for this unexpected finding is that the lexical classifiers associated pseudowords with connectivity patterns related to low-level lexical syntax matching in the absence of top-down semantic input, though this implies different syntax-matching processes under semantic contexts. Additional study using experimental tasks that include nonwords and illegal mathematical expressions is required to explore this interesting cross-domain overlap.

LIMITATIONS AND OPEN QUESTIONS

The reported results provide distributional statistics aggregating classifier performance over hundreds of replications using a particular set of experimental hyperparameters such as learning rate, batch size, Gaussian noise distribution, dropout probability, and network size. The random nature of several of these parameters and of the training procedure itself guarantees that the network performance was not contingent on a specific set of parameters and sequence of training events, but it is tempting to speculate that a different set of parameters might produce a different outcome. This is certainly the case, though just as it is trivially easy to select an inappropriate statistical test or experimental design, it is not interesting to observe that a poorly-selected set of hyperparameters can produce a poorly-performing set of models that lead to a different conclusion. Hyperparameter optimization techniques exist (e.g., Knudde et al., 2017), and it is possible that better performance is possible. However, because the data used for hyperparameter optimization should be removed from the training set, these techniques are more data and time intensive. As described earlier, network hyperparameters were based on those from the classifier used in McNorgan et al. (2020a), and the high classification performance did not justify the additional effort. Many hyperparameter changes will have uninteresting consequences, such as changing the amount training required to reach asymptotic performance, however interesting insights might be gained from parametric manipulations to, e.g., the number of hidden units or hidden layer regularization, that impact the model's ability to encode network motifs among the hidden units.

This study used intact groups, and though the poor readers had normal-range IQ (>100) consistent with a specific reading impairment, highly-skilled readers had significantly greater scores on all IQ subtests. The set of participants within the Math Set with non-overlapping reading skill likewise differed on all IQ subtests, raising the question of whether the classifiers were categorizing on the basis of general intelligence. Though the influence of general intelligence cannot be ruled out, parallel lexicality classification using shared parameters ensured that diagnosticity was contingent on relevance to lexicality, which is confirmed by the nonrandom lexicality classification of patterns from both datasets. Intelligence is generally viewed as a complex multidimensional construct, and though poor and highly-skilled readers may differ with respect to traits like working memory that correlate with both general intelligence and reading skill, the rhyming task on which the classifiers were trained cannot be argued to be representative of general cognitive processing. It is thus probable that the models learned the connectivity fingerprints of reading skill, rather than of general intelligence level. Multiple aspects of the classifier training—from the ICA reduction of input dimensionality to the parallel within- and between-subject classification decisions—prevented the reliance on idiosyncratic (i.e., participant-specific) features extracted from training patterns. Though cross-validation measures on the reading set established classifier generalizability, cross-validation using the mental multiplication task replicated the finding and generated novel insight into the relationship between lexical processing and other cognitive tasks. The shared processing elements between the tasks remain unknown, and future studies should explore the extent to which reading-skill related connectivity influences brain processing dynamics in other cognitive tasks or the default mode network.

Finally, as with any study using intact groups, this correlational study cannot claim a causal relationship between the characteristic connectivity patterns and reading skill. Moreover, the most diagnostic functional connections are elements of larger patterns of predictive networks. Thus, even if causality could be established using manipulations that temporarily impact interregional communication (e.g., TMS), modification of individual pathways may be necessary but not sufficient to impact reading ability. Moreover, it is unclear how predictions made from simulated impairment should be interpreted, given that disruption of an intact network would correspond to an acquired dyslexia, which has a different profile from the developmental dyslexia that is the focus of this study (Baddeley et al., 1982). Nonetheless, hubs for multiply-predictive functional connections, such as the left Jensen sulcus, are an intriguing target for experimental manipulations better suited for exploring causal relationships underlying brain-behavior correlations that may drive reading skill.

CONCLUSIONS

A multilayer perceptron classifier concurrently learned the functional connectivity fingerprints of poor and highly-skilled reading and of word and pseudoword processing within the functionally-defined reading network. These connectivity fingerprints were identifiable among functional connectivity measured in a mental multiplication task, and bore a very strong associative relationship with reading skill and a moderately strong associative relationship with lexicality processing. These results suggest that the manner in which reading skill reciprocally shapes functional connectivity in the reading network impacts dynamic brain organization in other cognitive domains, providing a path by which the uniquely human capacity for written language may influence human cognition in general.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://openneuro.org/datasets/ds001894/versions/1.3.1; https://openneuro.org/datasets/ds001486/versions/1.2.2.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of Northwestern University. Written informed consent to participate in the study was obtained from the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

CM was solely responsible for all elements of the experiment and manuscript.

REFERENCES

Allen, G. I. (2013). Automatic feature selection via weighted kernels and regularization. *J. Comp. Graph. Stat.* 22, 284–299. doi: 10.1080/10618600.2012.681213

Archibald, L. M., Cardy, J. O., Joanisse, M. F., and Ansari, D. (2013). Language, reading, and math learning profiles in an epidemiological sample of school age children. *PLoS ONE* 8:e77463. doi: 10.1371/journal.pone.0077463

Baddeley, A. D., Ellis, N. C., Miles, T. R., and Lewis, V. J. (1982). Developmental and acquired dyslexia: a comparison. *Cognition* 11, 185–199. doi: 10.1016/0010-0277(82)90025-7

Bailey, S., Hoeft, F., Aboud, K., and Cutting, L. (2016). Anomalous gray matter patterns in specific reading comprehension deficit are independent of dyslexia. *Ann. Dyslexia* 66, 256–274. doi: 10.1007/s11881-015-0114-y

Blau, V., Reithler, J., van Atteveldt, N. M., Seitz, J., Gerretsen, P., Goebel, R., et al. (2010). Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. *Brain* 133(Pt 3), 868–879. doi: 10.1093/brain/awp308

Boets, B., Op de Beeck, H. P., Vandermosten, M., Scott, S. K., Gillebert, C. R., Mantini, D., et al. (2013). Intact but less accessible phonetic representations in adults with dyslexia. *Science* 342, 1251–1254. doi: 10.1126/science.1244333

Bouhali, F., Thiebaut de Schotten, M., Pinel, P., Poupon, C., Mangin, J.-F., Dehaene, S., et al. (2014). Anatomical connections of the visual word form area. *J. Neurosci.* 34, 15402–15414. doi: 10.1523/JNEUROSCI.4918-13.2014

Caruana, R., Lawrence, S., and Giles, C. L. (2001). "Overfitting in neural nets: backpropagation, conjugate gradient, and early stopping," in *Paper presented at the Advances in Neural Information Processing Systems* (Como). doi: 10.1109/IJCNN.2000.857823

Chu, C., Hsu, A.-L., Chou, K.-H., Bandettini, P., and Lin, C. (2012). Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage* 60, 59–70. doi: 10.1016/j.neuroimage.2011.11.066

Dehaene, S., Cohen, L., Sigman, M., and Vinckier, F. (2005). The neural code for written words: a proposal. *Trends Cogn. Sci.* 9, 335–341. doi: 10.1016/j.tics.2005.05.004

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., et al. (2010). How learning to read changes the cortical networks for vision and language. *Science* 330, 1359–1364. doi: 10.1126/science.1194140

Del Tufo, S. N., and Myers, E. B. (2014). Phonemic restoration in developmental dyslexia. *Front. Neurosci.* 8:134. doi: 10.3389/fnins.2014.00134

Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53, 1–15. doi: 10.1016/j.neuroimage.2010.06.010

Diamantidis, N. A., Karlis, D., and Giakoumakis, E. A. (2000). Unsupervised stratification of cross-validation for accuracy estimation. *Artif. Intell* 116, 1–16. doi: 10.1016/S0004-3702(99)00094-6

Edwards, E. S., Burke, K., Booth, J. R., and McNorgan, C. (2018). Dyslexia on a continuum: a complex network approach. *PLoS ONE* 13:e0208923. doi: 10.1371/journal.pone.0208923

Esposito, F., Bertolino, A., Scarabino, T., Latorre, V., Blasi, G., Popolizio, T., et al. (2006). Independent component model of the default-mode brain function: Assessing the impact of active thinking. *Brain Res. Bull* 70, 263–269. doi: 10.1016/j.brainresbull.2006.06.012

Fletcher, J. M. (2005). Predicting math outcomes: reading predictors and comorbidity. *J. Learn. Disabil.* 38, 308–312. doi: 10.1177/00222194050380040501

Fraga González, G., Van der Molen, M. J. W., Žari,ć, G., Bonte, M., Tijms, J., Blomert, L., et al. (2016). Graph analysis of EEG resting state functional networks in dyslexic readers. *Clin. Neurophysiol.* 127, 3165–3175. doi: 10.1016/j.clinph.2016.06.023

Gonzalez-Castillo, J., and Bandettini, P. A. (2018). Task-based dynamic functional connectivity: recent findings and open questions. *Neuroimage* 180(Pt B), 526–533. doi: 10.1016/j.neuroimage.2017.08.006

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol* 6:e159. doi: 10.1371/journal.pbio.0060159

Hagmann, P., Sporns, O., Madan, N., Cammoun, L., Pienaar, R., Wedeen, V. J., et al. (2010). White matter maturation reshapes structural connectivity in the late developing human brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19067–19072. doi: 10.1073/pnas.1009073107

Hoeft, F., McCandliss, B. D., Black, J. M., Gantman, A., Zakerani, N., Hulme, C., et al. (2011). Neural systems predicting long-term outcome in dyslexia. *Proc. Natl. Acad. Sci. U.S.A.* 108, 361–366. doi: 10.1073/pnas.1008950108

Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., et al. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2035–2040. doi: 10.1073/pnas.0811168106

Horowitz-Kraus, T., DiFrancesco, M., Kay, B., Wang, Y., and Holland, S. K. (2015). Increased resting-state functional connectivity of visual- and cognitive-control brain networks after training in children with reading difficulties. *NeuroImage* 8, 619–630. doi: 10.1016/j.nicl.2015.06.010

Hyvärinen, A., and Bingham, E. (2003). Connection between multilayer perceptrons and regression using independent component analysis. *Neurocomputing* 50, 211–222. doi: 10.1016/S0925-2312(01)00705-6

Ioffe, S., and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv*:1502.03167.

Knudde, N., van der Herten, J., Dhaene, T., and Couckuyt, I. (2017). GPflowOpt: a Bayesian optimization library using tensorflow. *arXiv preprint arXiv*:1711.03845.

Koistinen, P., and Holmström, L. (1991). "Kernel regression and backpropagation training with noise," in *Paper presented at the Advances in Neural Information Processing Systems* (Singapore). doi: 10.1109/IJCNN.1991.170429

Korenberg, M. J., and Hunter, I. W. (1990). The identification of nonlinear biological systems: Wiener kernel approaches. *Ann. Biomed. Eng.* 18, 629–654. doi: 10.1007/BF02368452

Larsen, J., Hansen, L. K., Svarer, C., and Ohlsson, M. (1996). "Design and regularization of neural networks: the optimal use of a validation set," in *Paper presented at the Neural Networks for Signal Processing VI. Proceedings of the 1996 IEEE Signal Processing Society Workshop* (San Francisco, CA).

Lemley, J., Bazrafkan, S., and Corcoran, P. (2017). Smart augmentation learning an optimal data augmentation strategy. *IEEE Access* 5, 5858–5869. doi: 10.1109/ACCESS.2017.2696121

Li, W. (1990). Mutual information functions versus correlation functions. *J. Stat. Phys* 60, 823–837. doi: 10.1007/BF01025996

Lytle, M. N., McNorgan, C., and Booth, J. R. (2019). A longitudinal neuroimaging dataset on multisensory lexical processing in school-aged children. *Sci. Data* 6:329. doi: 10.1038/s41597-019-0338-5

McNorgan, C., Alvarez, A., Bhullar, A., Gayda, J., and Booth, J. R. (2011). Prediction of reading skill several years later depends on age and brain region: implications for developmental models of reading. *J. Neurosci.* 31, 9641–9648. doi: 10.1523/JNEUROSCI.0334-11.2011

McNorgan, C., Awati, N., Desroches, A. S., and Booth, J. R. (2014). Multimodal lexical processing in auditory cortex is literacy skill dependent. *Cereb. Cortex* 24, 2464–2475. doi: 10.1093/cercor/bht100

McNorgan, C., and Booth, J. R. (2015). Skill dependent audiovisual integration in the fusiform induces repetition suppression. *Brain Lang.* 141, 110–123. doi: 10.1016/j.bandl.2014.12.002

McNorgan, C., and Joanisse, M. F. (2014). A connectionist approach to mapping the human connectome permits simulations of neural activity within an artificial brain. *Brain Connect* 4, 40–52. doi: 10.1089/brain.2013.0174

McNorgan, C., Judson, C., Handzlik, D., and Holden, J. G. (2020a). Linking ADHD and behavioral assessment through identification of shared diagnistic task-based functional connections. *Front. Physiol.* 11:583005. doi: 10.3389/fphys.2020.583005

McNorgan, C., Randazzo-Wagner, M., and Booth, J. R. (2013). Cross-modal integration in the brain is related to phonological awareness only in typical readers, not in those with reading difficulty. *Front. Hum. Neurosci.* 7:388. doi: 10.3389/fnhum.2013.00388

McNorgan, C., Smith, G. J., and Edwards, E. S. (2020b). Integrating functional connectivity and MVPA through a multiple constraint network analysis. *Neuroimage* 208:116412. doi: 10.1016/j.neuroimage.2019.116412

Mills, J. R., and Jackson, N. E. (1990). Predictive significance of early giftedness: the case of precocious reading. *J. Educ. Psychol.* 82, 410–419. doi: 10.1037/0022-0663.82.3.410

Mokhtari, F., and Hossein-Zadeh, G.-A. (2013). Decoding brain states using backward edge elimination and graph kernels in fMRI connectivity networks. *J. Neurosci. Methods* 212, 259–268. doi: 10.1016/j.jneumeth.2012.10.012

Morken, F., Helland, T., Hugdahl, K., and Specht, K. (2017). Reading in dyslexia across literacy development: a longitudinal study of effective connectivity. *Neuroimage* 144, 92–100. doi: 10.1016/j.neuroimage.2016.09.060

Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of FMRI data. *Trends Cogn. Sci.* 10, 424–430. doi: 10.1016/j.tics.2006.07.005

Norton, E. S., Beach, S. D., and Gabrieli, J. D. E. (2015). Neurobiology of dyslexia. *Curr. Opin. Neurobiol.* 30, 73–78. doi: 10.1016/j.conb.2014.09.007

Paluš, M. (1997). Detecting phase synchronization in noisy systems. *Phys. Lett. A* 235, 341–351. doi: 10.1016/S0375-9601(97)00635-X

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., et al. (2018). A meta-analysis on the relation between reading and working memory. *Psychol. Bull* 144:48. doi: 10.1037/bul0000124

Perfetti, C. A., Liu, Y., Fiez, J., Nelson, J., Bolger, D. J., and Tan, L.-H. (2007). Reading in two writing systems: accommodation and assimilation of the brain's reading network. *Bilingualism* 10, 131–146. doi: 10.1017/S1366728907002891

Perfetti, C. A., and Tan, L.-H. (2013). Write to read: the brain's universal reading and writing network. *Trends Cogn. Sci.* 17, 56–57. doi: 10.1016/j.tics.2012.12.008

Poggio, T., Mhaskar, H., Rosasco, L., Miranda, B., and Liao, Q. (2017). Why and when can deep-but not shallow-networks avoid the curse of dimensionality: a review. *Int. J. Auto. Comp.* 14, 503–519. doi: 10.1007/s11633-017-1054-2

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., and Gabrieli, J. D. E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10, 15–35. doi: 10.1006/nimg.1999.0441

Price, C. J., and Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends Cogn. Sci.* 15, 246–253. doi: 10.1016/j.tics.2011.04.001

Pugh, K. R., Mencl, W. E., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Constable, R. T., et al. (2000). The angular gyrus in developmental dyslexia: task-specific differences in functional connectivity within posterior cortex. *Psychol. Sci.* 11, 51–56. doi: 10.1111/1467-9280.00214

Randazzo, M., Greenspon, E. B., Booth, J. R., and McNorgan, C. (2019). Children with reading difficulty rely on unimodal neural processing for phonemic awareness. *Front. Hum. Neurosci.* 13:390. doi: 10.3389/fnhum.2019.00390

Richardson, F. M., Seghier, M. L., Leff, A. P., Thomas, M. S. C., and Price, C. J. (2011). Multiple routes from occipital to temporal cortices during reading. *J. Neurosci.* 31, 8239–8247. doi: 10.1523/JNEUROSCI.6519-10.2011

Richlan, F. (2019). The functional neuroanatomy of letter-speech sound integration and its relation to brain abnormalities in developmental dyslexia. *Front. Hum. Neurosci.* 13:21. doi: 10.3389/fnhum.2019.00021

Richlan, F., Kronbichler, M., and Wimmer, H. (2009). Functional abnormalities in the dyslexic brain: a quantitative meta-analysis of neuroimaging studies. *Hum. Brain Mapp* 30, 3299–3308. doi: 10.1002/hbm.20752

Ruck, D. W., Rogers, S. K., Kabrisky, M., Oxley, M. E., and Suter, B. W. (1990). The multilayer perceptron as an approximation to a Bayes optimal discriminant function. *IEEE Trans. Neural Netw.* 1, 296–298. doi: 10.1109/72.80266

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1985). *Learning Internal Representations by Error Propagation*. San Diego, CA: University of California San Diego. doi: 10.21236/ADA164453

Shahin, A. J., Bishop, C. W., and Miller, L. M. (2009). Neural mechanisms for illusory filling-in of degraded speech. *Neuroimage* 44, 1133–1143. doi: 10.1016/j.neuroimage.2008.09.045

Shaywitz, S. E., and Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biol. Psychiatry* 57, 1301–1309. doi: 10.1016/j.biopsych.2005.01.043

Smith, G. J., Booth, J. R., and McNorgan, C. (2018). Longitudinal task-related functional connectivity changes predict reading development. *Front. Psychol.* 9, 1–13. doi: 10.3389/fpsyg.2018.01754

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.

Suárez-Pellicioni, M., Lytle, M., Younger, J. W., and Booth, J. R. (2019). A longitudinal neuroimaging dataset on arithmetic processing in school children. *Sci. Data* 6:190040. doi: 10.1038/sdata.2019.40

Temple, E., Poldrack, R. A., Salidis, J., Deutsch, G. K., Tallal, P., Merzenich, M. M., et al. (2001). Disrupted neural responses to phonological and orthographic processing in dyslexic children: an FMRI study. *Neuroreport* 12, 299–307. doi: 10.1097/00001756-200102120-00024

Turkeltaub, P. E., Gareau, L., Flowers, D. L., Zeffiro, T. A., and Eden, G. F. (2003). Development of neural mechanisms for reading. *Nat. Neurosci.* 6, 767–773. doi: 10.1038/nn1065

van der Mark, S., Bucher, K., Maurer, U., Schulz, E., Brem, S., Buckelmüller, J., et al. (2009). Children with dyslexia lack multiple specializations along the visual word-form (vwf) system. *Neuroimage* 47, 1940–1949. doi: 10.1016/j.neuroimage.2009.05.021

van der Mark, S., Klaver, P., Bucher, K., Maurer, U., Schulz, E., Brem, S., et al. (2011). The left occipitotemporal system in reading: disruption of focal fmri connectivity to left inferior frontal and inferior parietal language areas in children with dyslexia. *Neuroimage* 54, 2426–2436. doi: 10.1016/j.neuroimage.2010.10.002

Wagner, R. K. (1988). Causal relations between the development of phonological processing abilities and the acquisition of reading skills: a meta-analysis. *Merrill-Palmer Q.* 1982, 261–279.

Wolf, R. C., Sambataro, F., Lohr, C., Steinbrink, C., Martin, C., and Vasic, N. (2010). Functional brain network abnormalities during verbal working memory performance in adolescents and young adults with dyslexia. *Neuropsychologia* 48, 309–318. doi: 10.1016/j.neuropsychologia.2009.09.020

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.

ORIGINAL RESEARCH
published: 08 April 2021
doi: 10.3389/fncom.2021.650050

A Deep Learning Approach for Automatic Seizure Detection in Children With Epilepsy

Ahmed Abdelhameed and Magdy Bayoumi*

Department of Electrical and Computer Engineering, University of Louisiana at Lafayette, Lafayette, LA, United States

Over the last few decades, electroencephalogram (EEG) has become one of the most vital tools used by physicians to diagnose several neurological disorders of the human brain and, in particular, to detect seizures. Because of its peculiar nature, the consequent impact of epileptic seizures on the quality of life of patients made the precise diagnosis of epilepsy extremely essential. Therefore, this article proposes a novel deep-learning approach for detecting seizures in pediatric patients based on the classification of raw multichannel EEG signal recordings that are minimally pre-processed. The new approach takes advantage of the automatic feature learning capabilities of a two-dimensional deep convolution autoencoder (2D-DCAE) linked to a neural network-based classifier to form a unified system that is trained in a supervised way to achieve the best classification accuracy between the ictal and interictal brain state signals. For testing and evaluating our approach, two models were designed and assessed using three different EEG data segment lengths and a 10-fold cross-validation scheme. Based on five evaluation metrics, the best performing model was a supervised deep convolutional autoencoder (SDCAE) model that uses a bidirectional long short-term memory (Bi-LSTM) – based classifier, and EEG segment length of 4 s. Using the public dataset collected from the Children's Hospital Boston (CHB) and the Massachusetts Institute of Technology (MIT), this model has obtained 98.79 ± 0.53% accuracy, 98.72 ± 0.77% sensitivity, 98.86 ± 0.53% specificity, 98.86 ± 0.53% precision, and an F1-score of 98.79 ± 0.53%, respectively. Based on these results, our new approach was able to present one of the most effective seizure detection methods compared to other existing state-of-the-art methods applied to the same dataset.

Keywords: deep learning, epileptic seizure detection, EEG, autoencoders, classification, convolutional neural network (CNN), bidirectional long short term memory (Bi LSTM)

OPEN ACCESS

Edited by:
Saman Sargolzaei,
University of Tennessee at Martin,
United States

Reviewed by:
Nhan Duy Truong,
The University of Sydney, Australia
Antonio Dourado,
University of Coimbra, Portugal

****Correspondence:***
Ahmed Abdelhameed
ahmed.abdelhameed1@louisiana.edu

Received: *06 January 2021*
Accepted: *15 March 2021*
Published: *08 April 2021*

Citation:
Abdelhameed A and Bayoumi M (2021) A Deep Learning Approach for Automatic Seizure Detection in Children With Epilepsy. Front. Comput. Neurosci. 15:650050. doi: 10.3389/fncom.2021.650050

INTRODUCTION

Epilepsy is inevitably recognized to be one of the most critical and persistent neurological disorders affecting the human brain. It has spread to more than 50 million patients of various ages worldwide (World Health Organization, 2020) with approximately 450,000 patients under the age of 17 in the United States out of nearly 3 million American patients diagnosed with this disease (Epilepsy in Children, 2020). Epilepsy can be characterized apparently by its recurrent unprovoked seizures. A seizure is a period of anomalous, synchronous innervation of a population of neurons that may last from seconds to a few minutes. Epileptic seizures are ephemeral instances of partial or complete abnormal unintentional movements of the body that may also be combined with a loss of consciousness. While epileptic seizures rarely occur in each patient, their ensuing effects on the patients' emotions, social interactions, and physical communications make diagnosis and treatment of epileptic seizures of ultimate significance.

Electroencephalograms (EEGs; Schomer and Lopez da Silva, 2018) which have been around for a long time, are commonly used among neurologists to diagnose several brain disorders and in particular, epilepsy attributable to workable reasons, such as its availability, effortlessness, and low cost. EEG operates by positioning several electrodes along the surface of the human scalp and then recording and measuring the voltage oscillations emanating from the ion current flowing through the brain. These voltage oscillations, which correspond to the neuronal activity of the brain, are then transformed into multiple time series called signals. EEG is a very powerful non-invasive diagnostic tool since we can use it precisely to capture and denote epileptic signals that are characterized by spikes, sharp waves, or spike-and-wave complexities. As a result, EEG signals have been the most widely used in the clinical examination of various epileptic brain states, for both the detection and prediction of epileptic seizures.

By interpreting the recorded EEG signals visually, neurologists can substantially distinguish between epileptic brain activities during a seizure (ictal) state and normal brain activities between seizures (interictal) state. Over the last two decades, however, an abundance of automated EEG-based epilepsy diagnostic studies has been established. This was motivated by the exhausting and time-consuming nature of the human visual evaluation process that depends mainly on the doctors' expertise. Besides that, the need for objective, rapid, and effective systems for the processing of vast amounts of EEG recordings has become unavoidable to be able to diminish the possibility of misinterpretations. The availability of such systems would greatly enhance the quality of life of epileptic patients.

Following the acquisition and pre-processing of EEG raw signals, most of the automated seizure detection techniques consist of two key successive stages. The first stage concerns the extraction and selection of certain features of the EEG signals. In the second step, a classification system is then built and trained to utilize these extracted features for the detection of epileptic activities. The feature extraction step has a direct effect on the precision and sophistication of the developed automatic seizure detection technique. Due to the non-stationary property of the EEG signals, the feature extraction stage typically involves considerable work and significant domain-knowledge to study and analyze the signals either in the time domain, the frequency domain, or in the time-frequency domain (Acharya et al., 2013). Predicated on this research, it has become the mission of the system designer to devise the extraction of the best-representing features that can precisely discriminate between the epileptic brain states from the EEG signals of different subjects.

In the literature, several EEG signal features extracted by various methods have been proposed for seizure detection. For example (Song et al., 2012), used approximate entropy and sample entropy as EEG features, and integrated them with an extreme learning machine (ELM) for the automated detection of epileptic seizures. Chen et al. (2017) used non-subsampled wavelet–Fourier features for seizure detection. Wang et al. (2018) proposed an algorithm that combines wavelet decomposition and the directed transfer function (DTF) for feature extraction. Raghu et al. (2019) proposed using matrix determinant as a feature for the analysis of epileptic EEG signals. Certainly, even with the achievement of great results, it is not inherently guaranteed that the features derived through the intricate, and error-prone manual feature extraction methodology would yield the maximum possible classification accuracy. As such, it would be very fitting to work out how to build substantial systems that can automatically learn the best representative features from minimally preprocessed EEG signals while at the same time realize optimum classification performance.

The recent advances in machine learning science and particularly the deep learning techniques breakthroughs have shown its superiority for automatically learning very robust features that outperformed the human-engineered features in many fields such as speech recognition, natural language processing, and computer vision as well as medical diagnosis (Wang et al., 2020). Multiple seizure detection systems that used artificial neural networks (ANNs) as classifiers, after traditional feature extraction, were reported in previous work. For instance (Orhan et al., 2011), used multilayer perceptron (MLP) for classification after using discrete wavelet transform (DWT) and K-means algorithm for feature extraction. Samiee et al. (2015) also used MLP as a classifier after using discrete short-time Fourier transform (DSTFT) for feature extraction. In Jaiswal and Banka (2017), ANNs were evaluated for classification after using the local neighbor descriptive pattern (LNDP) and one-dimensional local gradient pattern (1D-LGP) techniques for feature extraction. Yavuz et al. (2018) performed cepstral analysis utilizing generalized regression neural network for EEG signals classification. On the other hand, convolutional neural networks (CNNs) were adopted for both automatic feature learning and classification. For example (Acharya et al., 2018), proposed a deep CNN consisting of 13 layers for automatic seizure detection. For the same purpose (Abdelhameed et al., 2018a), designed a system that combined a one-dimensional CNN with a bidirectional long short-term memory (Bi-LSTM) recurrent neural network. Ke et al. (2018); Zhou et al. (2018), and Hossain et al. (2019) also used CNN for feature extraction and classification. In Hu et al. (2019), CNN and support vector machine (SVM) were incorporated together for feature extraction and classification of EEG signals.

As reported, most of the deep learning algorithms that involved automatic feature learning have targeted single-channel epileptic EEG signals. It is therefore still important to research more data-driven algorithms that can handle more complex multichannel epileptic EEG signals.

In general, supervised learning is the most widely used technique for classifying EEG signals among all other machine learning techniques. Several researchers have recently experimented with semi-supervised deep learning strategies in which an autoencoder (AE) neural network can benefit from training using both unlabeled and labeled data to improve the efficacy of the classification process (Gogna et al., 2017;

Yuan et al., 2017; Abdelhameed and Bayoumi, 2018, 2019; She et al., 2018). Two approaches of using AEs have been used in the literature. The first one is the stacked AEs approach, where each layer of a neural network consisting of multiple hidden layers is trained individually using an AE in an unsupervised way. After that, all trained layers are stacked together and a softmax layer is attached to form a stacked network that is finally trained in a supervised fashion. The second approach uses deep AEs to pre-train all layers of the neural network simultaneously instead of that greedy layer-wise training. This latter approach still suffers from one particular drawback which is the necessity to train the semi-supervised deep learning model twice. One training episode is conducted in an unsupervised way using unlabeled training data that enables the AE to learn good initial parameters (weights). In the second episode, the first half of the pre-trained AE (the encoder network) attached to a selected classifier is trained as a new system in a supervised manner using labeled data to perform the final classification task.

Therefore, in this work, to address the limitation of the classification schemes alluded to above, a novel deep learning-based system that uses a two-dimensional supervised deep convolutional autoencoder (2D-SDCAE) is proposed for the detection of epileptic seizures in multichannel EEG signals recordings. The innovative approach in the proposed system is that the AE is trained only once in a supervised way to perform two tasks at the same time. The first task is to automatically learn the best features from the EEG signals and to summarize them in a succinct, low-dimensional, latent space representation while performing the classification task efficiently. The method of consolidating the simultaneous learning to perform both tasks in a single model, which is trained only once in a supervised way, has proven to have a good impact on improving the learning capabilities of the model and thus achieving better classification accuracy.

In addition to operating directly on raw EEG signal data, there are several advantages to our approach. First of all, the SDCAE is faster compared to conventional semi-supervised classification systems since it is trained only once. Second, to minimize the total number of network parameters, the proposed SDCAE uses convolutional layers for learning features instead of fully connected layers that are commonly used in regular MLP-based AEs. Third, the proposed system can be used in signal compression schemes as the original high-dimensional signals can be perfectly reconstructed from the low-dimensional latent representation using the second half of the AE (the decoder network). Finally, the training of AEs in a supervised way is more effective in learning more structured latent representation, making it very feasible to deploy very simple classifiers and still have very high-precision seizure detection systems. It is also worth noting that performance and hardware resource-saving have been taken into account to make the proposed system more suitable for real-time use and potential hardware implementation and deployment.

Two SDCAE models are designed to test our novel approach, and their performance for seizure detection in children is evaluated. Both models are used to classify EEG data segments to distinguish between ictal and interictal brain states. The first model is a two-dimensional deep convolution autoencoder (2D-DCAE) in which the convolutional layers of the encoder network are attached to a simple MLP network consisting of two fully connected hidden layers and one output layer for classification. The second system is also a 2D-DCAE but in this system, the convolutional layers of the encoder network are attached to one Bi-LSTM recurrent neural network layer to do the classification task. Besides, the performance of both proposed models is further compared to two regular deep learning models having the same layers' structure, except that the decoder network layers are completely removed. These two models are trained in a supervised manner to only do the classification task. By quantitatively evaluating the performance of the proposed models using different EEG segment lengths, our new approach of using SDCAE will prove to be a very good candidate for producing one of the most accurate seizure detection systems.

MATERIALS AND METHODS

Dataset

Patients' data obtained from the online Children's Hospital Boston–Massachusetts Institute of Technology (CHB–MIT) Database were used to assess and measure the efficacy of the proposed models. The dataset is recorded at Boston Children's Hospital and consists of long-term EEG scalp recordings of 23 pediatric patients with intractable seizures (Shoeb, 2009). 23 channels EEG signals recordings are collected using 21 electrodes whose names are specified by the International 10–20 electrode positioning system using the modified combinatorial nomenclature as shown in **Figure 1**. The signals are then sampled at 256 Hz and the band-pass filtered between 0 and 128 Hz.

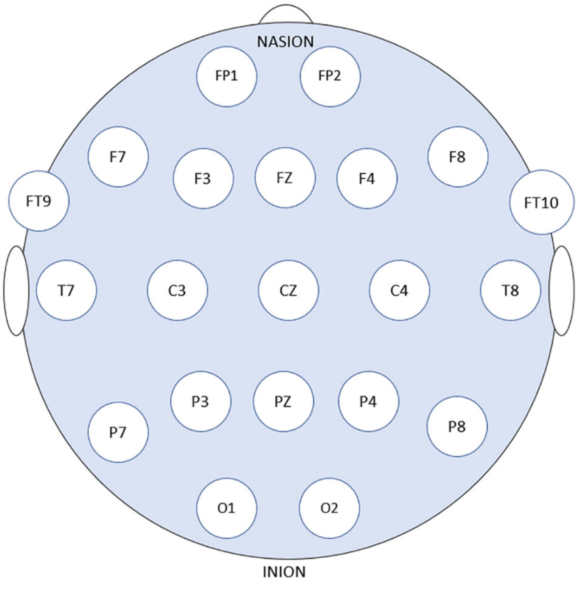

FIGURE 1 | 21 EEG electrode positions based on the 10–20 system using modified combinatorial nomenclature.

In this study, 16 out of the 23 pediatric patients are selected for the assessment of the classification models. More details about the selected patients are listed in **Table 1**. Seizures less than 10 s are too short so, all Chb16's seizures were not considered for testing (Gao et al., 2020). The seizures of the two patients (Chb12 and Chb13) were omitted due to the excess variations in channel naming and electrode positioning swapping. Four patients (Chb04, Chb15, Chb18, and Chb19) have been excluded since they are 16 years of age and older because the aim is to research seizure detection in young children.

Typically, epileptic patients have limited numbers of seizures that span much shorter times relative to seizure-free periods. A discrepancy between the number of ictal and interictal EEG data segments is often present. To surmount the bias in the training process of the classification models in which classifiers tend to favor the class with the largest number of segments, and as a design choice, the number of interictal segments is chosen to be equal to the number of ictal segments while forming the final dataset. Downsampling the original interictal dataset can be found in previous work as in Wei et al. (2019), Gao et al. (2020). Non-overlapped EEG segments of 1, 2, and 4 s duration were tested for evaluating the proposed models. A single EEG segment is represented as a matrix whose dimension is ($L \times N$) where L is the sequence length = 256 × segment duration and N is the number of channels. As an example, one 2-s segment is represented as a 512 × 23 matrix. The EEG dataset is then formed by putting all the ictal and interictal segments in one matrix whose dimension is ($2KL \times 23$) where K is the number of the ictal or interictal segments and L is as defined before.

Dataset Preparation

To prepare the EEG dataset before the training phase, all segments combined are pre-processed by applying z-score normalization for all channels at one to ensure that all values are standardized by having a zero mean (μ) and unit standard deviation (σ) using the Eq. (1)

$$x = \frac{x - \mu}{\sigma} \quad (1)$$

Next, as a batch, the whole dataset values are scaled to the [0, 1] range using Min–Max normalization to ensure that the original and the reconstructed segments have the same range of values. Finally, the channel's dimension of the segments is extended by one column to be more suitable for the AE to be used.

TABLE 1 | Seizure information of the selected patients.

Patient#	Gender-age	Number of seizures	Total seizures duration (s)
Chb01	F-11	7	442
Chb02	M-11	2	172
Chb03	F-14	7	402
Chb05	F-7	5	558
Chb06	F-1.5	10	153
Chb07	F-14.5	3	325
Chb08	M-3.5	5	919
Chb09	F-10	4	276
Chb10	M-3	7	447
Chb11	F-12	3	806
Chb14	F-9	8	169
Chb17	F-12	3	293
Chb20	F-6	8	294
Chb21	F-13	4	199
Chb22	F-9	3	204
Chb23	F-6	7	424
Total		86	6083

Proposed Systems Architecture

The objective of the article is to build accurate and reliable deep learning models for epileptic seizure detection based on differentiating between two classes of epileptic brain states, interictal and ictal. The proposed models automatically learn powerful features that help to achieve a high classification accuracy of minimally pre-processed EEG signals. Our target is to eliminate the overhead induced by the exhausting manual feature extraction process and also replacing complex systems that require long training times with a much simpler, faster, and more efficient system that benefits from the structure and functionality of AEs. An AE neural network consists of two subnetworks: an encoder and a decoder. The encoder network is used for compressing (encoding) the input information (EEG signals in our case) into a lower-dimensional representation and the decoder is used in a reverse way to decompress or reconstruct the original signal. AE-based compression is accomplished by continually training a network to reconstruct its input while trying to minimize a loss function between the original input and the reconstructed one. 2D-DCAE-based models are proposed for automatically learning inherent signal features from labeled EEG segments while being trained in a supervised way. **Figure 2** shows the block diagram of the first proposed model which consists of a 2D-DCAE where the encoder output, the latent space representation, is also fed into an MLP network to perform the classification task.

Figure 3 shows the block diagram of the second proposed model which consists of a 2D-DCAE but in this case, the encoder output which is the latent space representation is feed into a Bi-LSTM recurrent neural network to perform the classification task.

The performance of the two proposed models will be compared with two other models. One of the new models comprises a two-dimensional deep convolutional neural network (2D-DCNN) connected to an MLP, **Figure 4A**, while a 2D-DCNN is connected to a Bi-LSTM to form the second model, **Figure 4B**.

Two-Dimensional Deep Convolutional Autoencoder

Convolutional neural networks are a special class of feedforward neural networks that are very well-suited for processing multidimensional data like images or multi-channel EEG signals. Applications of CNNs in a variety of disciplines, such as computer vision and pattern recognition, have recorded very impressive outcomes (Krizhevsky et al., 2017). This is due to its great

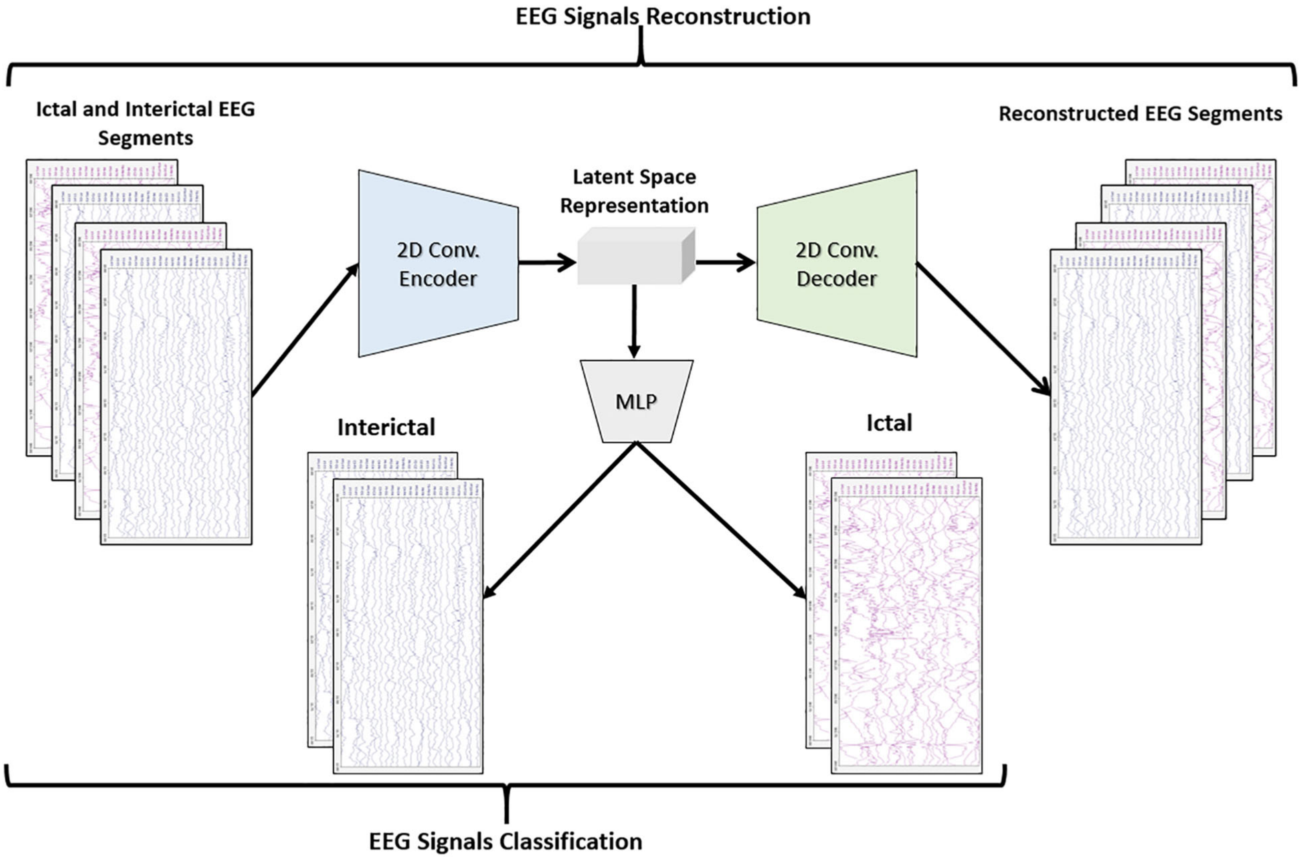
EEG Signals Reconstruction
Ictal and Interictal EEG Segments
2D Conv. Encoder
Latent Space Representation
2D Conv. Decoder
Reconstructed EEG Segments
MLP
Interictal
Ictal
EEG Signals Classification

FIGURE 2 | Block diagram of 2D-DCAE + MLP model for seizure detection.

ability to hierarchically learn excellent spatial features for the representation of data of different types. The parameter sharing and sparse connections properties of CNNs make them much more memory-savers compared to MLPs networks that consist of fully connected layers. As a result of these advantages, a two-dimensional convolution autoencoder stacked with convolution and pooling layers is proposed in this work rather than a standard AE that uses only fully connected layers.

The encoder subnetwork of the AE is a CNN consists of four convolutional layers and four max-pooling layers stacked interchangeably. The convolutional layers are responsible for learning the spatial and temporal features in the input EEG signals segments while the max-pooling layers are used for dimensionality reduction by downsampling. A single convolutional layer is made up of filters (kernels) consisting of trainable parameters (weights) that slide over and convolve with the input to generate feature maps where the number of feature maps equals the number of the applied filters. A configurable parameter (stride) controls how much the filter window is sliding over the input. The pooling layer performs down-sampling by lowering the dimension of the feature maps to reduce computational complexity. The low dimensional output of the encoding network is called latent space representation or bottleneck. On the other side, the decoder subnetwork consists of four convolutional layers and four upsampling layers which are also deployed interchangeably and are used to reconstruct the original input.

In all models, in the encoder network, the convolutional layers are configured with 32, 32, 64, and 64 filters, respectively. In the decoder network, the first three convolutional layers are configured with 64, 32, and 32 filters while the last layer has only one filter. All convolutional layers have a kernel size of 3 × 2, and a default stride value equals one. To keep the height and width of the feature maps at the same values, all convolutional layers are configured using the same padding technique. The activation function used in all convolutional layers, except the last layer, is the rectified linear unit (ReLU) defined in Eq. (2) because of its sparsity, computational simplicity, and sturdiness against noise in the input signals (Goodfellow et al., 2017).

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

where x is the weighted sum of the inputs and $f(x)$ is the ReLU activation function.

The final convolutional layer of the 2D-DCAE uses the sigmoid activation function defined in Eq. (3) to generate an

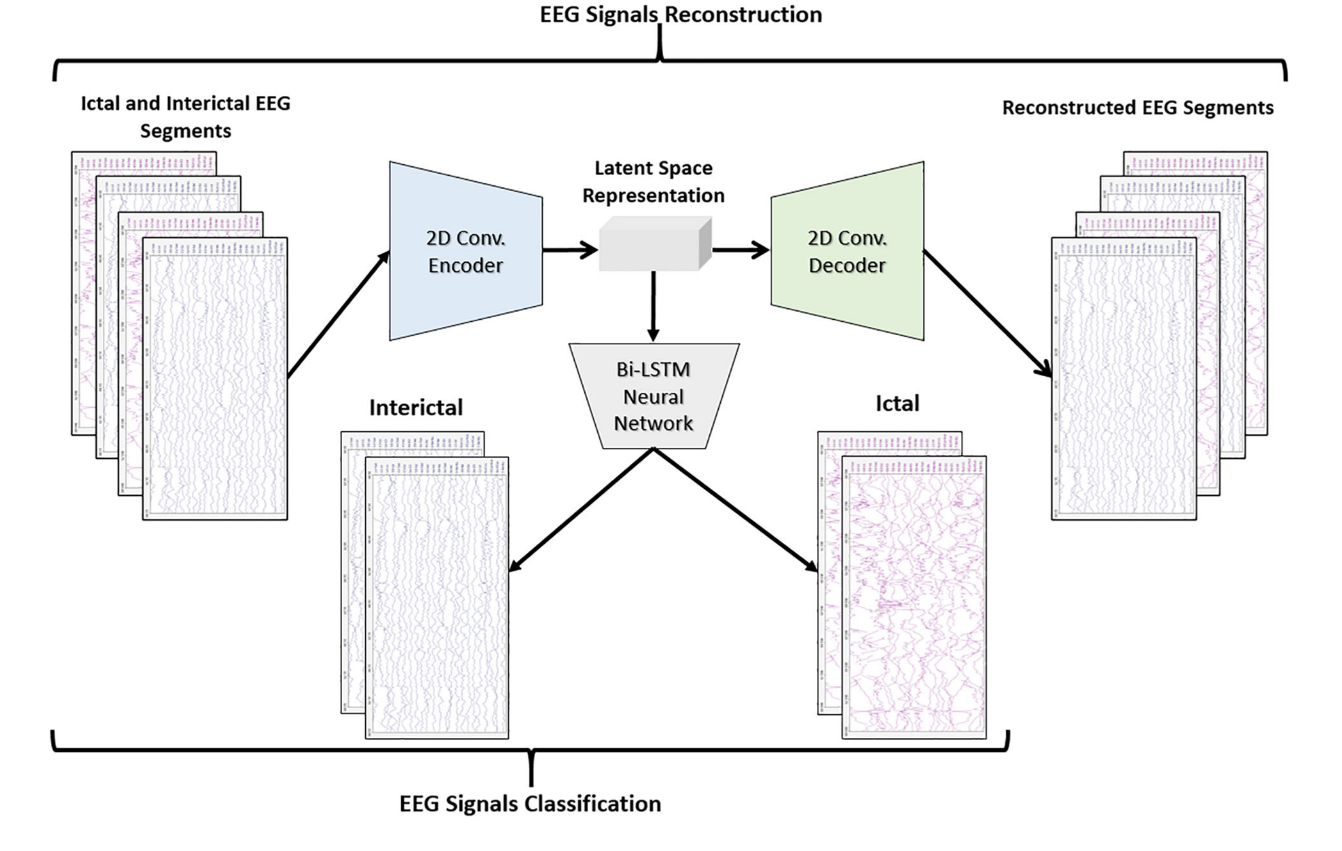
EEG Signals Reconstruction
Ictal and Interictal EEG Segments
Latent Space Representation
2D Conv. Encoder
2D Conv. Decoder
Reconstructed EEG Segments
Bi-LSTM Neural Network
Interictal
Ictal
EEG Signals Classification

FIGURE 3 | Block diagram of 2D-DCAE + Bi-LSTM model for seizure detection.

output in the range [0, 1].

$$y = \frac{1}{1 + e^{-x}} \tag{3}$$

where x is the weighted sum of the inputs and y is the output of the activation function.

All max-pooling layers are configured to perform input downsampling by taking the maximum value over windows of sizes (2, 2) except the last layer that uses a window of size (2, 3). The first upsampling layer does its job by interpolating the rows and columns of the input data using a size (2, 3) while the last three upsampling layers use (2, 2) sizes.

Our models apply the Batch Normalization (batch norm) technique for speeding up and stabilizing the training process and to ensure high performance. The batch norm transform (Ioffe and Szegedy, 2015) is defined as:

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

where an input vector x_i is normalized within a mini-batch $B = \{x_1, x_2 ... x_m\}$ having a mean $\mu_B$and variance σ_B^2. β and γ are two parameters that are learned jointly and used to scale and shift the normalized value while ∈ is a constant added for numerical stability. Four batch normalization layers are deployed between the four convolutional and max-pooling layers of the encoder subnetwork.

Proposed Classification Models

Two-Dimensional Deep Convolution Autoencoder + MLP

In the first proposed model depicted in **Figure 5**, the output of the decoder subnetwork (the latent space representation) is also converted from its multi-dimensional form to a vector using a flatten layer and then fed into an MLP network-based classifier. The MLP network consists of two hidden fully connected layers having 50 and 32 neurons (units), respectively. Both layers use the Relu activation function. The output layer of the MLP has a sigmoid activation function whose output represents the probability that an input EEG segment belongs to one of the classes.

Two-Dimensional Deep Convolution Autoencoder + Bi-LSTM

Long short-term memory (LSTM) is a particular architecture of recurrent neural networks. It was developed to solve numerous problems that vanilla RNNs suffer during training using backpropagation over Time (BPTT) (Mozer, 1989) such as information morphing and exploding and vanishing gradients (Bengio et al., 1994). By proposing the concept of memory

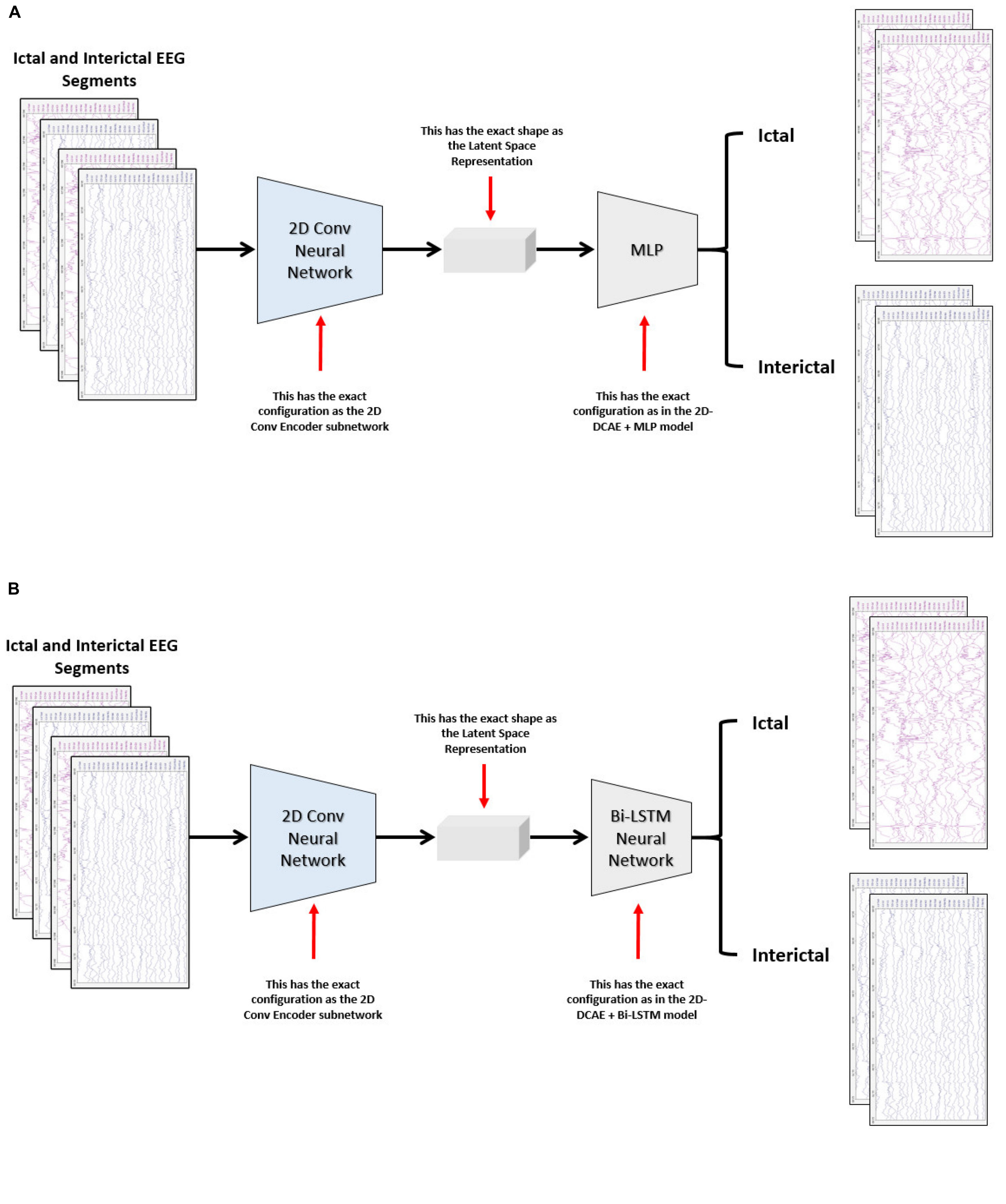
A
Ictal and Interictal EEG Segments
2D Conv Neural Network
This has the exact shape as the Latent Space Representation
MLP
Ictal
Interictal
This has the exact configuration as the 2D Conv Encoder subnetwork
This has the exact configuration as in the 2D-DCAE + MLP model
B
Ictal and Interictal EEG Segments
2D Conv Neural Network
This has the exact shape as the Latent Space Representation
Bi-LSTM Neural Network
Ictal
Interictal
This has the exact configuration as the 2D Conv Encoder subnetwork
This has the exact configuration as in the 2D-DCAE + Bi-LSTM model

FIGURE 4 | Block diagram of the two-dimensional deep convolutional neural network-based models for seizure detection, **(A)** 2D-DCNN + MLP model and **(B)** 2D-DCNN + Bi-LSTM model.

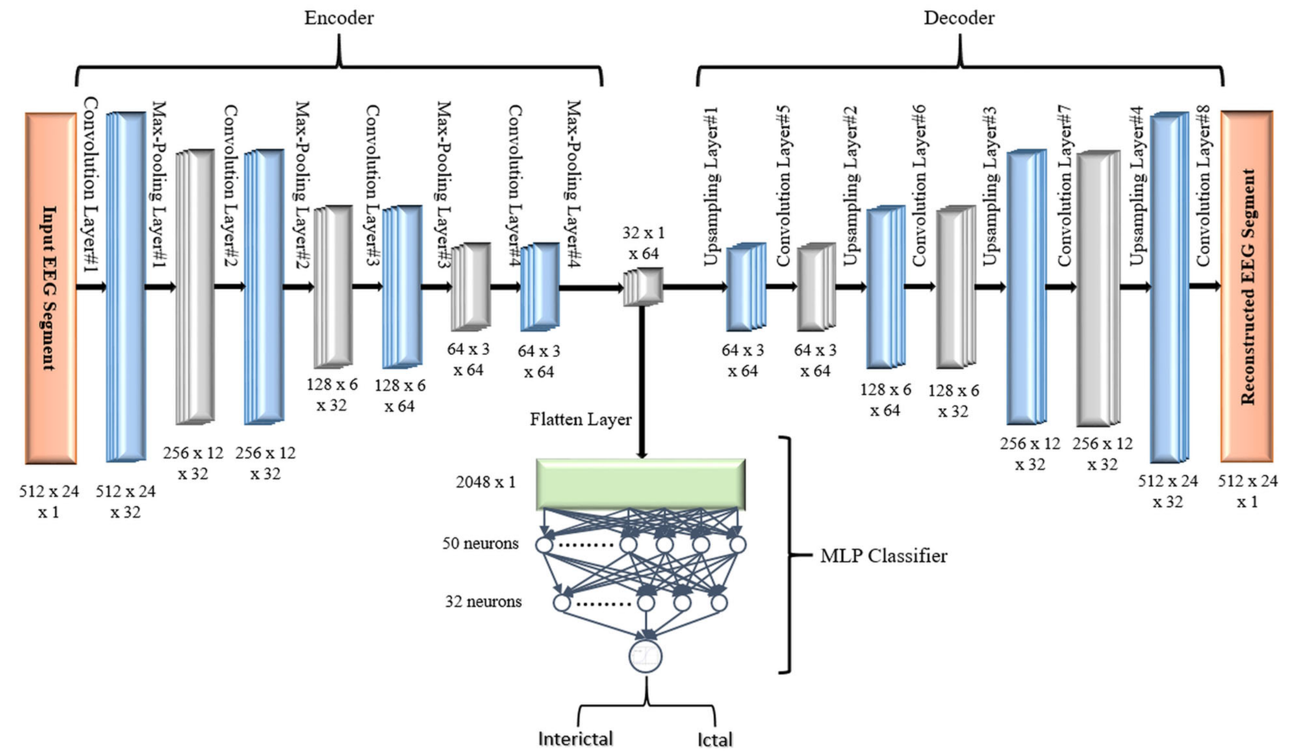

FIGURE 5 | Proposed 2D-DCAE + MLP architecture (assuming that an EEG segment length is 2 s).

cells (units) with three controlling gates, LSTMs are capable of maintaining gradients values calculated by backpropagation during network training while preserving long-term temporal dependencies between inputs (Hochreiter and Schmidhuber, 1997). **Figure 6** shows the structure of a single LSTM cell.

The following equations show how information is processed inside the LSTM cell.

$$f_t = \sigma(W_f . \left[h_{t-1},\ \ x_t\right] + b_f) \tag{5}$$

$$i_t = \sigma(W_i . \left[h_{t-1},\ x_t\right] + b_i) \tag{6}$$

$$o_t = \sigma\left(W_o . \left[h_{t-1},\ x_t\right] + b_o\right) \tag{7}$$

$$\widetilde{c}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{8}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c}_t \tag{9}$$

$$h_t = o_t \odot tanh(c_t) \tag{10}$$

where x_t is the input at time t in a sequence $X = (x_1, x_2,\ x_3,\ .,\ x_n)$ of n time steps. h_{t-1} and c_{t-1} are the hidden state output and cell state at the previous time step, respectively. h_t and c_t are the current hidden state and cell state. f_t,i_t, and o_t are the forget, input, and output gates. W and b represent the weights and biases matrices and vectors while σ is the sigmoid (logistic) function and ⊙ is the Hadamard product operator. The memory cell starts operation by selecting which information to keep or forget from the previous states using the forget gate f_t. Then, the cell calculates the candidate state $\tilde{c}_t$. After that, using the prior cell state c_{t-1} and the input gate i_t, the cell decides what further information to write to the current state c_t. Finally, the output gate o_t calculates how much state information

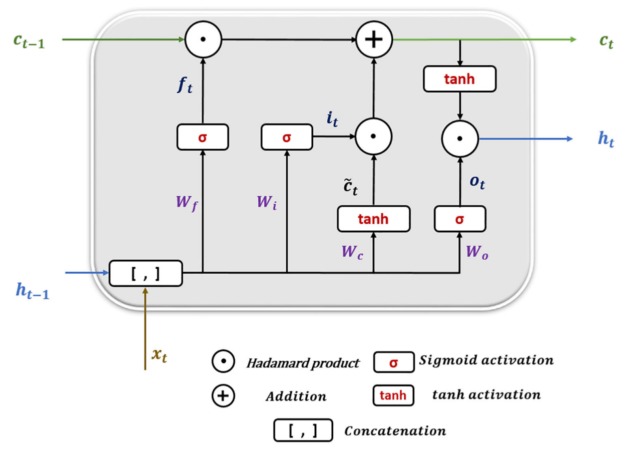

FIGURE 6 | Long short-term memory cell structure.

h_t will be transported to the next time step. Note that, in **Figure 6**, the biases and the multiplication operations between the matrix of the concatenated input and the hidden state, and the weight matrices are not shown to make the figure simpler.

In the second proposed model, the output of the decoder subnetwork is fed into a Bi-LSTM recurrent neural network-based classifier as shown in **Figure 7**.

For classification, a single-layer Bi-LSTM network consisting of two LSTM blocks (cells) is used in this model. The Bi-LSTM network architecture is similar to the standard unidirectional LSTM architecture, except that both LSTM blocks process the output of the encoder, reshaped as a sequence, simultaneously in two opposite directions instead of one direction. After passing through the entire input sequence, the average of the two outputs of both blocks concatenated together is computed and used for the classification task. Bi-LSTMs are useful in that they take into account the temporal dependence between the current input at a certain time and its previous and subsequent counterparts, which offers a strong advantage for enhancing the classification results (Abdelhameed et al., 2018b). **Figure 8** shows a single-layer Bi-LSTM network unrolled over n time steps.

The Bi-LSTM layer is configured to have 50 units and to overcome overfitting, the dropout regularization technique is used with a value of 0.1. As in the first model, the sigmoid activation function is used to predict the EEG segment class label.

Loss Functions and Optimizer

As the SDCAE is performing the two tasks of input reconstruction and classification simultaneously, both proposed models are designed to minimize two losses during network training. The first loss is the supervised classification loss (CL) between the predicted and actual class labels. The binary cross-entropy, defined in Eq. (11), is chosen as the loss function.

$$CL = -\frac{1}{N}\sum_{i=0}^{N-1} y_i \cdot \log(\hat{y}_i) \; + \; (1-y_i)\cdot\log(1-\hat{y}_i) \qquad (11)$$

where $\hat{y}_i$ is the predicted model output for a single EEG segment, and y_i is the corresponding actual class label in a training batch equals N.

The second loss is the loss of reconstruction (RL) between the input EEG segments and their reconstructed equivalents decoded by the DCAE and the mean square error defined in Eq. (12) is utilized for calculating this loss.

$$RL = \frac{1}{N}\sum_{i=0}^{N-1}\frac{1}{mn}\sum_{j=0}^{m-1}\sum_{k=0}^{n-1}\left(y_{jk} - \hat{y}_{jk}\right)^2 \qquad (12)$$

Where an y_{jk} is the original value at the position indexed by j, k in an input EEG segment matrix of size ($m \times n$), $\hat{y}_{jk}$ is the reconstructed value and N is the number of segments defined as before.

There is no much difference between training a deep learning model with a single output or a deep learning model with multiple outputs. In the latter case as in our proposed SDCAE models, the total loss (TL) of a model is calculated as the weighted summation of the CL and the reconstruction loss (RL) as in Eq. (13)

$$TL = w_c \times CL \; + \; w_r \times RL \qquad (13)$$

where w_c and w_r are the weights and can have any values in the interval (0,1]. In our design, w_c is chosen to be 0.5 while w_r equals to 1.

The backpropagation of the loss in both subnetworks starts by calculating two partial derivatives (gradients): $\frac{\partial TL}{\partial CL}$ and $\frac{\partial TL}{\partial RL}$. All other gradients are then calculated using the chaining rule and the weights and biases are then updated in the same way as typical deep learning models.

Different optimizers such as Stochastic Gradient Descent (SGD; Bottou, 2004), root mean square propagation (RMSProp; Tieleman and Hinton, 2012), ADADELTA (Zeiler, 2012), and Adam (Kingma and Ba, 2014) have been tested while training the SDCAE. Eventually, based on different models' performances, Adam optimizer was the chosen optimizer with a learning rate set at 0.0001.

Data Selection and Training

The performance of the two proposed SDCAE seizure detection models (DCAE + MLP), and (DCAE + Bi-LSTM) is evaluated against that of two regular deep learning models (DCNN + MLP), and (DCNN + Bi-LSTM) using EEG segments of three different lengths. That means a total number of twelve models will be tested and assessed using various performance measures.

A stratified 10-fold cross-validation methodology (He and Ma, 2013). is used to prepare the dataset for training and to evaluate the performance of all models to test their strength and reliability while classifying unseen data. In this methodology, the investigated EEG dataset (containing both interictal and ictal data segments) is randomly divided into ten equal subsamples or folds where the balanced distribution of both classes (ictal and interictal) is preserved within each fold. One ten percent of the dataset (a subsample) is marked as the testing set (testing fold) while the remaining nine folds of the dataset collectively are used as the training set. The cross-validation process is repeated for ten iterations, with each of the 10-folds used exactly once as the testing set. Within each iteration, all models are trained for 200 epochs using a batch size of 32. The average and standard deviation of the classification results of the 10 iterations are calculated to produce the final estimations for different evaluation measures.

Models Performance Evaluation

Various statistical metrics commonly used in the literature such as accuracy, sensitivity (recall), specificity, precision, and F1-score (Sokolova and Lapalme, 2009) have been calculated to assess the classification efficiency of the models against the testing set, in each of the ten iterations of the 10-fold cross-validation. These evaluation metrics are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \qquad (14)$$

$$Sensitivity\left(Recall\right) = \frac{TP}{TP + FN} \times 100\% \qquad (15)$$

$$Specificity = \frac{TN}{TN + FP} \times 100\% \tag{16}$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{17}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \tag{18}$$

where P denotes the number of positive (ictal) EEG segments while N denotes the number of negative (interictal) EEG segments). TP and TN are the numbers of true positives and true negatives while FP and FN are the numbers of false positives and false negatives, respectively. In this study, accuracy is defined as the percentage of the correctly classified EEG segments belonging to any state (ictal or interictal), sensitivity is the percentage of correctly classified ictal state EEG segments, specificity is the percentage of correctly classified interictal state EEG segments, while precision determines how many of the EEG segments classified as belonging to the ictal state are originally ictal state EEG segments. Finally,

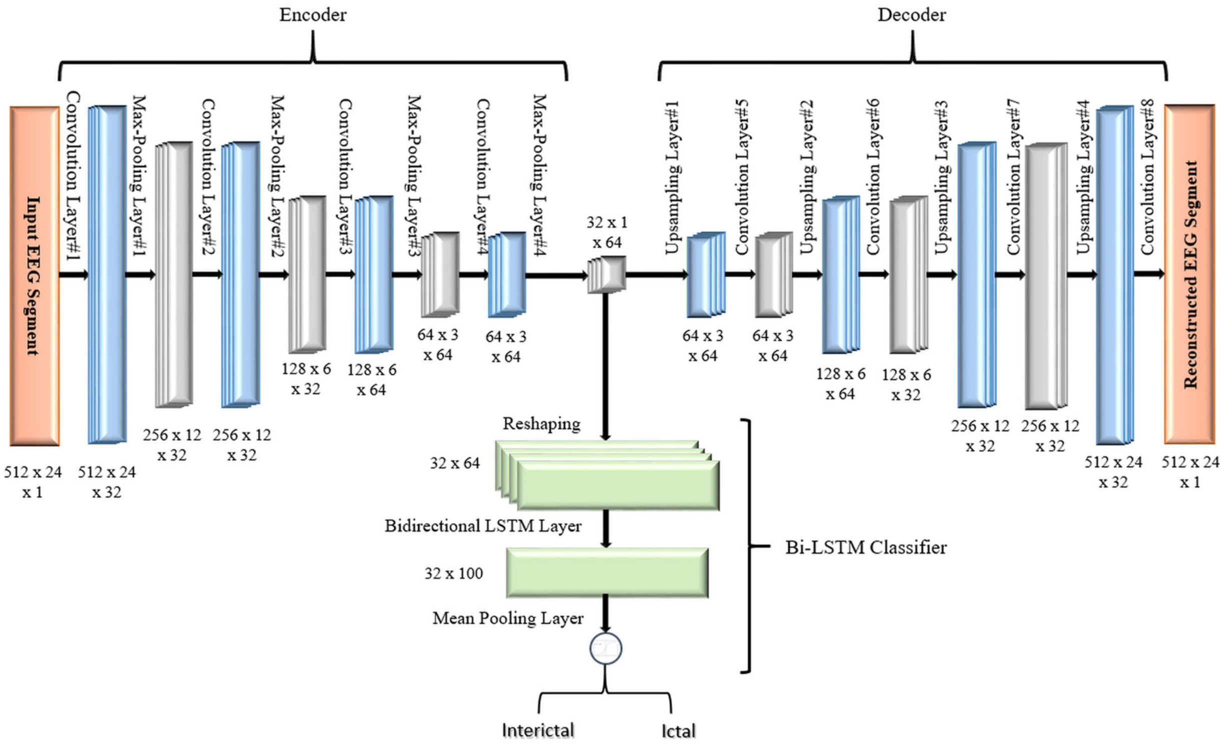

FIGURE 7 | Proposed 2D-DCAE + Bi-LSTM architecture (assuming that an EEG segment length is 2 s).

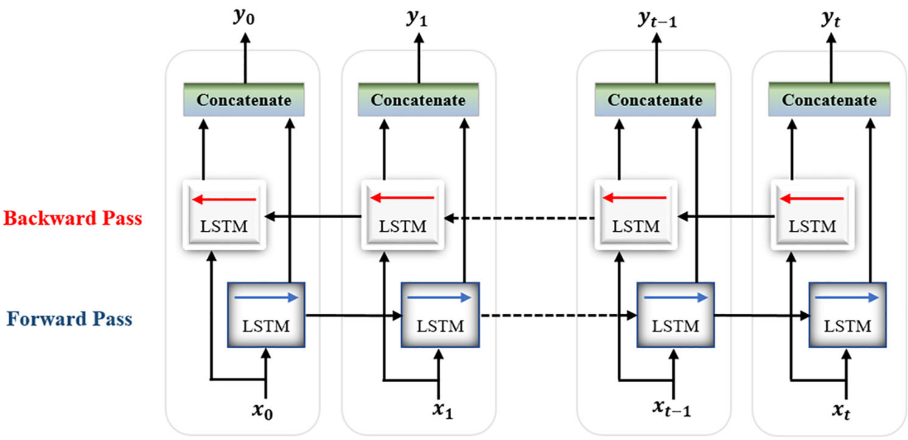

FIGURE 8 | Unrolled single-layer Bi-LSTM network.

the F1-score combines the values of precision and recall in a single metric.

Models Implementation

The Python programming language along with many supporting libraries and in particular, the Tensorflow machine learning library's Keras deep learning API, has been used to develop our models. Due to the variations in the hardware resources and different GPU specifications, we have chosen not to include the computational times of training and testing the proposed models as a metric in our comparisons especially since we are developing our models using external resources provided by Google Colaboratory online environment that runs on Google's cloud servers.

RESULTS

For each of the four models, **Figure 9** shows the ranges of values of the five performance metrics calculated based on the 10-Fold cross-validation classification results of the EEG segments of lengths 1, 2, and 4 s. The mean and standard deviation of all metrics over the 10-folds are then calculated and summarized in **Table 2**.

The same results are interpreted visually in **Figure 10**.

DISCUSSION

As can be seen from the results, for all EEG segment lengths and evaluation metrics, the two proposed SDCAE models (DCAE + MLP and DCAE + Bi-LSTM) have outperformed the other two models (DCNN + MLP and DCNN + Bi-LSTM) that do not use AEs. Furthermore, as highlighted in **Table 2**, using a segment length of 4 s, the DCAE + Bi-LSTM model has achieved the highest performance in terms of all evaluation metrics among all other combinations of models. It is also interesting to see that in all SDCAE models, a 4-s EEG segment length is the best choice to get the best classification performance. Generally, it can be noticed that all models that utilized a Bi-LSTM for classification have accomplished better results compared to their counterpart models that use MLP-based

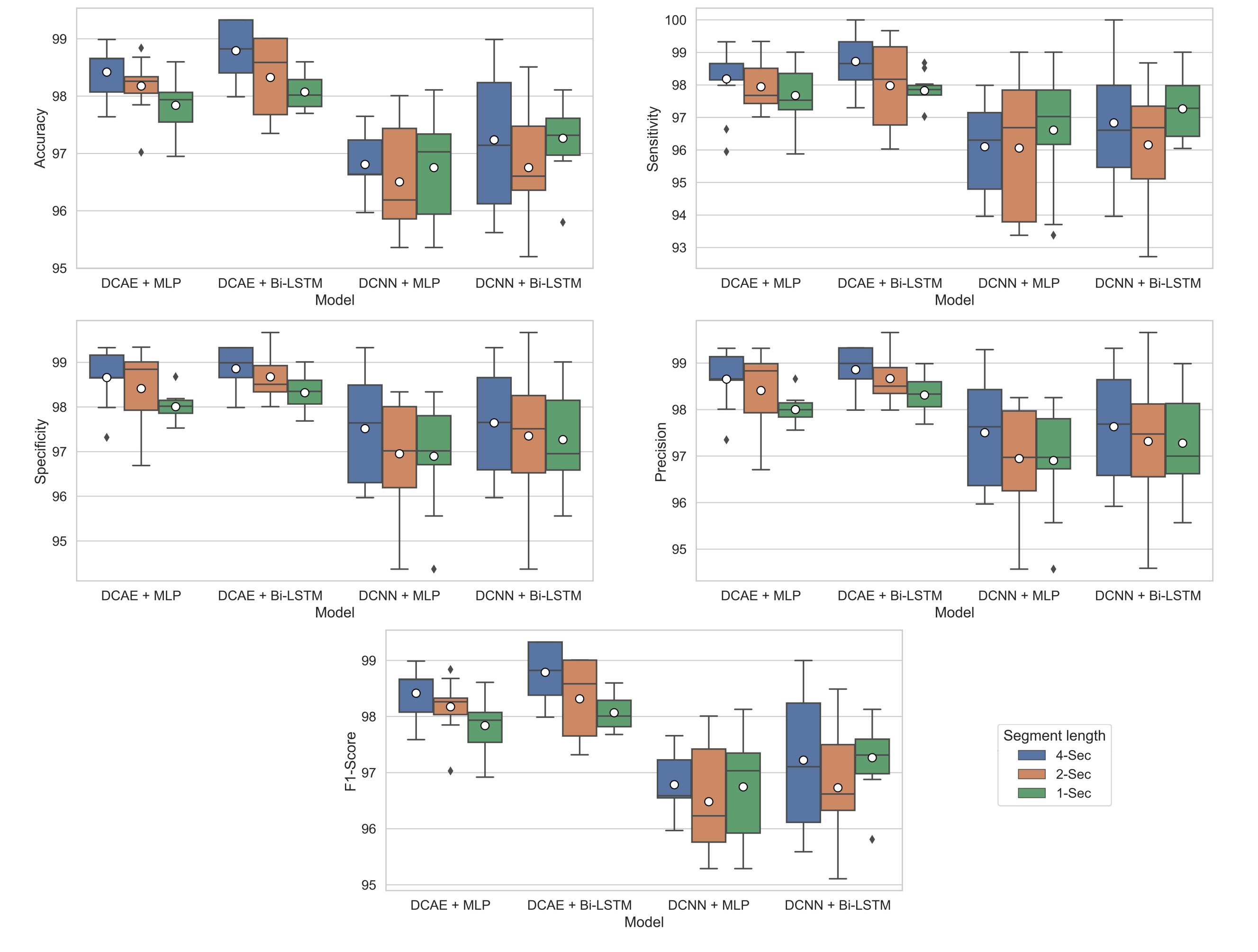

FIGURE 9 | Boxplots showing ranges of performance metrics percentages calculated based on the 10-fold cross-validation results.

TABLE 2 | Classification results using different EEG segment lengths.

EEG segment length	Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-score (%)
1 s	DCAE + MLP	97.84 ± 0.44	97.67 ± 0.88	98.01 ± 0.30	98.00 ± 0.30	97.84 ± 0.46
	DCAE + Bi-LSTM	98.07 ± 0.31	97.83 ± 0.52	98.32 ± 0.43	98.31 ± 0.42	98.07 ± 0.32
	DCNN + MLP	96.75 ± 0.88	96.61 ± 1.79	96.90 ± 1.18	96.91 ± 1.11	96.75 ± 0.90
	DCNN + Bi-LSTM	97.27 ± 0.65	97.27 ± 0.95	97.27 ± 1.09	97.28 ± 1.06	97.27 ± 0.64
2 s	DCAE + MLP	98.18 ± 0.48	97.95 ± 0.71	98.41 ± 0.90	98.41 ± 0.88	98.18 ± 0.48
	DCAE + Bi-LSTM	98.33 ± 0.71	97.98 ± 1.34	98.68 ± 0.50	98.67 ± 0.50	98.32 ± 0.72
	DCNN + MLP	96.51 ± 0.95	96.06 ± 2.12	96.95 ± 1.21	96.95 ± 1.14	96.48 ± 0.98
	DCNN + Bi-LSTM	96.76 ± 1.02	96.16 ± 1.98	97.35 ± 1.54	97.32 ± 1.46	96.73 ± 1.05
4 s	DCAE + MLP	98.42 ± 0.48	98.19 ± 1.02	98.66 ± 0.61	98.65 ± 0.59	98.42 ± 0.50
	DCAE + Bi-LSTM	98.79 ± 0.53	98.72 ± 0.77	98.86 ± 0.53	98.86 ± 0.53	98.79 ± 0.53
	DCNN + MLP	96.81 ± 0.50	96.10 ± 1.43	97.52 ± 1.20	97.50 ± 1.16	96.79 ± 0.51
	DCNN + Bi-LSTM	97.24 ± 1.20	96.84 ± 1.79	97.65 ± 1.28	97.63 ± 1.28	97.22 ± 1.21

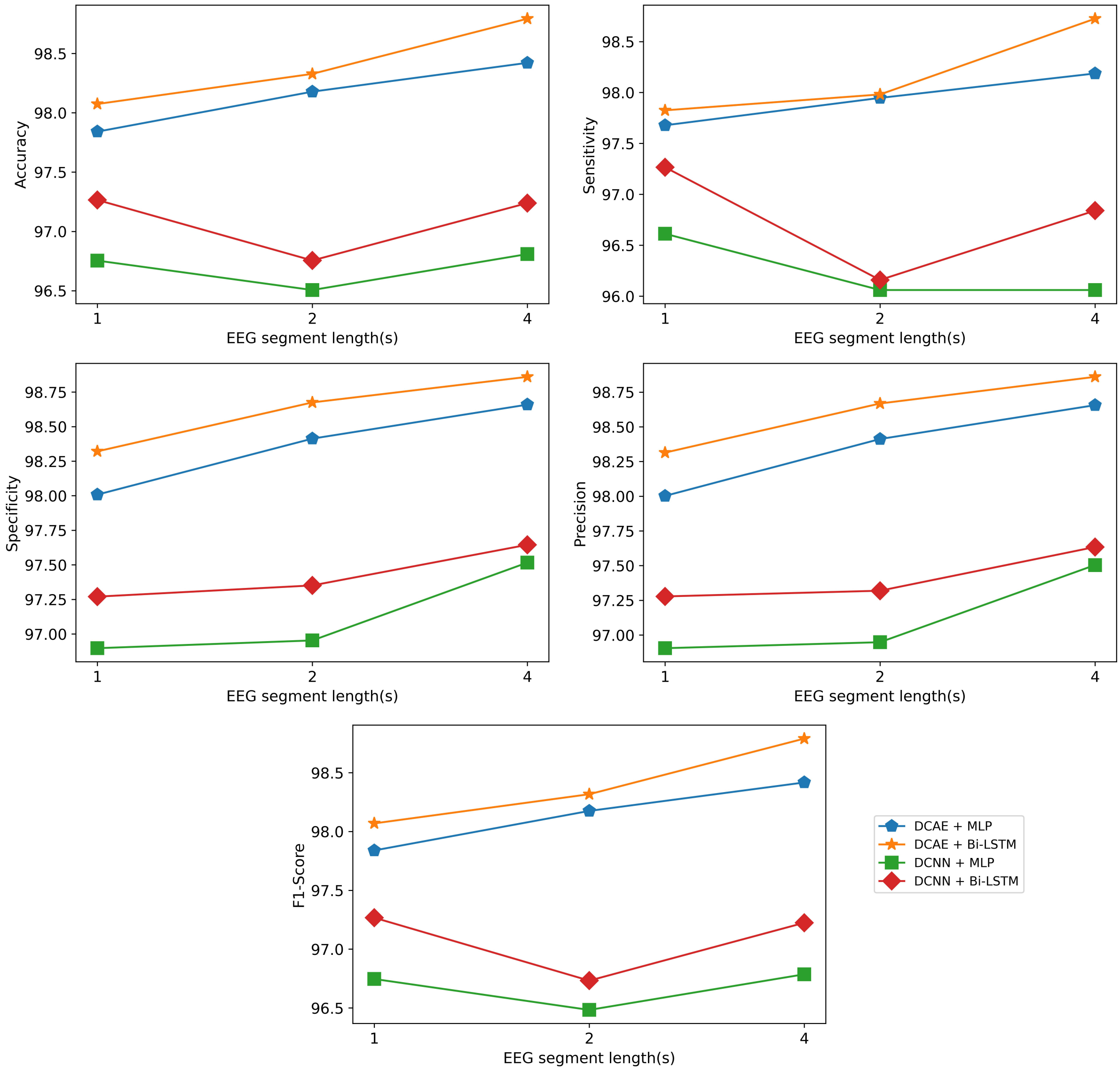

FIGURE 10 | Visualization of the classification results of the models using different EEG segment lengths.

classifiers using the same EEG segment lengths. That can be explained as Bi-LSTM networks are more capable to learn better temporal patterns from the generated latent space sequence better than MLP networks. Finally, by comparing the standard deviations in the evaluation metrics values for all models, it is clear that the results of the SDCAE models mostly have less dispersion compared to the other models, which means that the SDCAE models' performance is more consistent across all cross-validation iterations.

Figure 11 shows the classification accuracy, CL, and RL curves for the training and testing datasets obtained while training the winning model (DCAE + Bi-LSTM) in one of the iterations of the 10-fold cross-validation.

Statistical Analysis

The non-parametric Kruskal–Wallis *H* test (Kruskal and Wallis, 1952) is used to test the statistical significance of the classification results of the two proposed models (DCAE + MLP and DCAE + Bi-LSTM). For simplicity, the test results for comparing the evaluation metrics of the models obtained using an EEG segment

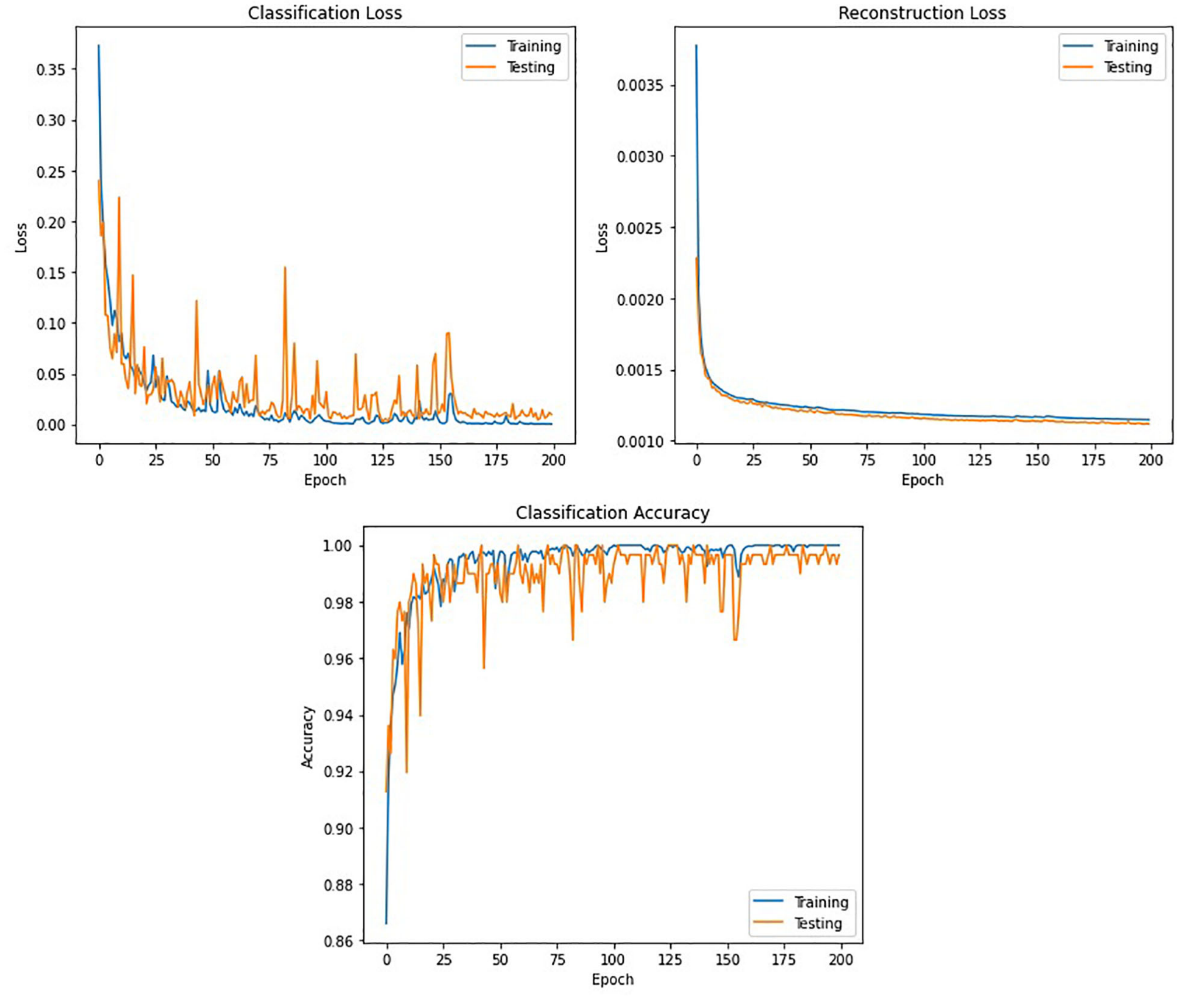

FIGURE 11 | Accuracy and loss curves against the number of epochs obtained while training the DCAE + Bi-LSTM model.

TABLE 3 | Comparison between our best performing model and previous methods using the same dataset.

Methods	Features extraction	Data selection	Accuracy (%)	Sensitivity (%)	Specificity (%)
Yuan et al., 2017	STFT + SSDAE	Random	93.82	N/A	N/A
Ke et al., 2018	MIC+VGGNet	Fivefold CV	98.1	98.85	97.47
Zhou et al., 2018	FFT+CNN	Sixfold CV	97.5	96.86	98.15
Hossain et al., 2019	2D-CNN	Random	98.05	90	91.65
Proposed Work	(DCAE + Bi-LSTM)	10-Fold CV	98.79	98.72	98.86

length of 4 s will be demonstrated. When comparing DCAE + Bi-LSTM with the two models (DCNN + MLP and DCNN + Bi-LSTM), the Kruskal–Wallis *H* test produced *p*-value = 0.0005 for accuracy, *p*-value = 0.02 for sensitivity, *p*-value = 0.025 for specificity, *p*-value = 0.005 for precision, and *p*-value = 0.001 for F1-score. Also, when comparing DCAE + MLP with the same two models, the statistical test showed *p*-value = 0.003 for accuracy, *p*-value = 0.011 for sensitivity, *p*-value = 0.083 for specificity, *p*-value = 0.019 for precision, and *p*-value = 0.002 for F1-score. For all performance assessment metrics, nearly all comparisons yielded a *p*-value lower than 0.05, apart from only one *p*-value for specificity. This shows the disparity in the statistical significance between the outcomes of all the proposed models.

Comparison With Other Methods

In the literature, not all previous work uses the same set of metrics for evaluating the performance of the seizure classification algorithms. So, comparisons based on the most commonly used metrics which are accuracy, sensitivity, and specificity, will only be provided in this section. **Table 3** summarizes the comparison between our best performing model and some state-of-the-art methods that use deep neural networks for feature extraction and classification of seizures. In Yuan et al. (2017), various stacked sparse denoising autoencoders (SSDAE) have been tested and compared for feature extraction and classification after preprocessing using short-time Fourier transform (STFT). The best accuracy they obtained was 93.82% using a random selection of training and testing datasets. Ke et al. (2018) combined global maximal information coefficient (MIC) with visual geometry group network (VGGNet) for feature extraction and classification. Using fivefold cross-validation, they achieved 98.1% accuracy, 98.85% sensitivity, and 97.47% specificity. Using fast Fourier transform (FFT) for frequency domain analysis and CNN, the authors (Zhou et al., 2018) performed patient-specific classifications between the ictal and interictal signals. Relying on sixfold cross-validation, the average of the evaluation metrics for all patients was 97.5% accuracy, 96.86% sensitivity, and 98.15% specificity. Finally, in Hossain et al. (2019), the authors used a 2D-CNN model to extract spectral and temporal characteristics of EEG signals and used them for patient-specific classification using a random selection of training and testing datasets. They got 98.05% accuracy, 90% sensitivity, and 91.65% specificity for the cross-patient results. Following the previous comparison, the results obtained by our model have shown to be superior to some of the state-of-the-art systems which all lack the proper statistical analysis for significance testing.

CONCLUSION

A novel deep-learning approach for the detection of seizures in pediatric patients is proposed. The novel approach uses a 2D-SDCAE for the detection of epileptic seizures based on classifying minimally pre-processed raw multichannel EEG signal recordings. In this approach, an AE is trained in a supervised way to classify between the ictal and interictal brain state EEG signals to exploit its capabilities of performing both automatic feature learning and classification simultaneously with high efficiency. Two SDCAE models that use Bi-LSTM and MLP networks-based classifiers were designed and tested using three EEG data segment lengths. The performance of both proposed models is compared to two regular deep learning models having the same layers' structure, except that the decoder network layers are completely removed. The twelve models are trained and assessed using a 10-fold cross-validation scheme and based on five evaluation metrics, the best performing model was the SDCAE model that uses a Bi-LSTM and 4 s EEG segments. This model has achieved an average of 98.79% accuracy, 98.72% sensitivity, 98.86% specificity, 98.86% precision, and finally an F1-score of 98.79%. The comparison between this SDCAE model and other state-of-the-art systems using the same dataset has shown that the performance of our proposed model is superior to that of most existing systems.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://physionet.org/content/chbmit/1.0.0/.

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 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

AA conceived the presented idea, conducted the analysis, and produced the figures. MB supervised the findings of this work. Both authors discussed the results and contributed to the final manuscript.

REFERENCES

Abdelhameed, A. M., and Bayoumi, M. (2018). "Semi-supervised deep learning system for epileptic seizures onset prediction," in *Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL. doi: 10.1109/icmla.2018.00191

Abdelhameed, A. M., and Bayoumi, M. (2019). Semi-supervised EEG signals classification system for epileptic seizure detection. *IEEE Signal Process. Lett.* 26, 1922–1926. doi: 10.1109/lsp.2019.2953870

Abdelhameed, A. M., Daoud, H. G., and Bayoumi, M. (2018a). "Deep convolutional bidirectional LSTM recurrent neural network for epileptic seizure detection," in *Proceedings of the 2018 16th IEEE International New Circuits and Systems Conference (NEWCAS)*, Montreal, QC. doi: 10.1109/newcas.2018.8585542

Abdelhameed, A. M., Daoud, H. G., and Bayoumi, M. (2018b). "Epileptic seizure detection using deep convolutional autoencoder," in *Proceedings of the 2018 IEEE International Workshop on Signal Processing Systems (SiPS)*, Cape Town. doi: 10.1109/sips.2018.8598447

Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., and Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Comput. Biol. Med.* 100, 270–278. doi: 10.1016/j.compbiomed.2017.09.017

Acharya, U. R., Sree, S. V., Swapna, G., Martis, R. J., and Suri, J. S. (2013). Automated EEG analysis of epilepsy: a review. *Knowled. Based Syst.* 45, 147–165. doi: 10.1016/j.knosys.2013.02.014

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transact. Neur. Netw.* 5, 157–166. doi: 10.1109/72.279181

Bottou, L. (2004). *Stochastic Learning. Advanced Lectures on Machine Learning, LNAI*, Vol. 3176. Berlin: Springer, 146–168.

Chen, G., Xie, W., Bui, T. D., and Krzyżak, A. (2017). Automatic epileptic seizure detection in EEG using nonsubsampled wavelet–fourier features. *J. Med. Biol. Eng.* 37, 123–131. doi: 10.1007/s40846-016-0214-0

Epilepsy in Children (2020). *Diagnosis & Treatment HealthyChildren.org*. Available online at: https://www.healthychildren.org/English/health-issues/conditions/seizures/Pages/Epilepsy-in-Children-Diagnosis-and-Treatment.aspx (accessed December 15, 2020).

Gao, Y., Gao, B., Chen, Q., Liu, J., and Zhang, Y. (2020). Deep convolutional neural network-based epileptic electroencephalogram (EEG) signal classification. *Front. Neurol.* 11:375. doi: 10.3389/fneur.2020.00375

Gogna, A., Majumdar, A., and Ward, R. (2017). Semi-supervised stacked label consistent autoencoder for reconstruction and analysis of biomedical signals. *IEEE Transact. Biomed. Eng.* 64, 2196–2205. doi: 10.1109/tbme.2016.2631620

Goodfellow, I., Bengio, Y., and Courville, A. (2017). *Deep Learning*. Cambridge, MA: The MIT Press.

He, H., and Ma, Y. (2013). *Imbalanced Learning Foundations, Algorithms, and Applications*. Hoboken, NJ: IEEE Press.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neur. Comput.* 9, 1735–1780.

Hossain, M. S., Amin, S. U., Alsulaiman, M., and Muhammad, G. (2019). Applying deep learning for epilepsy seizure detection and brain mapping visualization. *ACM Transact. Multimed. Comput. Commun. Appl.* 15, 1–17. doi: 10.1145/3241056

Hu, W., Cao, J., Lai, X., and Liu, J. (2019). Mean amplitude spectrum based epileptic state classification for seizure prediction using convolutional neural networks. *J. Ambient Intell. Human. Comput.* doi: 10.1007/s12652-019-01220-6

Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv* [Preprint]. arXiv:1502.03167.

Jaiswal, A. K., and Banka, H. (2017). Local pattern transformation based feature extraction techniques for classification of epileptic EEG signals. *Biomed. Signal Process. Control* 34, 81–92. doi: 10.1016/j.bspc.2017.01.005

Ke, H., Chen, D., Li, X., Tang, Y., Shah, T., and Ranjan, R. (2018). Towards brain big data classification: epileptic EEG identification with a lightweight VGGNet on global MIC. *IEEE Access* 6, 14722–14733. doi: 10.1109/access.2018.2810882

Kingma, D., and Ba, J. (2014). Adam: a method for stochastic optimization. *ArXiv* [Preprint]. arXiv:1412.6980.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi: 10.1145/3065386

Kruskal, W. H., and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *J. Am. Statist. Associat.* 47, 583–621. doi: 10.1080/01621459.1952.10483441

Mozer, M. C. (1989). A focused backpropagation algorithm for temporal pattern recognition. *Comp. Syst.* 3, 349–381.

Orhan, U., Hekim, M., and Ozer, M. (2011). EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Exp. Syst. Appl.* 38, 13475–13481. doi: 10.1016/j.eswa.2011.04.149

Raghu, S., Sriraam, N., Hegde, A. S., and Kubben, P. L. (2019). A novel approach for classification of epileptic seizures using matrix determinant. *Exp. Syst. Appl.* 127, 323–341. doi: 10.1016/j.eswa.2019.03.021

Samiee, K., Kovacs, P., and Gabbouj, M. (2015). Epileptic seizure classification of EEG time-series using rational discrete short-time fourier transform. *IEEE Transact. Biomed. Eng.* 62, 541–552. doi: 10.1109/tbme.2014.2360101

Schomer, D. L., and Lopez da Silva, H. F. (2018). *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. New York, NY: Oxford University Press.

She, Q., Hu, Bo, Luo, Z., Nguyen, T., and Zhang, L. (2018). A hierarchical semi-supervised extreme learning machine method for EEG recognition. *Med. Biol. Eng. Comput.* 57, 147–157. doi: 10.1007/s11517-018-1875-3

Shoeb, A. H. (2009). *Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Sokolova, M., and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Inform. Process. Manag.* 45, 427–437. doi: 10.1016/j.ipm.2009.03.002

Song, Y., Crowcroft, J., and Zhang, J. (2012). Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine. *J. Neurosci. Methods* 210, 132–146. doi: 10.1016/j.jneumeth.2012.07.003

Tieleman, T., and Hinton, G. (2012). Lecture 6.5-RMSprop: divide the gradient by a running average of its recent magnitude. *COURSERA Neur. Netw. Mach. Learn.* 4, 26–31.

Wang, D., Ren, D., Li, K., Feng, Y., Ma, D., Yan, X., et al. (2018). Epileptic seizure detection in long-term EEG recordings by using wavelet-based directed transfer function. *IEEE Transact. Biomed. Eng.* 65, 2591–2599. doi: 10.1109/tbme.2018.2809798

Wang, X., Zhao, Y., and Pourpanah, F. (2020). Recent advances in deep learning. *Int. J. Mach. Learn. Cybern.*11, 747–750. doi: 10.1007/s13042-020-01096-5

Wei, X., Zhou, L., Zhang, Z., Chen, Z., and Zhou, Y. (2019). Early prediction of epileptic seizures using a long-term recurrent convolutional network. *J. Neurosci. Methods* 327:108395. doi: 10.1016/j.jneumeth.2019.108395

World Health Organization (2020). *Epilepsy*. Available online at: https://www.who.int/en/news-room/fact-sheets/detail/epilepsy (accessed December 5, 2020).

Yavuz, E., Kasapbaşı, M. C., Eyüpoğlu, C., and Yazıcı, R. (2018). An epileptic seizure detection system based on cepstral analysis and generalized regression neural network. *Biocybern. Biomed. Eng.* 38, 201–216. doi: 10.1016/j.bbe.2018.01.002

Yuan, Y., Xun, G., Jia, K., and Zhang, A. (2017). "A multi-view deep learning method for epileptic seizure detection using short-time fourier transform," in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, New York, NY. doi: 10.1145/3107411.3107419

Zeiler, M. D. (2012). ADADELTA: an adaptive learning rate method. *arXiv* [Preprint]. arXiv:1212.5701.

Zhou, M., Tian, C., Cao, R., Wang, B., Niu, Y., Hu, T., et al. (2018). Epileptic Seizure detection based on EEG signals and CNN. *Fronti. Neuroinform.* 12:95. doi: 10.3389/fninf.2018.00095

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.

frontiers
in Computational Neuroscience

ORIGINAL RESEARCH
published: 08 April 2021
doi: 10.3389/fncom.2021.654315

ASD-SAENet: A Sparse Autoencoder, and Deep-Neural Network Model for Detecting Autism Spectrum Disorder (ASD) Using fMRI Data

*Fahad Almuqhim and Fahad Saeed**

Knight Foundation School of Computing and Information Sciences, Florida International University, Miami, FL, United States

Autism spectrum disorder (ASD) is a heterogenous neurodevelopmental disorder which is characterized by impaired communication, and limited social interactions. The shortcomings of current clinical approaches which are based exclusively on behavioral observation of symptomology, and poor understanding of the neurological mechanisms underlying ASD necessitates the identification of new biomarkers that can aid in study of brain development, and functioning, and can lead to accurate and early detection of ASD. In this paper, we developed a deep-learning model called *ASD-SAENet* for classifying patients with ASD from typical control subjects using fMRI data. We designed and implemented a sparse autoencoder (SAE) which results in optimized extraction of features that can be used for classification. These features are then fed into a deep neural network (DNN) which results in superior classification of fMRI brain scans more prone to ASD. Our proposed model is trained to optimize the classifier while improving extracted features based on both reconstructed data error and the classifier error. We evaluated our proposed deep-learning model using publicly available Autism Brain Imaging Data Exchange (ABIDE) dataset collected from 17 different research centers, and include more than 1,035 subjects. Our extensive experimentation demonstrate that *ASD-SAENet* exhibits comparable accuracy (70.8%), and superior specificity (79.1%) for the whole dataset as compared to other methods. Further, our experiments demonstrate superior results as compared to other state-of-the-art methods on 12 out of the 17 imaging centers exhibiting superior generalizability across different data acquisition sites and protocols. The implemented code is available on GitHub portal of our lab at: https://github.com/pcdslab/ASD-SAENet.

Keywords: ASD, fMRI, autoencoder, sparse autoencoder, ABIDE, deep-learning, classification, diagnosis

OPEN ACCESS

Edited by:
Fahimeh Mamashli,
Massachusetts General Hospital and
Harvard Medical School,
United States

Reviewed by:
Jussi Alho,
Massachusetts General Hospital and
Harvard Medical School,
United States
Nataliia Kozhemiako,
Brigham and Women's Hospital and
Harvard Medical School,
United States

****Correspondence:***
Fahad Saeed
fsaeed@fiu.edu

Received: *15 January 2021*
Accepted: *08 March 2021*
Published: *08 April 2021*

Citation:
Almuqhim F and Saeed F (2021)
ASD-SAENet: A Sparse Autoencoder,
and Deep-Neural Network Model for
Detecting Autism Spectrum Disorder
(ASD) Using fMRI Data.
Front. Comput. Neurosci. 15:654315.
doi: 10.3389/fncom.2021.654315

1. INTRODUCTION

More than 1.5 Million children (Baio et al., 2018) in the US are affected by heterogenous Autism Spectrum Disorder (ASD) which has wide range of symptoms or characteristics such as limited communication (including verbal and non-verbal), limited social interaction, and may exhibit repeated or limited interests or activities (American Psychiatric Association, 2013). Individuals with ASD have numerous challenges in daily life, and often develop comorbidities such as depression, anxiety disorder, or ADHD which may further complicate the diagnostic processes especially for young children (Mizuno et al., 2019). Although some symptoms are generally recognizable between 1 and 2 years of age; numerous children are not formally diagnosed with ASD diagnosis until they are much older (Stevens et al., 2016).

To date, the diagnostic process for individuals with ASD is based purely on behavioral descriptions of symptomology (DSM-5/ICD-10) (Nickel and Huang-Storms, 2017) from informants observing children with the disorder across different settings (e.g., home, school). Early cognitive, language, and social interventions for children (under 24 months old) with ASD has shown to be especially effective (Bradshaw et al., 2015), and a delayed diagnosis can have more disastrous effects in the life of the child. Help in terms of assisted learning or speech therapies is often available to these children (especially in low-income demographics), only *after* a diagnosis has been administered (Boat and Wu, 2015), making an early diagnosis even more urgent. ASD is associated with altered brain development in the early childhood but there are no reliable biomarkers that can be used for diagnosis (Lord et al., 2018). Collectively, our evolving understanding of the shared and distinct behavioral features that characterize ASD highlights the need for further inquiry into mechanisms behind ASD brain development, and functioning. The shortcomings of current clinical approaches (National Collaborating Centre for Mental Health, 2009), and the poor understanding of the neurological mechanisms underlying ASD necessitates the identification of new biomarkers and computational techniques that can aid clinicians, and neuroscientists alike to understand the distinct way ASD brain works as compared to a typical brain.

In the recent decade, advances in neuroimaging technologies are providing a critical step, and has made it possible to measure the functional and structural changes associated with ASD (Just et al., 2007). Functional magnetic resonance imaging (fMRI) is commonly used to detect biomarker patterns for brain disorders (Just et al., 2007; Dichter, 2012; Botvinik-Nezer et al., 2020), and has gained extensive attention for ASD biomarker discovery, and classification (Iidaka, 2015; Plitt et al., 2015; Li et al., 2018; Xiao et al., 2018; El-Gazzar et al., 2019b; Wang et al., 2019). The fMRI data is shown to provide significant insights, and can demonstrate both the hypo- and hyper-connectivity in the ASD brain development (Di Martino et al., 2014; Lau et al., 2019), and can be used to study different origination theories related to ASD (Just et al., 2007). In fMRI studies, functional connectivity is based on the correlation of the activation time series in pairs of brain areas and are studied for both ASD and healthy brains. However, it is not possible to detect subtle biomarker patterns using conventional computational and statistical methods (Eslami and Saeed, 2019; Haweel et al., 2020; Nogay and Adeli, 2020). Machine learning algorithms have been successful in identifying biomarkers from functional Magnetic Resonance Imaging (fMRI) datasets for biomarker discovery, and classification of various brain disorders (Deshpande et al., 2015; Sarraf and Tofighi, 2016; Dvornek et al., 2017; Eslami and Saeed, 2018, 2019; El-Gazzar et al., 2019a; Yao and Lu, 2019). Effective modeling of ASD brain connectivity using fMRI data may lead to biomarker detection, and consequently better understanding of the brain neural activity associated with ASD.

In this study, we focus on designing and developing a machine-learning model that can distinguish and classify fMRI data from ASD subjects, and from typical control (TC). We focus on designing a deep learning algorithm that can extract, and distinguish between the functional features associated with ASD fMRI brain scans as compared to healthy typical controls. To this end, we have designed and implemented a deep-learning model, called *ASD-SAENet*, consisting of a sparse autoencoder (SAE) which lowers the dimensionality of our input features. The sparsity of SAE helps in extracting the features from high-dimensional imaging data while ensuring that limited sample size does not lead to overfitting (Ng, 2011). Extraction of feature(s) is then followed by a deep-neural network with 2-hidden layers, and softmax layer at the output. Our extensive experimentation using ABIDE-I datasets show that ASD-SAENet achieves an average accuracy of 70.8% improving upon our earlier state-of-the-art work- (Eslami et al., 2019), as well as other methods (Heinsfeld et al., 2018). In this study, we further demonstrate that ASD-SAENet model outperforms other methods (Heinsfeld et al., 2018; Eslami et al., 2019) in more than 12 of the 17 individual imaging sites showing superior generalizability across different data acquisition sites, protocols, and processes.

The structure of the paper is as follows: In section 2, we discuss the related work, and the associated state-of-the-art tools. In section 3, we propose the design of the machine-learning algorithm, the feature extraction, and the classification processes for the proposed ASD-SAENet method. Experimental results, datasets, and comparison against state-of-the-art methods are discussed in section 4. Section 5 illustrates the discussion, conclusions, and future work.

2. RELATED WORK

Detecting, and finding biomarkers from imaging datasets such as fMRI has attracted significant attention in recent years. One of the key reasons of this increased attention, apart from the significance of finding quantitative biomarkers for ASD that can lead to new neuroscientific knowledge discovery, is the availability of publicly accessible Autism Brain Imaging Data Exchange (ABIDE) datasets collected from 17 different sites (Craddock et al., 2013) resulting in numerous studies (Abraham et al., 2017; Fredo et al., 2018; Khosla et al., 2018; El-Gazzar et al., 2019a; Parikh et al., 2019; Sherkatghanad et al., 2019).

Several studies used a *subset* of ABIDE dataset include El-Gazzar et al. (2019b) developed a 3D convolutional neural network and a 3D convolutional LSTM to classify ASD subjects from heathy subjects using fMRI data. The study was evaluated on a subset of ABIDE dataset containing 184 subjects collected from NYU and UM sites achieving 77% accuracy. In another work, Guo et al. (2017) proposed a deep neural network that has several stacked sparse autoencoders to lower dimensional features and a deep neural network (DNN) model to classify ASD patients from TC patients. The model was trained and tested on the dataset of UM site from ABIDE dataset, achieving 86.36% accuracy. Using the whole dataset from ABIDE, Brown et al. (2018) developed a model based on BrainNetCNN with an element-wise layer attached as the first step, and a data-driven structural priors. This model was evaluated on 1,013 subjects where 539 were heathy and 474 were ASD subjects and achieved an accuracy of 68.7%. Machine learning methods such as Support Vector Machine (SVM), and Random Forests have also been used to classify ASD subjects. Kazeminejad and Sotero (2019) obtained an accuracy of 95% by developing a feature selection pipeline based on graph theoretical metrics, and SVM method for classification. However, these results (Kazeminejad and Sotero, 2019) were based on a *subset* of ABIDE data set with subjects older than 30 years, and its generalizability is unknown for subjects that are children.

Deep learning techniques such as Deep learning network (DNN), Autoencoders, and Convolutional Neural Network (CNN) have gained an extensive attention for ASD biomarker detection, and classification studies (Khosla et al., 2018; Li et al., 2018; Wang et al., 2019; Yao and Lu, 2019). Heinsfeld et al. (2018) used a deep learning method which consists of two stacked denoising autoencoders, and a multi-layer classifier which achieved an average of 70% accuracy using the full ABIDE dataset. They also executed the model for each site, and they achieved an average accuracy of 52%. Recently, we (Eslami et al., 2019) proposed a deep-learning model called *ASD-DiagNet* which is the current state-of-the-art method in the field used by multiple studies (Mostafa et al., 2019a,b; Bilgen et al., 2020; Niu et al., 2020). ASD-DiagNet consists of an autoencoder for lowering the features dimensionality, and a single layer perceptron for classification decision. The method was trained with an expanded training data using data augmentation technique which then achieved 70.3% accuracy using the complete ABIDE dataset. In Eslami et al. (2019), accuracy for each of the 17 sites from ABIDE dataset resulted in average accuracy of 63.2%. However, we also reported that *ASD-DiagNet* exhibited 82% as the maximum accuracy for some of the sites which was 28% higher than any other method. However, clearly higher accuracy depicted by *ASD-DiagNet* still requires that the accuracy is generalizable across different sites with different MRI machines, and various data acquisition protocols, and pre-processing workflows. In this paper, we demonstrate that the proposed ASD-SAENet model exhibits comparable average accuracy relative to other methods but results in higher accuracy for 12 out of the 17 data acquisition centers. We will use these two studies (Heinsfeld et al., 2018; Eslami et al., 2019) to compare, and evaluate our proposed model.

3. MATERIALS AND METHODS

3.1. Functional Magnetic Resonance Imaging and ABIDE Dataset

Functional Magnetic Resonance Imaging (fMRI) is a non-invasive brain imaging technique that allows capturing brain activity over time. The data of fMRI is represented by measuring the blood-oxygen-level-dependent (BOLD) volume of each small cubic called voxel at a given time point. Therefore, the data consists of a time series of each voxel representing its activity over time. For brain disorders, resting state fMRI (rs-fMRI) is commonly used which is scanning the brain image while the subject is resting. In this paper, we used the ABIDE-I dataset that is provided by the ABIDE initiative. This dataset consists 1,035 rs-fMRI data with 505 ASD subject, and 530 healthy control subjects that are collected from 17 different sites. The dataset was preprocessed and downloaded from (http://preprocessed-connectomes-project.org/abide/). We used the preprocessed data using the Configurable Pipeline for the Analysis of Connectomes C-PAC pipeline (Craddock et al., 2013) which is parcellated into 200 region of interests (ROIs) using Craddock 200 (CC200) functional parcellation (Craddock et al., 2012). For each region, the average voxels' BOLDs is calculated. The preprocessing steps also include skull-striping, slice time correction, motion correction, and nuisance signal regression. Each site used different parameters, and scanners for brain imaging, such as repetition time (TR), echo time (TE), and flip angle degree. **Table 1** shows the parameters of each site.

3.2. Feature Extraction

Craddock 200 (CC200) (Craddock et al., 2012) atlas divides the brain into 200 regions. Time series of each regions was extracted. Pearson's correlation coefficient is used to calculate the functional correlations of the ROIs. The following equation was used to obtain the correlation between two different time series data of each region i, and j of length T.

$$p_{i,j} = \frac{\sum_{t=1}^{T}(i_t - \bar{i})(j_t - \bar{j})}{\sqrt{\sum_{t=1}^{T}(i_t - \bar{i})^2}\sqrt{\sum_{t=1}^{T}(j_t - \bar{j})^2}} \tag{1}$$

TABLE 1 | The scanning parameters of ABIDE dataset for each site show the different in MRI Scanner, TR (Repetition Time), TE (Echo Time), Flip Angle, and Age which may result in difference in data acquisition as well as the pre- and post-processing of fMRI data.

Site	MRI scanner	TR (ms)	TE (ms)	Flip angle (degree)	Age (year)
Caltech	SIEMENS	2,000	30	75	17–56.2
CMU	SIEMENS	2,000	30	73	19–40
KKI	PHILLIPS	2,500	30	75	8–12.8
Leuven	PHILLIPS	1,656	33	90	12.1–32
MaxMun	SIEMENS	3,000	30	80	7–58
NYU	SIEMENS	2,000	15	90	6.5–39.1
OHSU	SIEMENS	2,500	30	90	8–15.2
OLIN	SIEMENS	1,500	27	60	10–24
PITT	SIEMENS	1,500	25	70	9.3–35.2
SBL	PHILLIPS	2,200	30	80	20–64
SDSU	GE	2,000	30	90	8.7–17.2
Stanford	GE	2,000	30	80	7.5–12.9
Trinity	PHILLIPS	2,000	28	90	12–25.9
UCLA	SIEMENS	3,000	28	90	8.4–17.9
UM	GE	2,000	30	90	8.2–28.8
USM	SIEMENS	2,000	28	90	8.8–50.2
Yale	SIEMENS	2,000	25	60	7–17.8

These differences may lead the machine-learning models learn site-specific variations leading many machine-learning models give better average accuracy (for whole ABIDE data set) than the site-specific accuracy.

where $\bar{i}$, and $\bar{j}$ are the mean of the time series i and j respectively. A matrix C_{n*n} is obtained after computing all pair-wise correlations. Since we used CC200 atlas which divides the brain into n = 200 regions, it generates a matrix of 200 × 200. Due to the symmetry of the matrix with regard to the diagonal, we only consider the right upper triangle of the matrix, and flatten it to one-dimensional vector as features. These pairs result in $(n)(n-1)/2 = 19,900$ values for each vector. In order to reduce the dimensionality of the input, we adopted the same technique as in Eslami et al. (2019) and only considered 1/4 largest and 1/4 smallest of the average correlations, resulting in a vector of 9,950 values as the input for each subject.

3.3. Model Architecture: Feature Selection and Classification

In order to reduce the dimensionality of the input, we developed an autoencoder model. Autoencoder (AE) neural network is form of unsupervised learning that uses a feed-forward neural network with encoding, and decoding architecture. It is trained to get an input x and then reconstruct x' to be as similar to the input x as possible. There are several types of autoencoders, such as sparse autoencoder (Ng, 2011), a stacked autoencoder (Vincent et al., 2010), and denoising autoencoder (Vincent et al., 2008). Autoencoders can fail to reconstruct the raw data since it might fall into copying task specially when there is a large data space. The lower-out put dimensions of a sparse autoencoder can force the autoencoder to reconstruct the raw data from useful features instead of copying it (Goodfellow et al., 2016b). For this study, we choose a sparse autoencoder which will be used to extract useful patterns with lower dimensionality. These feature vectors are then fed to our deep neural network model which consists of two hidden layers, and a softmax output layer.

The overview of our model is shown in **Figure 1**. The bottleneck of the sparse autoencoder is used as input vector to the deep neural network. In the figure, neurons labeled as (+1) are the bias units added to the feed-forward neural network through the cost function. This step will force the AE to better reconstruct the input x without falling into overfitting. Our proposed sparse

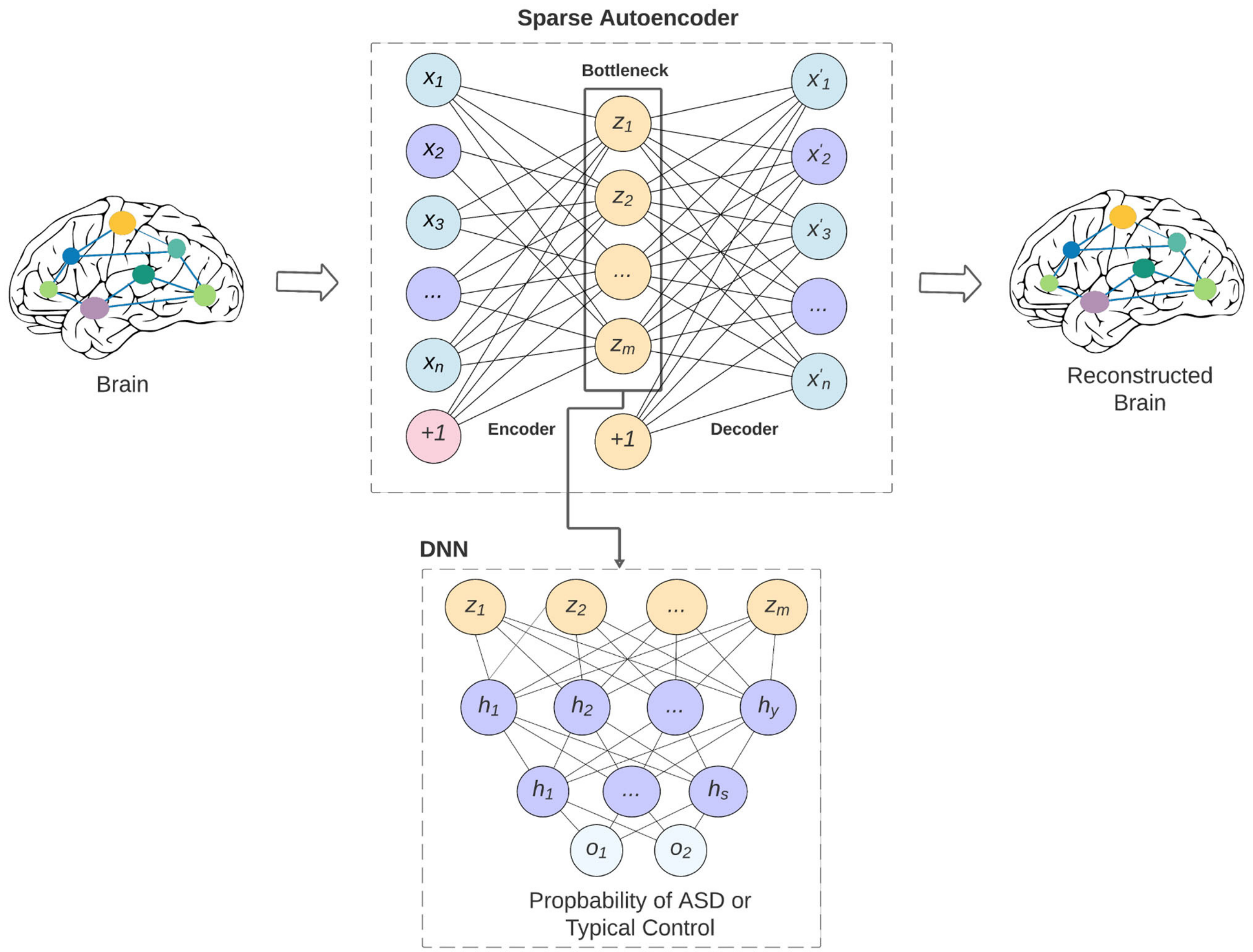

FIGURE 1 | An overview of our proposed model and how the sparse autoencoder is used as feature selection to the deep neural network. The limitation of the feature lead to better generalizability of the model across various data-acquisition sites, and may lead to better interpretability of the models.

autoencoder's (SAE) cost function consists of three parts that are discussed below.

Given a dataset of N training samples $(x_1, x_2, ...x_n)$, where x_i represents the i^{th} input. The developed SAE is trained to reconstruct the input x_i with the function $h_{W,b}(x_i)$ to be as close to x_i as possible. The three parts of the cost function are mean squared error, weight decay, and sparsity term. The first two parts of the cost function, the mean squared error of all N training samples, and the weight decay can be defined as follows:

$$J_{sparse}(W, b) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\|h_{W,b}(x^i) - x^i\| \tag{2}$$

$$+\frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(W_{ji}^l)^2 \tag{3}$$

The Equation (3) defines the weight decay, which helps to avoid overfitting. A small value of λ may lead to overfitting, while a large value of λ may lead to underfitting. Thus, we performed several empirical experiments to select lambda to achieve the best fit of this term.

The third part of the cost function is the sparsity term, which is used to apply activations to the hidden layer of the autoencoder model to prevent overfitting. It can limit the number of regions that are considered in the hidden layer. The following equation defined the average activated value of the hidden layer were a denotes to the activation function which is rectifier (ReLU):

$$\hat{p}_j = \frac{1}{N}\sum_{i=1}^{N}(a_j^2(x^i) \tag{4}$$

Now, the sparsity term is calculated to make $\hat{p}_j$ as close to p as possible, where p is the sparsity parameter. The benefit of this parameter is to deviate $\hat{p}_j$ from p which will result to activate and deactivate neurons on the hidden layer. This term is defined using Kullback-Leibler divergence as follows:

$$\sum_{j=1}^{s_l}KL(p\|\hat{p}_j) = \sum_{j=1}^{s_l}\left[plog\frac{p}{\hat{p}_j} + (1-p)log\frac{1-p}{1-\hat{p}_j}\right] \tag{5}$$

Finally, the cost function of our SAE model after adding all the three parts is defined as follows:

$$J_{sparse}(W, b) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\|h_{W,b}(x^i) - x^i\| + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(W_{ji}^l)^2 + \beta\sum_{j=1}^{s_l}KL(p\|\hat{p}_j) \tag{6}$$

where β is the sparse penalty term.

The SAE is used to reduce the dimensional representation of the input where the size of the input is 9,500 features. The bottleneck of the SAE provides useful features that can be used as inputs for our deep neural network classifier. The size of the bottleneck is 4,975 hidden units. The classifier consists of two hidden layers, and an output layer where the units sizes are 4,975, 2,487, 500, and 2, respectively. The output layer is a softmax regression (Goodfellow et al., 2016a) which represents the probability of each class. To avoid overfitting, we used dropout between the fully connected neural networks. Then we take the maximum probability between the two classes as the final decision of the classifier. We used Cross Entropy for calculating the cost function of the classifier, and added a weight decay term.

The SAE is trained to minimize its cost function described above, and the deep neural network classifier is trained by taking the bottleneck of the SAE as inputs. The SAE and the classifier were trained simultaneously which results in feature extraction which improves while optimizing the classifier's decision. The training process is completed in 30 iterations and the batch size is 8. The sparsity parameter p, the weight decay λ, and the sparse penalty term β were chosen to be 0.05, 0.0001, and 2, respectively. We fine-tuned the deep neural network classifier on the last 10 iterations to adjust the parameters of the classifier, and minimize the cost function of the softmax while the parameters of SAE are frozen. Adam optimizer (Kingma and Ba, 2014) is used to update the parameters based on the computed gradients. Our ASD-SAENet model can be seen in **Figure 2**. All the experiments reported in this paper were performed using a Linux server with Ubuntu OS. The server has a processor of Intel Xeon E5-2690 v3 at 2.60 GHz. The total RAM is 54 GBs. The server also contains an NVIDIA Tesla K80 running CUDA version 10.2 and PyTorch library to perform our deep learning model.

3.4. Model Validation

Due to the limitation of the sample data, our model was evaluated using k-fold cross validation technique in which the dataset is randomly split into k equal sized samples, and one of these is used for getting the classification performance. This process is repeated k times to ensure that the model is not overfitted (Moore, 2001).

4. EXPERIMENTS AND RESULTS

In our experiments, ASD-SAENet was evaluated in two different scenarios. First, the whole dataset containing 1,035 subjects were used to evaluate our model, and then we tested the model on each site separately. Evaluating each site separately demonstrates how our model performs on small datasets, and how it generalizes across different data acquisition sites and MRI machines. Due to the limitation of the sample data, our model was evaluated using k-fold cross validation technique in which the dataset is randomly split into k equal sized samples, and one of these is used for getting the classification performance. This process is repeated k times to ensure that the model is not overfitted (Moore, 2001). The details of each strategy are explained in the following subsections.

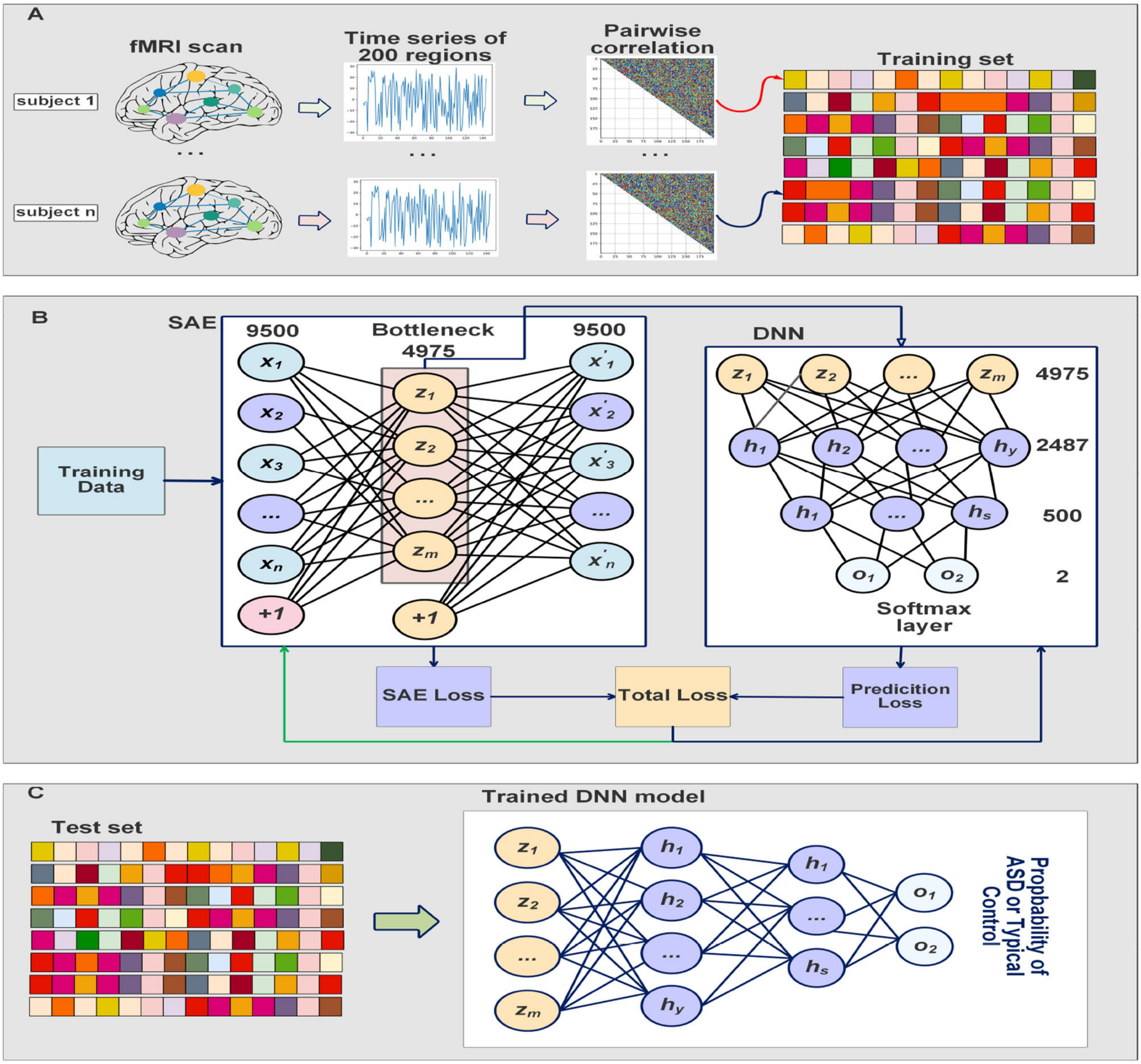

FIGURE 2 | (A) Shows the preprocessing steps which include extract time series from fMRI scans, calculate the Pearson's correlations, and then 1/4 smallest and largest average pairwise correlations were selected for feature vectors. **(B)** Explain how our model is trained at the same time to improve feature selection while obtaining optimal classification model. The DNN classifier input is the bottleneck of the SAE. **(C)** Shows the testing process where the input subject is fed into the trained SAE, and then the DNN will take the bottleneck to make the classification using softmax layer.

4.1. Average Accuracy for the ABIDE Dataset

In this experiment, we chose k as 10 to perform 10-fold cross validation using the whole dataset. We compare our proposed model with ASD-DiagNet model (Eslami et al., 2019), and the method proposed by Heinsfeld et al. (2018). **Table 2** shows the comparison of accuracy, sensitivity, specificity, and the running time with these state-of-the-art tools. The result shows that ASD-SAENet achieves 70.8% which is comparable to the average accuracy of these methods.

4.2. Accuracy for Each Data Acquisition Site

In these experiments, we performed a 5-fold cross validation for each site because of the limitation of the size of the data. **Table 3** shows accuracy, sensitivity, and specificity of ASD-SAENet for each site, and **Table 4** compares the accuracy with other approaches. These results demonstrate that our proposed model outperforms other state-of-the-art methods, and exhibits better accuracy for 12 out of the 17 sites. The average accuracy achieved by ASD-SAENet model was 64.42% which is comparable to other

TABLE 2 | This table shows the comparison between ASD-SAENet and the state-of-the-art methods using the whole ABIDE dataset.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	Running time
ASD-SAENet	**70.8**	62.2	**79.1**	52.74 min
ASD-DiagNet (Eslami et al., 2019)	70.3	68.3	72.2	41.14 min
Heinsfeld et al., 2018	70	**74**	63	7 h

The bold values indicate the best accuracy, sensitivity, and specificity for the three models.

TABLE 3 | This table shows accuracy, sensitivity, and specificity using our proposed ASD-SAENet model for each imaging site of ABIDE dataset.

Site	Accuracy (%)	Sensitivity (%)	Specificity (%)
Caltech	56.7	68.3	45
CMU	70.6	93.3	46.6
KKI	72.6	60	82
Leuven	64.6	50.6	77.1
MaxMun	47.5	49	54
NYU	72	67.9	75
OHSU	72	50	90.3
OLIN	66.6	81.6	46.6
PITT	**73.1**	78.6	64.6
SBL	56.6	60	53.3
SDSU	64.2	53.3	65.9
Stanford	53.2	36.6	70
Trinity	57.5	48	64
UCLA	68.3	72.3	64.1
UM	67.8	79	58.3
USM	70	63.5	71.5
Yale	66	69.3	64
Average	64.6	63.6	64.2

The bold value indicates the best accuracy among all the sites.

TABLE 4 | This table shows the comparison between ASD-SAENet and state-of-the-art methods ASD-DiagNet (Eslami et al., 2019), and Heinsfeld et al. (2018) for each site of ABIDE dataset.

Site	Data size		Accuracy (%)		
	ASD	Typical control	ASD-SAENet	ASD-DiagNet (Eslami et al., 2019)	Heinsfeld et al., 2018
Caltech	19	18	**56.7**	52.8	52.3
CMU	14	13	**70.6**	68.5	45.3
KKI	20	28	**72.6**	69.5	58.2
Leuven	29	34	**64.6**	61.3	51.8
MaxMun	24	28	47.5	48.6	**54.3**
NYU	75	100	**72**	68	64.5
OHSU	12	14	72	**82**	74
OLIN	19	15	**66.6**	65.1	44
PITT	29	27	**73.1**	67.8	59.8
SBL	15	15	**56.6**	51.6	46.6
SDSU	14	22	**64.2**	63	63.6
Stanford	19	20	53.2	**64.2**	48.5
Trinity	22	25	57.5	54.1	**61**
UCLA	54	44	68.3	**73.2**	57.7
UM	66	74	**67.8**	64.2	57.6
USM	46	25	**70**	68.2	62
Yale	28	28	**66**	63.8	57.6
	Average		**64.6**	63.8	56.1

The bold numbers indicate the best accuracy for the three models. As can be seen that ASD-SAENet exhibits better average-accuracy, as well as superior accuracy for 12 out of the 17 sites that are part of the ABIDE benchmark.

methods as well. The fact that ASD-SAENet model exhibits better accuracy for more number of sites as compared to both state-of-the-art methods shows the robustness and generalizability of our proposed model.

5. CONCLUSIONS AND DISCUSSIONS

More than 1.5 Million children in the US are affected by heterogeneous Autism spectrum disorder (ASD) which has wide range of symptoms and characteristics such as limited communication (including verbal and non-verbal), limited social interaction, and may exhibit repeated or limited interests or activities. The diagnostic challenges using clinical techniques have resulted in significant interest in identifying a biomarkers, and consequently an objective test that correctly classifies children with and without the disorder earlier than the current timeline. However, before any such test can be administered at clinical level; sufficient understanding of the neurobiological underpinning of ASD is essential. In the recent decade, advances in neuroimaging technologies are providing a critical step, and has made it possible to measure that pathological changes associated with ASD brain. Imaging techniques such as structural MRI, and functional MRI (to detect the alterations in function, connectivity of the brain) can be used to detect the changes in the brain. The underlying fMRI data has the features that can be used to distinguish between ASD and healthy controls. However, the subtle changes in the ASD brain as compared to the healthy controls make it impossible to identify, and detect biomarkers using conventional computational or statistical methods. Advanced machine-learning solutions offer a systematic approach to developing automatic solutions for objective classification, and learn the subtle patterns in the data that might be specific to ASD brains.

In this paper, we have designed, and developed a deep-learning method, called *ASD-SAENet* for classifying brain scans that exhibit ASD from healthy controls scans. Our novel deep-learning model utilizes sparse autoencoders which are more open to interpretability, and may advance our understanding of the neurobiological underpinning of the ASD brain. The fMRI data used for training, and evaluating our deep-learning model is provided by ABIDE consortium, which has been collected from 17 different MRI data acquisition imaging centers. Our proposed model uses the Pearson's correlations of 200 regions of the brain as features which are fed into a sparse autoencoder to lower the dimensionality of the features. These features are then fed to the two-hidden layer deep-learning network with softmax function as output layer. Our proposed sparse autoencoder, and the deep-network is trained simultaneously for feature selection and improving classifier decision. Any further training to improve the classifier was done by executing more iterations with autoencoder kept at a constant state. Two major sets of experiments were performed for evaluation of our proposed model: First, we used the whole dataset and performed 10-fold cross-validation. We achieved 70.8% in 51 min which is significantly shorter than 6 hours required by other methods (Heinsfeld et al., 2018) while resulting in better accuracy. Second, we tested our method on each site using 5-fold cross-validation. The sparse autoencoder, coupled with limited amount of fMRI data that is available for ASD, demonstrates a computationally light-weight machine-learning module for ASD biomarker identification, and classification. Our extensive experimentation has shown that the proposed ASD-SAENet model gives higher accuracy for 12 out of the 17 centers that are part of ABIDE benchmark. Combined with the average accuracy of ASD-SAENet closer to the accuracy of state-of-the-art models (Heinsfeld et al., 2018; Eslami et al., 2019) conclusively shows that ASD-SAENet is a generalizable model, and is more robust to different data acquisition, various MRI machines, and (pre- and post-processing) protocols that are followed to acquire the fMRI data. Robustness of our ASD-SAENet model is a significant improvement to the variance observed in many existing state-of-the-art machine-learning models for ASD classification, and clearly suitable for further development for clinical usage in the future.

The variation in site-specific accuracies (shown in **Tables 3**, **4**) can be explained by the variation in different data acquisition protocols (Power et al., 2012) involving different scanners, parameters, age range, as well as the post-data acquisition protocols that are followed by different groups Botvinik-Nezer et al. (2020). For example, in our results, the highest accuracy was on the PITT site dataset, which used Siemens scanner, repetition time of 1,500 ms, echo time of 25 ms, flip angle of 70 degrees, and an age range of 9.3–35.2. The lowest accuracy with almost similar data size was on MaxMun dataset which used same MRI scanner, different parameters (i.e., 3,000 for repetition time, 30 for echo time, and 80 for flip angle degree), and a huge gap of age range. Our results also show that different parameters and scanners can affect the quality of the data, and hence the performance of the deep-learning models. For example, most of the sites that achieved around 70% with our model were using Siemens scanner, repetition time is between 1,500 and 2,500, and age-range not highly variable. The results also demonstrate that there is a correlation between the echo time and the flip angle degree, and their sum between 95 and 105 gives the better performance for our deep-learning models. This empirical finding may show that Siemens scanner can work well when there is a correlation between echo time and flip angle degree. However, more studies and experiments are needed for confirmation, and how these data-acquisition parameters effect the features that are extracted by our deep-learning models.

Our extensive experimentation demonstrate that ASD-SAENet exhibits similar accuracy (70.8 vs. 70.3%) to our earlier proposed ASD-DiagNet (Eslami et al., 2019) model, superior specificity (79.1 vs. 72.2%) but slight decrease in sensitivity (62.2 vs. 68.3%). We attribute this slight decrease in sensitivity to the usage of sparse autoencoder in which only a small number of the hidden units are allowed to be active at the same time, and may miss some features. However, the strength of the model outweighs the small decrease in the sensitivity by exhibiting superior specificity. We also show that the number of true-negative rate is comparatively less than other state of the art methods; leading to classifiers that could be used in real-world i.e., since most of the population is not ASD, typical control people should be correctly identified as not having the condition. Additional advantages of the *ASD-SAENet* is the unique statistical features, and low computational cost which will help in identifying the feature importance estimates for our future studies. Absence of advance computational techniques such as interpretable deep-learning models that can process multimodal datasets such as sMRI and fMRI is a major technical hurdle in identifying ASD biomarkers. Investigation of these multimodal deep-learning models combined with the sparse autoencoder based classification strategy will allow us to device methods which will make the interpretation of the deep-learning models possible leading to better understanding of the neurobiological underpinning of the Autism Spectrum Disorder.

5.1. Limitations

This study has some limitations. First, although comparable to other published studies, the present study has a modest sample size for training and evaluation for our deep-learning model. Second, our model shows superior generalizability across multiple data acquisition sites but the features that might be most effective in classification cannot be determined due to non-interpretability of the deep-learning model. Third, there might be potential group differences due to head movement when using fMRI functional connectivity measures as input. The authors believe that such differences cannot be systematic due to variance in the subjects, scanning sites, and procedures that are site dependent. However, there is no way to demonstrate that the subtle changes picked up by deep-learning models are due to neurological difference, or due to head movements. We also believe that if a model achieves reliable classification accuracy (especially across different sites) despite such noise generated from different equipment and demographics shows promise for machine learning applications to ASD diagnosis and understanding. All of these limitations are at par with other machine-learning methods that have been proposed for biomarkers identification, and classification.

DATA AVAILABILITY STATEMENT

The datasets used in this study can be found in the ABIDE repository (http://preprocessed-connectomes-project.org/abide/).

AUTHOR CONTRIBUTIONS

FS conceived the study. FS and FA designed the machine-learning model and wrote and edited the manuscript. FA completed its implementations and reported results. All authors contributed to the article and approved the submitted version.

FUNDING

Research reported in this paper was partially supported by NIGMS of the National Institutes of Health (NIH) under award number R01GM134384. The content was solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. FS was additionally supported by the NSF CAREER award OAC 1925960. In addition, FA was supported by Imam Mohammad Ibn Saud Islamic University (IMSIU) under scholarship number 44946.

REFERENCES

Abraham, A., Milham, M. P., Di Martino, A., Craddock, R. C., Samaras, D., Thirion, B., et al. (2017). Deriving reproducible biomarkers from multi-site resting-state data: an autism-based example. *NeuroImage* 147, 736–745. doi: 10.1016/j.neuroimage.2016.10.045

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., et al. (2018). Prevalence of autism spectrum disorder among children aged 8 years–autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *MMWR Surveill. Summar*. 67:1. doi: 10.15585/mmwr.ss6706a1

Bilgen, I., Guvercin, G., and Rekik, I. (2020). Machine learning methods for brain network classification: application to autism diagnosis using cortical morphological networks. *arXiv preprint arXiv:2004.13321*. doi: 10.1016/j.jneumeth.2020.108799

Boat, T. F., Wu, J. T., and National Academies of Sciences, Engineering, and Medicine and others. (2015). "Clinical characteristics of intellectual disabilities," in *Mental Disorders and Disabilities Among Low-Income Children* (Washington, DC: National Academies Press)

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., et al. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* 582, 84–88. doi: 10.1038/s41586-020-2314-9

Bradshaw, J., Steiner, A. M., Gengoux, G., and Koegel, L. K. (2015). Feasibility and effectiveness of very early intervention for infants at-risk for autism spectrum disorder: a systematic review. *J. Autism Dev. Disord*. 45, 778–794. doi: 10.1007/s10803-014-2235-2

Brown, C. J., Kawahara, J., and Hamarneh, G. (2018). "Connectome priors in deep neural networks to predict autism," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (Washington, DC), 110–113. doi: 10.1109/ISBI.2018.8363534

Craddock, C., Benhajali, Y., Chu, C., Chouinard, F., Evans, A., Jakab, A., et al. (2013). The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Neuroinformatics* 4. doi: 10.3389/conf.fninf.2013.09.00041

Craddock, R. C., James, G. A., Holtzheimer, P. E. III, Hu, X. P., and Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. Brain mapp*. 33, 1914–1928. doi: 10.1002/hbm.21333

Deshpande, G., Wang, P., Rangaprakash, D., and Wilamowski, B. (2015). Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data. *IEEE Trans. Cybern*. 45, 2668–2679. doi: 10.1109/TCYB.2014.2379621

Di Martino, A., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* 19, 659–667. doi: 10.1038/mp.2013.78

Dichter, G. S. (2012). Functional magnetic resonance imaging of autism spectrum disorders. *Dialog. Clin. Neurosci*. 14:319. doi: 10.31887/DCNS.2012.14.3/gdichter

Dvornek, N. C., Ventola, P., Pelphrey, K. A., and Duncan, J. S. (2017). "Identifying autism from resting-state fMRI using long short-term memory networks," in *International Workshop on Machine Learning in Medical Imaging* (Quebec City, QC: Springer), 362–370. doi: 10.1007/978-3-319-67389-9_42

El Gazzar, A., Cerliani, L., van Wingen, G., and Thomas, R. M. (2019a). "Simple 1-D convolutional networks for resting-state fMRI based classification in autism," in *2019 International Joint Conference on Neural Networks (IJCNN)* (Budapest), 1–6. doi: 10.1109/IJCNN.2019.8852002

El-Gazzar, A., Quaak, M., Cerliani, L., Bloem, P., van Wingen, G., and Thomas, R. M. (2019b). "A hybrid 3dcnn and 3dc-lstm based model for 4d spatio-temporal fMRI data: an abide autism classification study," in *OR 2.0 Context-Aware Operating Theaters and Machine Learning in Clinical Neuroimaging* (Shenzhen: Springer), 95–102. doi: 10.1007/978-3-030-32695-1_11

Eslami, T., Mirjalili, V., Fong, A., Laird, A. R., and Saeed, F. (2019). ASD-diagnet: a hybrid learning approach for detection of autism spectrum disorder using fMRI data. *Front. Neuroinform*. 13:70. doi: 10.3389/fninf.2019.00070

Eslami, T., and Saeed, F. (2018). "Similarity based classification of ADHD using singular value decomposition," in *Proceedings of the 15th ACM International Conference on Computing Frontiers* (Ischia), 19–25. doi: 10.1145/3203217.3203239

Eslami, T., and Saeed, F. (2019). "Auto-ASD-network: a technique based on deep learning and support vector machines for diagnosing autism spectrum disorder using fMRI data," in *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics* (Niagara Falls, ON), 646–651. doi: 10.1145/3307339.3343482

Fredo, A., Jahedi, A., Reiter, M., and Müller, R.-A. (2018). Diagnostic classification of autism using resting-state fMRI data and conditional random forest. *Age* 12, 6–41.

Goodfellow, I., Bengio, Y., and Courville, A. (2016a). "6.2. 2.3 softmax units for multinoulli output distributions," in *Deep Learning* (MIT Press), 180–184.

Goodfellow, I., Bengio, Y., Courville, A., and Bengio, Y. (2016b). *Deep Learning, Vol. 1*. Cambridge, MA: MIT Press.

Guo, X., Dominick, K. C., Minai, A. A., Li, H., Erickson, C. A., and Lu, L. J. (2017). Diagnosing autism spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method. *Front. Neurosci*. 11:460. doi: 10.3389/fnins.2017.00460

Haweel, R., Shalaby, A., Mahmoud, A., Seada, N., Ghoniemy, S., Ghazal, M., et al. (2020). A robust DWT-CNN based cad system for early diagnosis of autism using task-based fMRI. *Med. Phys*. doi: 10.1002/mp.14692

Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., and Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the abide dataset. *NeuroImage* 17, 16–23. doi: 10.1016/j.nicl.2017.08.017

Iidaka, T. (2015). Resting state functional magnetic resonance imaging and neural network classified autism and control. *Cortex* 63, 55–67. doi: 10.1016/j.cortex.2014.08.011

Just, M. A., Cherkassky, V. L., Keller, T. A., Kana, R. K., and Minshew, N. J. (2007). Functional and anatomical cortical underconnectivity in autism: evidence from an fMRI study of an executive function task and corpus callosum morphometry. *Cereb. Cortex* 17, 951–961. doi: 10.1093/cercor/bhl006

Kazeminejad, A., and Sotero, R. C. (2019). Topological properties of resting-state fMRI functional networks improve machine learning-based autism classification. *Front. Neurosci*. 12:1018. doi: 10.3389/fnins.2018.01018

Khosla, M., Jamison, K., Kuceyeski, A., and Sabuncu, M. R. (2018). "3D convolutional neural networks for classification of functional connectomes," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, eds D. Stoyanov, G. Carneiro, Z. Taylor, and T. Syeda-Mahmood (Granada: Springer), 137–145. doi: 10.1007/978-3-030-00889-5_16

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lau, W. K., Leung, M.-K., and Lau, B. W. (2019). Resting-state abnormalities in autism spectrum disorders: a meta-analysis. *Sci. Rep*. 9, 1–8. doi: 10.1038/s41598-019-40427-7

Li, H., Parikh, N. A., and He, L. (2018). A novel transfer learning approach to enhance deep neural network classification of brain functional connectomes. *Front. Neurosci*. 12:491. doi: 10.3389/fnins.2018.00491

Lord, C., Elsabbagh, M., Baird, G., and Veenstra-Vanderweele, J. (2018). Autism spectrum disorder. *Lancet* 392, 508–520. doi: 10.1016/S0140-6736(18)31129-2

Mizuno, Y., Kagitani-Shimono, K., Jung, M., Makita, K., Takiguchi, S., Fujisawa, T. X., et al. (2019). Structural brain abnormalities in children and adolescents with comorbid autism spectrum disorder and attention-deficit/hyperactivity disorder. *Transl. Psychiatry* 9, 1–7. doi: 10.1038/s41398-019-0679-z

Moore, A. W. (2001). *Cross-Validation for Detecting and Preventing Overfitting*. School of Computer Science Carneigie Mellon University.

Mostafa, S., Tang, L., and Wu, F.-X. (2019a). Diagnosis of autism spectrum disorder based on eigenvalues of brain networks. *IEEE Access* 7, 128474–128486. doi: 10.1109/ACCESS.2019.2940198

Mostafa, S., Yin, W., and Wu, F.-X. (2019b). "Autoencoder based methods for diagnosis of autism spectrum disorder," in *International Conference on Computational Advances in Bio and Medical Sciences* (Miami, FL: Springer), 39–51. doi: 10.1007/978-3-030-46165-2_4

National Collaborating Centre for Mental Health (2009). *Attention Deficit Hyperactivity Disorder: Diagnosis and Management of ADHD in Children, Young People and Adults*. British Psychological Society.

Ng, A. (2011). *Sparse Autoencoder*. CS294A Lecture notes, 72, 1–19.

Nickel, R. E., and Huang-Storms, L. (2017). Early identification of young children with autism spectrum disorder. *Indian J. Pediatr*. 84, 53–60. doi: 10.1007/s12098-015-1894-0

Niu, K., Guo, J., Pan, Y., Gao, X., Peng, X., Li, N., et al. (2020). Multichannel deep attention neural networks for the classification of autism spectrum disorder using neuroimaging and personal characteristic data. *Complexity* 2020:1357853. doi: 10.1155/2020/1357853

Nogay, H. S., and Adeli, H. (2020). Machine learning (ML) for the diagnosis of autism spectrum disorder (ASD) using brain imaging. *Rev. Neurosci*. 1. doi: 10.1515/revneuro-2020-0043

Parikh, M. N., Li, H., and He, L. (2019). Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data. *Front. Comput. Neurosci*. 13:9. doi: 10.3389/fncom.2019.00009

Plitt, M., Barnes, K. A., and Martin, A. (2015). Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. *NeuroImage* 7, 359–366. doi: 10.1016/j.nicl.2014.12.013

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59, 2142–2154. doi: 10.1016/j.neuroimage.2011.10.018

Sarraf, S., and Tofighi, G. (2016). Classification of Alzheimer's disease using fMRI data and deep learning convolutional neural networks. *arXiv preprint arXiv:1603.08631*.

Sherkatghanad, Z., Akhondzadeh, M., Salari, S., Zomorodi-Moghadam, M., Abdar, M., Acharya, U. R., et al. (2019). Automated detection of autism spectrum disorder using a convolutional neural network. *Front. Neurosci*. 13:1325. doi: 10.3389/fnins.2019.01325

Stevens, T., Peng, L., and Barnard-Brak, L. (2016). The comorbidity of ADHD in children diagnosed with autism spectrum disorder. *Res. Autism Spectr. Disord*. 31, 11–18. doi: 10.1016/j.rasd.2016.07.003

Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P.-A. (2008). "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th International Conference on Machine Learning* (Helsinki), 1096–1103. doi: 10.1145/1390156.1390294

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., and Bottou, L. (2010). Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res*. 11, 3371–3408. doi: 10.5555/1756006.1953039

Wang, C., Xiao, Z., Wang, B., and Wu, J. (2019). Identification of autism based on SVM-RFE and stacked sparse auto-encoder. *IEEE Access* 7, 118030–118036. doi: 10.1109/ACCESS.2019.2936639

Xiao, Z., Wang, C., Jia, N., and Wu, J. (2018). Sae-based classification of school-aged children with autism spectrum disorders using functional magnetic resonance imaging. *Multim. Tools Appl*. 77, 22809–22820. doi: 10.1007/s11042-018-5625-1

Yao, Q., and Lu, H. (2019). "Brain functional connectivity augmentation method for mental disease classification with generative adversarial network," in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)* (Xi'an: Springer), 444–455. doi: 10.1007/978-3-030-31654-9_38

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.

frontiers
in Computational Neuroscience

REVIEW
published: 05 May 2021
doi: 10.3389/fncom.2021.670489

Can Deep Learning Hit a Moving Target? A Scoping Review of Its Role to Study Neurological Disorders in Children

*Saman Sargolzaei**

Department of Engineering, College of Engineering and Natural Sciences, University of Tennessee at Martin, Martin, TN, United States

Neurological disorders dramatically impact patients of any age population, their families, and societies. Pediatrics are among vulnerable age populations who differently experience the devastating consequences of neurological conditions, such as attention-deficit hyperactivity disorders (ADHD), autism spectrum disorders (ASD), cerebral palsy, concussion, and epilepsy. System-level understanding of these neurological disorders, particularly from the brain networks' dynamic perspective, has led to the significant trend of recent scientific investigations. While a dramatic maturation in the network science application domain is evident, leading to a better understanding of neurological disorders, such rapid utilization for studying pediatric neurological disorders falls behind that of the adult population. Aside from the specific technological needs and constraints in studying neurological disorders in children, the concept of development introduces uncertainty and further complexity topping the existing neurologically driven processes caused by disorders. To unravel these complexities, indebted to the availability of high-dimensional data and computing capabilities, approaches based on machine learning have rapidly emerged a new trend to understand pathways better, accurately diagnose, and better manage the disorders. Deep learning has recently gained an ever-increasing role in the era of health and medical investigations. Thanks to its relatively more minor dependency on feature exploration and engineering, deep learning may overcome the challenges mentioned earlier in studying neurological disorders in children. The current scoping review aims to explore challenges concerning pediatric brain development studies under the constraints of neurological disorders and offer an insight into the potential role of deep learning methodology on such a task with varying and uncertain nature. Along with pinpointing recent advancements, possible research directions are highlighted where deep learning approaches can assist in computationally targeting neurological disorder-related processes and translating them into windows of opportunities for interventions in diagnosis, treatment, and management of neurological disorders in children.

Keywords: ADHD, ASD, cerebral palsy, concussion, deep learning, epilepsy, network neuroscience, neurological disorders

OPEN ACCESS

Edited by:
Xiaoli Li,
Beijing Normal University, China

Reviewed by:
HengJin Ke,
Wuhan University, China
He Chen,
Beijing Normal University, China

****Correspondence:***
Saman Sargolzaei
saman@utm.edu

Received: *21 February 2021*
Accepted: *09 April 2021*
Published: *05 May 2021*

Citation:
Sargolzaei S (2021) Can Deep Learning Hit a Moving Target? A Scoping Review of Its Role to Study Neurological Disorders in Children. Front. Comput. Neurosci. 15:670489. doi: 10.3389/fncom.2021.670489

1. BACKGROUND AND HISTORY

The brain, as the body commander in chief, evolves by passing through multiple developmental and maturation stages, from the neurogenesis (neuron production stage) and migration (neuron translocation to the neocortex stage) to the differentiation (neurons integration in specialized neural networks by forming axonal connections) (Stiles and Jernigan, 2010). Before transition into the postnatal phase, a massive systematic synaptic exuberance and pruning occurs. After birth, axonal myelination (the process of sheath formation) comes into play that significantly improves axonal integrity and conductance. In support of emerging brain connectivity, synaptic exuberance and pruning continue to occur in parallel with myelination. Tackling neurological disorders by studying the brain structure and function has long been in the neuroscientific research spotlight.

In 1906, an exciting milestone was set for the history of modern neuroscience when the Noble Prize of physiology and medicine was shared between two people, let us indeed reemphasize, between two opponent theories, proposed by Camillo Golgi (1843–1926) and Santiago Ramon y Cajal (1852–1934) (Swanson et al., 2007). Their opposition is rooted in their opinions of how neurons, as the critical units of nervous systems, intercommunicate. Golgi's reticular theory formulated the nervous system as a continuous organization, wherein Cajal's neuron doctrine expressed it as a contiguous organization. While scholarly activities have been on the rise more in favor of the contiguous organization, a growing belief in networks' role has persistently emerged in studying the brain under normal and pathology. The emergence of brain connectivity networks is an essential aspect of brain plasticity, also known as neuroplasticity, a brain adaptation process from experience and learning. During development, brain plasticity rises as it is exposed to environmental events (Rosenzweig and Bennett, 1996). Neuroimaging studies have confirmed the dynamic evolution of these cortical networks through use-induced plasticity to learn and improve functions (Hua and Smith, 2004).

Our current understanding of the connectivity role is the direct consequence of advancements in two main inter-related domains, hardware and computational algorithms. The breakthrough of network-level studies of brain activities was expedited by a series of technology inventions that began by Electroencephalography (EEG) (Britton et al., 2016) and continued with Magnetoencephalography (MEG) (Cohen, 1968), X-ray, Computed Tomography (CT), and Positron Emission Tomography (PET) in the 70's (Raichle, 2009). In the late 70s, Magnetic Resonance Imaging (MRI) was introduced, further boosting brain-related discoveries. The expedition offered by the MRI technology itself was obscured until the advent of different MRI sequences. The MRI sequence refers to a particular harmony set between radio-frequency pulses and magnetic gradients, leading to capturing a specific perspective from the tissue appearance. In 1990 (Ogawa et al., 1990), oxygen consumption and supply to cerebral regions was, for the first time, considered as an endogenous contrast agent for recording brain functional activities. The method, called Blood-Oxygen-Level-Dependant (BOLD) functional MRI, has since been a dominant MRI method of choice in resting-state functional MRI (rs-fMRI) task-based functional MRI, both based on the regional association of brain activity with oxygen supply and consumption. Shifting the focus from functional imaging, The knowledge of brain microstructure diffusion properties has driven diffusion-weighted imaging (Basser et al., 1994) to streamline structural imaging. In parallel to hardware-related technological advancements, computational and visualization power has immensely matured following the development of computing backbones such as C/C++ and python programming languages and the ease of virtual memorization and parallel processing. The development of tools like AFNI (Cox, 1996), FreeSurfer (Fischl, 2012), FSL (Jenkinson et al., 2012), and SPM (Penny et al., 2011) has complemented these inter-related efforts. The rising investment in portraying the brain and its role in diseases and disorders is reflected amid initiatives such as the Brain Initiative or the International Brain Initiative.

The field has been immensely grown since Golgi and Ramon y Cajal's neuroanatomical work that emphasized the role of the functional interplay among regional brain structures. Nevertheless, the growth has happened at a lower rate within discoveries related to the knowledge we have about children's brain networks. **Figure 1** captures a synopsis of scholarly activities within the "brain network" domain over the years highlighted with milestones of technology advancements. Contrasted by the age of 19 years old, the graphs show the annual number of full-text articles found in the PubMed search engine with [(brain) AND (network)] term, normalized by the total number of publications found by [(brain)] keyword. From the early 1970s, there has been a steady increase in brain research's scholarly activities, particularly under the umbrella of the brain network. However, the increase in trends shows a slower rate for studies in 19 years old and younger age population.

2. CHALLENGE AND PROSPECT

Natural and technical challenges (**Figure 2**) could be deemed for the slower trend of scholarly activities related to pediatric brain studies. The natural, or to better frame it, the inherent difficulty in studying the developmental brain is the development factor's attachment. The development of a complex biological entity, such as the human brain, entails significant synergies among phased development stages, from the genetic blueprint to the formation of brain structure on the foundation of millions of brain cells, including neuronal and glial cells (Gibb and Kovalchuk, 2018). It becomes even further intricate when adding the element of functional network formation to the developmental stages. Brain circuitry emergence is the result of brain cell interactions over time. It is a set of highly dynamic processes, including iterative formation and elimination of synapses and stabilization of relevant synaptic connections to mature functional brain activity in conjunction with brain structure (Kuczmarski, 2002; Hua and Smith, 2004). Aside from the genetic blueprint, it is recognized to depend upon other factors, including nutrition,

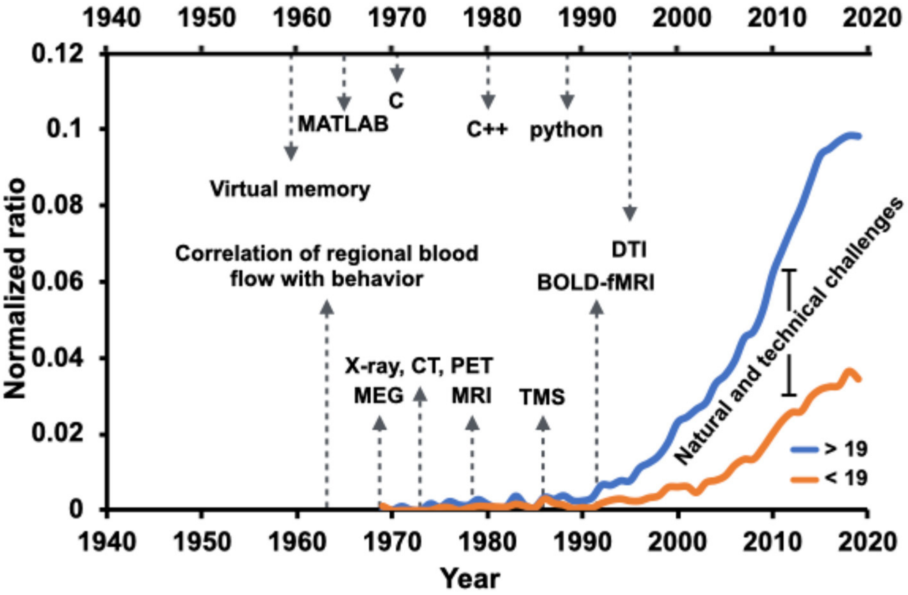

FIGURE 1 | Brain network-related scholarly activities in the pediatric population (younger than 19 years old of age) have paced at a slower rate compared to the adult population (more aged than 19 years old of age), a possible reason being the natural and technical challenges. Normalized ratio of the PubMed-indexed full text articles including [(brain) AND (network)] term, normalized by the total number of publications found by [(brain)] keyword, overlaid with advancement milestones in the field of technology and computation. The annual ratio is visualized for two age groups, younger and older than 19 years old.

environmental exposures, and experiences such as interactions with people (Robinson et al., 2017).

The amendment of disease-driven processes to the ensemble of normal brain development processes introduces further uncertainty and complexity, particularly for studies concerned with accurate delineation of normal development operations from those operations driven by pathology. The natural challenge is to an extent that cautious needs to be practiced before generalizing findings in neuroimaging studies of children and adolescents (Santosh, 2000). While neuroimaging, especially non-invasive ones, has become increasingly popular in the era of brain scientific investigation (Morita et al., 2016), its use is confined by issues such as non-compliance (Schlund et al., 2011; Raschle et al., 2012; Thieba et al., 2018). Non-compliance refers to the interruption or failure of the experiment, often due to subject movement or failure to abide by instruction due to anxiety. The issue arises even further for the clinical neuroimaging studies in children and adolescents to the extent that it necessitates proper considerations and study protocols (Greene et al., 2016).

The unique challenges of clinical studies currently extend beyond non-compliance to procedural difficulties, technical obstacles, and data processing methodological limitations. In this regard, The development and utilization of age-appropriate equipment can further streamline clinical studies in children and adolescents. Along with procedural and technical considerations, specific adjustments, age-dependent protocol developments, and analytical fine-tuning are essential for correct implications of neuroimaging studies in the pediatric population (Sargolzaei et al., 2015).

Figure 2 highlights an example of the technical considerations that must be practiced when the neuroimaging study population is pediatrics. It features the task of automatic intracranial volume (ICV) estimation in brain research. ICV estimation is a critically required task in neuroimaging studies. While the task may be fulfilled by manual inspection and landmarks identification, it is a tedious and laborious process. Automatic estimation involves the utilization of neuroimaging software packages and the implementation of a proper ICV estimation routine. As it is apparent from the shown figure, different software packages led into different ICV estimation when contrasted to the reference estimation of such quantity under different condition (control sample vs. pediatric patients with epilepsy)(Sargolzaei et al., 2015). The study emphasizes methodological challenges for pediatric neuroimaging studies, pointing out the necessity of guided decision-making in selecting developed tools under different circumstances.

To overcome the above-summarized challenges, indebted to the availability of high-dimensional data and increased computing capacities, approaches based on graph theory and machine learning have rapidly emerged a new trend to understand pathways better, accurately diagnose, and adequately manage the disorders. Deep learning has recently gained an ever-increasing role in the era of health and medical investigations under the umbrella of machine learning-based solutions to studies of neurological disorders. For the field of neurological disorders in children, our prospect is at the conjunction of applied deep learning engaged with graph theory on a personalized scale.

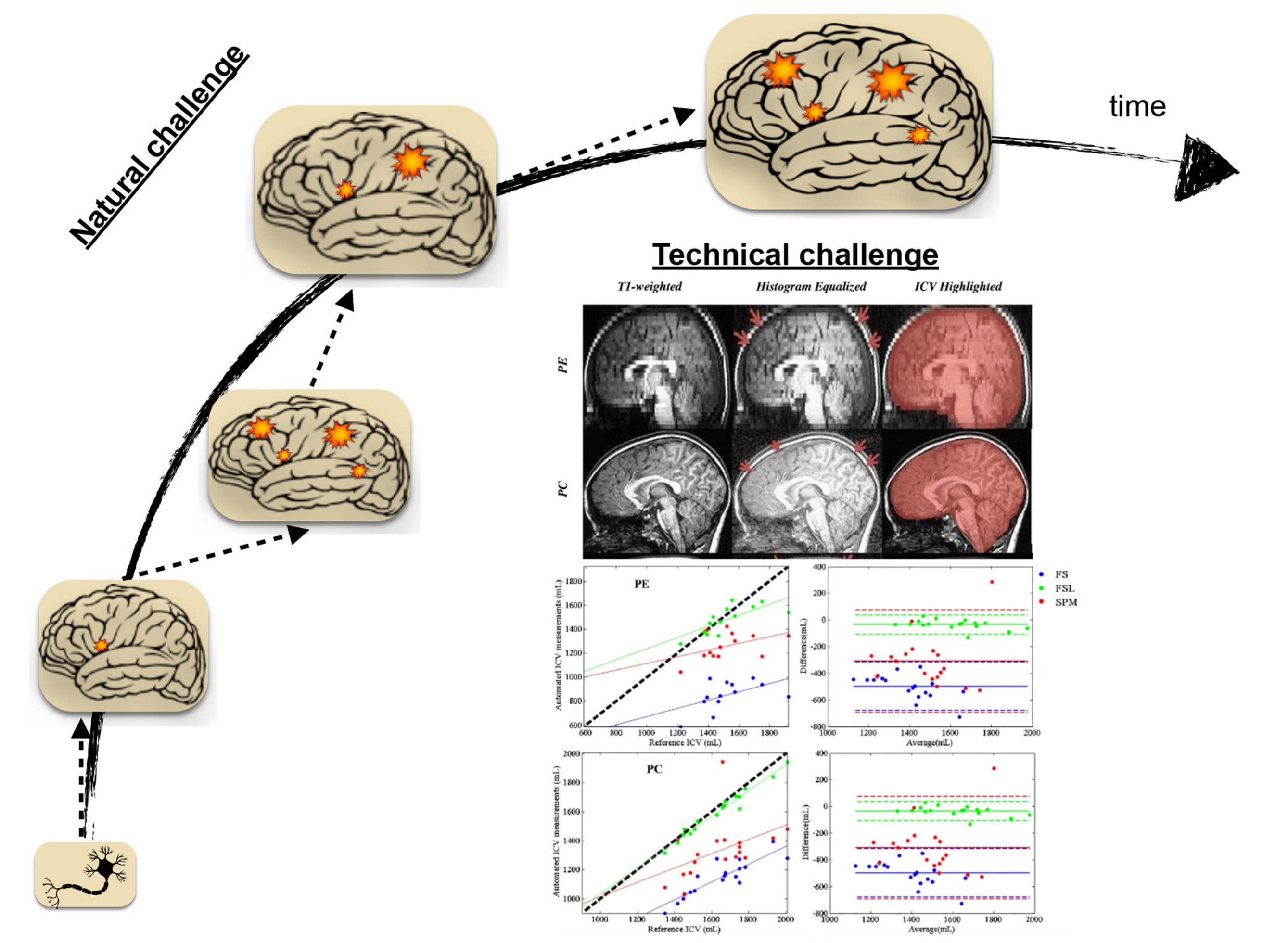

FIGURE 2 | Natural and technical challenges in the face of studying neurological disorders in the pediatric population. Over time, brain development involves many ongoing processes from the cellular to the network level. In the presence of disease or disorder, The ensemble of normal brain development processes is occluded by dynamic processes led by disease or condition, making the clinical studies naturally challenging in pediatric age groups. Technical challenges, including procedural difficulties and data processing methodological limitations, are also surrounding the brain's studies in the pediatric population. The presented technical problem here is reproduced with permission from Figures 2, 4 of Sargolzaei et al. (2015) for the task of intracranial volume estimation (ICV) using automated tools in two pediatric age group of human subjects (PE, Pediatric Epilepsy; PC, Pediatric Control).

3. DEEP LEARNING, SOMETHING NEW OR THE NEW FACE OF THE OLD

Granting machines the gift of intelligence is an ever-increasing appetite for humankind, and the learning skill is standing at the forefront of this intellectual gift. Through decades of artificial intelligence field evolution, machines learned through human-generated rules and accurately engineered features. This approach, conveniently referred to as the classical approach, has phenomenally succeeded in multiple application domains, yet extracting an optimized set of features and rules is not always cumbersome. Further tasks, such as recognition, which is intuitive for humans, remained a challenge for machines to learn via classical approaches that rely upon the existence of high-level abstract features (Goodfellow et al., 2016).

The deep learning approach's neurobiological spirit roots back in experimental studies (Hubel and Wiesel, 1959; Métin and Frost, 1989; Roe et al., 1992) discovering the mutually critical roles of intracortical circuit specifications and inputs to such circuits in the learning process of brain neural networks. Learning about the experience-dependant plasticity of the brain (Simons and Land, 1987; Kirkwood et al., 1995; Crist et al., 2001; Trachtenberg et al., 2002) inspires the methodological possibility of letting the machine also acquire knowledge by experience. Thorough experience-dependent knowledge acquisition involves exposure to a vast amount of such experiences, accompanied by corresponding outcomes, and the capacity to automatically form an intra-circuitry that maps these experiences to their corresponding outcome. Translating these requirements to artificial neural networks' jargon entails

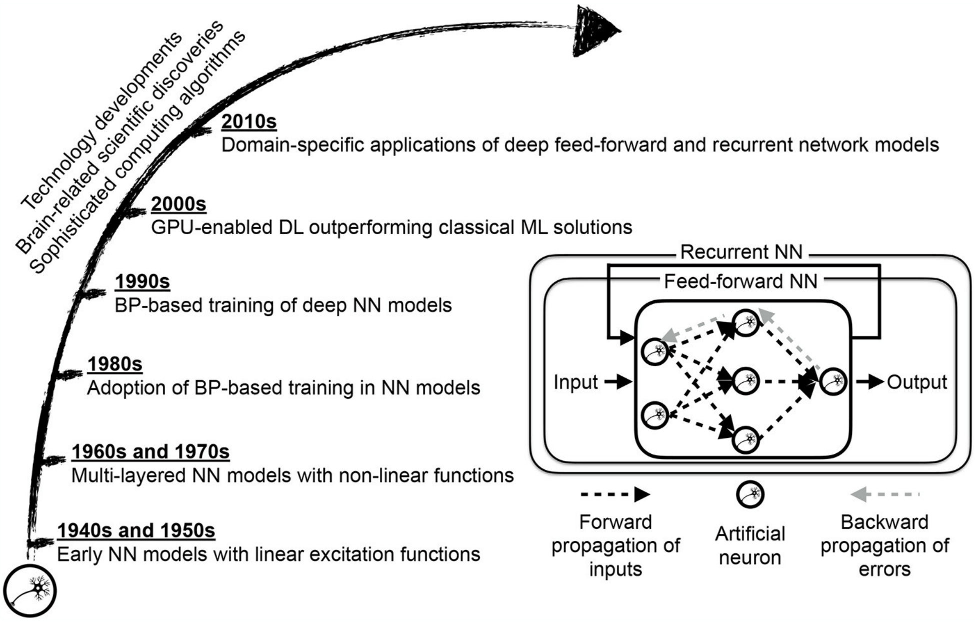
Technology developments
Brain-related scientific discoveries
Sophisticated computing algorithms
2010s
Domain-specific applications of deep feed-forward and recurrent network models
2000s
GPU-enabled DL outperforming classical ML solutions
1990s
BP-based training of deep NN models
1980s
Adoption of BP-based training in NN models
1960s and 1970s
Multi-layered NN models with non-linear functions
1940s and 1950s
Early NN models with linear excitation functions
Recurrent NN
Feed-forward NN
Input
Output
Forward propagation of inputs
Artificial neuron
Backward propagation of errors

FIGURE 3 | Deep learning (DL) approach maturation motivated by scientific discoveries of brain structure and function and enabled by computing technology and algorithms' immense progress. The core of a neural network (NN) model is an artificial neuron, resembling a biological neuron in a limited way, evaluating the collective input signals followed by conditional excitement. While early citations of NN models, as a linear regression variation, dates back to the 1800s, NN-based learning was not widespread until the development of sophisticated training algorithms based on back-propagation of error signals, and gradient descent optimization (Schmidhuber, 2015). The transition from classical NN-based learning, where the network examines extracted features from data to deep NN-based learning, where the network scans raw data representation, requires equipping NN with self-exploring capacity through multiple layers of abstraction. The graphical processing unit (GPU) based computing partially enabled such a self-exploratory multi-layered representation learning to the extent that deep learning becomes the choice method in numerous domain-specific applications (LeCun et al., 2015).

a specialized architecture, ultra-large datasets, and sophisticated linear and non-linear training algorithms, formulating the deep learning approach to the machine learning task. The specialized network architecture allows sequential processing of the supplied raw data units through multiple layers without the need for human-based feature design and engineering.

The deep learning approach (**Figure 3**) lets the network itself find an optimized allocation of credits to basic units (neurons) of these layers in sketching the desired outcome (LeCun et al., 2015; Schmidhuber, 2015). The network depth offers a way to break down a complex raw input into simplified units of information, mainly through convolution and sampling processes, and collectively map the input to the provided network output. Self-exploration mapping is the backbone of architectures such as deep convolutional neural networks (CNN) and recurrent neural networks (RNN), with the latter one empowering the learning of sequential data such as text and speech. Alongside other application sectors, healthcare has reported breakthroughs achieved with deep learning adoption in neuroimaging, genetics, oncology, radiation therapy, and drug discovery, to name a few (Wang et al., 2018, 2020; Boldrini et al., 2019; David et al., 2019; Serag et al., 2019; Tang et al., 2019; Zhu et al., 2019; Chen et al., 2020; Zhang et al., 2020).

4. DEEP LEARNING AND NEUROLOGICAL DISORDERS IN CHILDREN

To review the scope of deep learning application in diagnosis, treatment and management of neurological disorders in children, we surveyed existing peer-reviewed literature. The search was performed using PubMed (https://pubmed.ncbi.nlm.nih.gov), and IEEE Xplore (https://ieeexplore.ieee.org/Xplore/home.jsp). The search queries were set to [(deep learning) AND ((children) OR (child) OR (pediatric) OR (paediatric))] for PubMed search, and to [((“All Metadata”:Deep learning) AND ((“All Metadata”:children) OR (“All Metadata”:child) OR (“All Metadata”:pediatric)))] for IEEE Xplore search. Setting the inclusion criteria to peer-reviewed journal articles reporting the deep learning approach utilization in studying, diagnosing, or managing neurological disorders in the pediatric (up to 19 years of age) population led to a total of 22 records. Included articles were examined to retrieve the following information, study neurological condition (*Condition*), study population age range (*Age range*), study goal (*Goal*), modality of the used data (*Modality*), used deep learning model (*Model*), and backbone implementation environment (*Environment*). **Table 1** summarizes included studies.

TABLE 1 | A summary of found peer-reviewed journal articles utilizing deep learning to study, manage, diagnose, and prognosis of neurological conditions in pediatric (18 years or younger) populations.

Condition*	References	Age range**	Goal	Modality***	Model****	Environment*****
ADHD	Muñoz-Organero et al., 2018	[6–15]ye	Distinguishing medicated from non-medicated activity pattern	Tri-axial accelerometer	CNN	MATLAB
	Amado-Caballero et al., 2020	[6–15]ye	Automatic diagnosis of combined type ADHD	Activity recordings	CNN	MATLAB
	Muñoz-Organero et al., 2019	[6–16]ye	Assessment of movement patterns	Tri-axial accelerometer	LSTM-based RNN	
	Chen et al., 2019	[9–12]ye	Detecting spatiotemporal anomalies	EEG	CNN	
ASD	Bahado-Singh et al., 2019b	[24–79]ho	Identification of early biomarkers	Leucocyte DNA	Deep neural network	R h2o
	Hazlett et al., 2017	[6–24]mo	Prediction of autism diagnosis	MRI	Deep neural network	MATLAB
	Eni et al., 2020	[23–73]mo	Prediction of ASD severity	Speech	CNN	
	Aghdam et al., 2018	[5–10]ye	Classification of ASD from typically development control	rs-fMRI and sMRI	DBN	Theano
	Ghafouri-Fard et al., 2019	[6–14]ye	ASD status detection	Genomic data	Deep neural network	Keras
CP	Bahado-Singh et al., 2019a	[24–79]ho	CP prediction	Leucocyte DNA	Deep neural network	R h2o
Concussion	Boshra et al., 2019	[15–20]yy	Classification of concussed from control	EEG/ERP	CNN	Tensorflow v1.8.0
CHD brain dysmaturation	Ceschin et al., 2018	[35–45]we	Dysplastic cerebelli detection	MRI	CNN	Theano
Epilepsy	O'Shea et al., 2020	[39–42]we gestational	Seizure detection	EEG	CNN	
	Ansari et al., 2019	Neonates	Classifying epileptic from non-epileptic seizures	EEG	CNN with RF	MATLAB
	Bernardo et al., 2018	[2–60]mo	Detection of fast-ripples	EEG	CNN	Theano
	Daoud and Bayoumi, 2019	[3–15]ye	Early detection of pre-ictal state	EEG	CNN	
	Lin et al., 2020	[7–16]ye	Distinguishing patients without ED from controls	EEG	CNN	
	Lee et al., 2020	[4–16]ye	Tract classification for preoperative analysis	MRI-DWI	CNN	PyTorch 0.2
	Xu et al., 2019	[6–17]ye	Preoperative detection of axonal pathways	MRI-DWI	CNN	PyTorch 0.2
Brain tumor	Quon et al., 2020	92mo median	Posterior fossa tumor detection	MRI	CNN-ResNeXt	Keras
Schizophrenia	Aristizabal et al., 2020	[9–16]ye	Risk assessment	EEG	RNN	Keras
TSC	Sánchez Fernández et al., 2020	[5–16]ye	Cortical tubers detection	MRI	CNN	TensorFlow

**Condition is the neurological condition studied in each reference (ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; CP, cerebral palsy; CHD, congenital heart disease; TSC, Tuberous sclerosis complex). **Age range data is estimated and rounded from the given information for each reference (ye, years; ho, hours; mo, months; we, weeks). ***Modality refers to the data types used for learning (EEG, electroencephalograph, DNA, Deoxyribonucleic acid; MRI, magnetic resonance imaging; rs-fMRI, resting-state functional MRI; ERP, event-related potential; DWI, diffusion-weighted imaging). ****Model describes the core model utilized for deep learning given provided information in each reference (CNN, convolutional neural network; LSTM, long short-term memory; RNN, recurrent neural network; DBN, deep belief network; RF, random forest). *****Environment is where the deep learner being implemented in each referenced study (MATLAB by MathWorks, R interface for H2O package for large-scale machine learning, Theano, python library supporting large-scale multidimensional computations, Keras, A python deep learning API, TensorFlow, An open-source machine learning platform, PyTorch, an open-source machine learning framework.*

Perceived from **Table 1**, The deep learning architectures, in particular convolutional neural networks, have reported promises in a wide variety of applications, including automatic diagnosis, biomarker identification, early detection, within-condition type classification, and risk assessment for a variety of neurological conditions in children. The survey reveals the applicability potential of deep learning solutions beyond neuroimaging modalities, such as EEG and MRI, extending its use to other information sources such as speech and activity-based modalities within the pediatric clinical population. The latter inference from the survey is of utmost importance due to the challenges discussed before regarding methodological constraints of studies in pediatrics. While python-based libraries are the dominant library of choice reported in the current studies sample, MATLAB and R language libraries for large-scale machine learning implementations are also cited.

While deep learning solutions are relatively free of conventional feature engineering requirements, they are entangled with their choice of architecture and learning parameters, enforcing the need for careful implementation and selection of such internal architecture and its relevant parameters. The consequences of varying architecture are reported to have an impact on the deep learner performance for the use of convolutional neural network (Xu et al., 2019) and deep belief network (Aghdam et al., 2018). Therefore, the inclusion of benchmark routines, mainly reporting different architectures for optimum architecture selection, can enhance study findings.

Our survey search has also identified a rising interest in using deep learners to overcome inherent challenges, low tissue contrast, and dynamic appearance variation for brain imaging tasks in infants (Guo et al., 2014; Zhang et al., 2015; Kawahara et al., 2017; Mostapha and Styner, 2019; Sun et al., 2019) and fetus (Girault et al., 2019; Dou et al., 2020). A common goal for most surveyed articles is to conduct more extensive validation studies that involve multi-site investigations to make deep-learning-based computational tools more suitable as the non-invasive and inexpensive real-time clinical tool of choice.

5. VIEWPOINTS AND CONCLUSION

The current scoping review concentrated the scope on the rising use of deep learning to study, diagnose, and prognosis for children's neurological disorders. The hallmark of neurological disorders is the presence of brain dysfunction. Presentation of behavioral and psycho-behavioral symptoms is a consequence of such dysfunction. Neuroimaging studies have confirmed the dynamic evolution of these functional cortical networks through use-induced plasticity. In a similar context, neurological disorders have been shown to leave the pathophysiological signature on brain networks. Network neuroscience, in this regard, has significantly matured and been extensively utilized in studying neurological disorders; however, the pace of brain network understanding and its role in driving neurological conditions presentation in children falls behind that of the older age population.

We discussed inherent challenges in the face of studying neurological disorders in the pediatric population, contrasting it to the task of hitting a moving target. Brain plasticity is a complex and heterogeneous phenomenon, reflected as a multi-faceted maturation. The pace of brain development process variation occurs at a relatively higher rate in the early stages. Overlaying such a dynamic with uncertainties imposed by neurological disease and disorders turns the task of studying pediatric neurological studies into a difficult one. Inspired by the way our brain sees and learns the world, computational advancement gave birth to a new horizon, called deep learning, promising revolutionary insights into the field of medicine and biology.

Our scoping review of the current state of deep learning utilization in children's neurological disorders has identified a rising interest from scientific investigators for deep learning in various classification tasks related to children's neurological disorders. Data volume, along with, to a lesser extent, data quality, remains among major barriers to incorporate deep learning for studying neurological disorders (Miotto et al., 2018; Valliani et al., 2019). The emergence of deep learning will further continue in the era of pediatric clinical studies because of its lesser reliance upon the existence of engineered features.

Aiming at the development of the generalized deep learner to the level of clinical efficacy requires taking a steeper, and yet clear, road. It involves conducting more extensive multi-sites validation studies and performing benchmark evaluations of deep learning architectures to intensify their utilization further. Studying less utilized architectures such as recurrent neural networks and auto-encoders to learn the joint temporal dynamics of the disease and development in one shot will form a valuable complement to the existing efforts. Legal, privacy, and ethical challenges, with respect to the use of deep learners in healthcare (Miotto et al., 2018; Valliani et al., 2019), remain at place with greater emphasis for the vulnerable populations such as pediatrics. Efforts such as distributed deep learning schemes (Zeng et al., 2018; Remedios et al., 2019) can partially help resolving such dilemma. While deep learning is reported to outperform traditional machine learning algorithms, joined solutions need to be evaluated in cases where deep learners may supplement the established solutions.

Like hitting a moving target, studying neurological disorders in the developing brain is a cumbersome task. However, as tracking a moving target before hitting it can increase success, performing multi time-points longitudinal studies in the pediatric population can improve our chances of understanding and defeating the disease. Therefore, longitudinal modeling and analysis of children's neurological disorders using deep learning architectures can unravel windows of opportunities to hit the moving target.

AUTHOR CONTRIBUTIONS

SS contributed to all stages of study, contributed to manuscript revision, read, and approved the submitted version.

REFERENCES

Aghdam, M. A., Sharifi, A., and Pedram, M. M. (2018). Combination of rs-fMRI and sMRI data to discriminate autism spectrum disorders in young children using deep belief network. *J. Digit. Imaging* 31, 895–903. doi: 10.1007/s10278-018-0093-8

Amado-Caballero, P., Casaseca-de-la Higuera, P., Alberola-Lopez, S., Andres-de Llano, J. M., Villalobos, J. A. L., Garmendia-Leiza, J. R., et al. (2020). Objective ADHD diagnosis using convolutional neural networks over daily-life activity records. *IEEE J. Biomed. Health Inform*. 24, 2690–2700. doi: 10.1109/JBHI.2020.2964072

Ansari, A. H., Cherian, P. J., Caicedo, A., Naulaers, G., De Vos, M., and Van Huffel, S. (2019). Neonatal seizure detection using deep convolutional neural networks. *Int. J. Neural Syst*. 29:1850011. doi: 10.1142/S0129065718500119

Aristizabal, D. A., Fernando, T., Denman, S., Robinson, J. E., Sridharan, S., Johnston, P. J., et al. (2020). Identification of children at risk of schizophrenia via deep learning and EEG responses. *IEEE J. Biomed. Health Inform*. 25, 69–76. doi: 10.1109/JBHI.2020.2984238

Bahado-Singh, R. O., Vishweswaraiah, S., Aydas, B., Mishra, N. K., Guda, C., and Radhakrishna, U. (2019a). Deep learning/artificial intelligence and blood-based dna epigenomic prediction of cerebral palsy. *Int. J. Mol. Sci*. 20:2075. doi: 10.3390/ijms20092075

Bahado-Singh, R. O., Vishweswaraiah, S., Aydas, B., Mishra, N. K., Yilmaz, A., Guda, C., et al. (2019b). Artificial intelligence analysis of newborn leucocyte epigenomic markers for the prediction of autism. *Brain Res*. 1724:146457. doi: 10.1016/j.brainres.2019.146457

Basser, P. J., Mattiello, J., and LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophys. J*. 66, 259–267. doi: 10.1016/S0006-3495(94)80775-1

Bernardo, D., Nariai, H., Hussain, S. A., Sankar, R., Salamon, N., Krueger, D. A., et al. (2018). Visual and semi-automatic non-invasive detection of interictal fast ripples: a potential biomarker of epilepsy in children with tuberous sclerosis complex. *Clin. Neurophysiol*. 129, 1458–1466. doi: 10.1016/j.clinph.2018.03.010

Boldrini, L., Bibault, J.-E., Masciocchi, C., Shen, Y., and Bittner, M.-I. (2019). Deep learning: a review for the radiation oncologist. *Front. Oncol*. 9:977. doi: 10.3389/fonc.2019.00977

Boshra, R., Ruiter, K. I., DeMatteo, C., Reilly, J. P., and Connolly, J. F. (2019). Neurophysiological correlates of concussion: deep learning for clinical assessment. *Sci. Rep*. 9, 1–10. doi: 10.1038/s41598-019-53751-9

Britton, J. W., Frey, L. C., Hopp, J. L., Korb, P., Koubeissi, M. Z., Lievens, W. E., et al. (2016). *Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants*. Chicago, IL: American Epilepsy Society.

Ceschin, R., Zahner, A., Reynolds, W., Gaesser, J., Zuccoli, G., Lo, C. W., et al. (2018). A computational framework for the detection of subcortical brain dysmaturation in neonatal MRI using 3d convolutional neural networks. *Neuroimage* 178, 183–197. doi: 10.1016/j.neuroimage.2018.05.049

Chen, C., Qin, C., Qiu, H., Tarroni, G., Duan, J., Bai, W., et al. (2020). Deep learning for cardiac image segmentation: a review. *Front. Cardiovasc. Med*. 7:25. doi: 10.3389/fcvm.2020.00025

Chen, H., Song, Y., and Li, X. (2019). Use of deep learning to detect personalized spatial-frequency abnormalities in EEGs of children with ADHD. *J. Neural Eng*. 16:066046. doi: 10.1088/1741-2552/ab3a0a

Cohen, D. (1968). Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents. *Science* 161, 784–786. doi: 10.1126/science.161.3843.784

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res*. 29, 162–173. doi: 10.1006/cbmr.1996.0014

Crist, R. E., Li, W., and Gilbert, C. D. (2001). Learning to see: experience and attention in primary visual cortex. *Nat. Neurosci*. 4, 519–525. doi: 10.1038/87470

Daoud, H., and Bayoumi, M. A. (2019). Efficient epileptic seizure prediction based on deep learning. *IEEE Trans. Biomed. Circ. Syst*. 13, 804–813. doi: 10.1109/TBCAS.2019.2929053

David, L., Arús-Pous, J., Karlsson, J., Engkvist, O., Bjerrum, E. J., Kogej, T., et al. (2019). Applications of deep-learning in exploiting large-scale and heterogeneous compound data in industrial pharmaceutical research. *Front. Pharmacol*. 10:1303. doi: 10.3389/fphar.2019.01303

Dou, H., Karimi, D., Rollins, C. K., Ortinau, C. M., Vasung, L., Velasco-Annis, C., et al. (2020). A deep attentive convolutional neural network for automatic cortical plate segmentation in fetal MRI. *arXiv preprint arXiv:2004.12847*. doi: 10.1109/TMI.2020.3046579

Eni, M., Dinstein, I., Ilan, M., Menashe, I., Meiri, G., and Zigel, Y. (2020). Estimating autism severity in young children from speech signals using a deep neural network. *IEEE Access* 8, 139489–139500. doi: 10.1109/ACCESS.2020.3012532

Fischl, B. (2012). Freesurfer. *Neuroimage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021

Ghafouri-Fard, S., Taheri, M., Omrani, M. D., Daaee, A., Mohammad-Rahimi, H., and Kazazi, H. (2019). Application of single-nucleotide polymorphisms in the diagnosis of autism spectrum disorders: a preliminary study with artificial neural networks. *J. Mol. Neurosci*. 68, 515–521. doi: 10.1007/s12031-019-01311-1

Gibb, R., and Kovalchuk, A. (2018). "Chapter 1: brain development," in *The Neurobiology of Brain and Behavioral Development*, eds R. Gibb, and B. Kolb (Academic Press), 3–27. doi: 10.1016/B978-0-12-804036-2.00001-7

Girault, J. B., Munsell, B. C., Puechmaille, D., Goldman, B. D., Prieto, J. C., Styner, M., et al. (2019). White matter connectomes at birth accurately predict cognitive abilities at age 2. *Neuroimage* 192, 145–155. doi: 10.1016/j.neuroimage.2019.02.060

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press. Available online at: http://www.deeplearningbook.org

Greene, D. J., Black, K. J., and Schlaggar, B. L. (2016). Considerations for MRI study design and implementation in pediatric and clinical populations. *Dev. Cogn. Neurosci*. 18, 101–112. doi: 10.1016/j.dcn.2015.12.005

Guo, Y., Wu, G., Commander, L. A., Szary, S., Jewells, V., Lin, W., et al. (2014). "Segmenting hippocampus from infant brains by sparse patch matching with deep-learned features," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Boston, MA: Springer), 308-315. doi: 10.1007/978-3-319-10470-6_39

Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., et al. (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature* 542, 348–351. doi: 10.1038/nature21369

Hua, J. Y., and Smith, S. J. (2004). Neural activity and the dynamics of central nervous system development. *Nat. Neurosci*. 7, 327–332. doi: 10.1038/nn1218

Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol*. 148:574. doi: 10.1113/jphysiol.1959.sp006308

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Kawahara, J., Brown, C. J., Miller, S. P., Booth, B. G., Chau, V., Grunau, R. E., et al. (2017). Brainnetcnn: convolutional neural networks for brain networks; towards predicting neurodevelopment. *Neuroimage* 146, 1038–1049. doi: 10.1016/j.neuroimage.2016.09.046

Kirkwood, A., Lee, H.-K., and Bear, M. F. (1995). Co-regulation of long-term potentiation and experience-dependent synaptic plasticity in visual cortex by age and experience. *Nature* 375, 328–331. doi: 10.1038/375328a0

Kuczmarski, R. J. (2002). *2000 CDC Growth Charts for the United States: Methods and Development*. Number 246. Department of Health and Human Services, Centers for Disease Control.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Lee, M.-H., O'Hara, N., Sonoda, M., Kuroda, N., Juhasz, C., Asano, E., et al. (2020). Novel deep learning network analysis of electrical stimulation mapping-driven diffusion MRI tractography to improve preoperative evaluation of pediatric epilepsy. *IEEE Trans. Biomed. Eng*. 67, 3151–3162. doi: 10.1109/TBME.2020.2977531

Lin, L.-C., Ouyang, C.-S., Wu, R.-C., Yang, R.-C., and Chiang, C.-T. (2020). Alternative diagnosis of epilepsy in children without epileptiform discharges using deep convolutional neural networks. *Int. J. Neural Syst*. 30:1850060. doi: 10.1142/S0129065718500600

Métin, C., and Frost, D. O. (1989). Visual responses of neurons in somatosensory cortex of hamsters with experimentally induced retinal projections to somatosensory thalamus. *Proc. Natl. Acad. Sci. U.S.A*. 86, 357–361. doi: 10.1073/pnas.86.1.357

Miotto, R., Wang, F., Wang, S., Jiang, X., and Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Brief. Bioinform*. 19, 1236–1246. doi: 10.1093/bib/bbx044

Morita, T., Asada, M., and Naito, E. (2016). Contribution of neuroimaging studies to understanding development of human cognitive brain functions. *Front. Hum. Neurosci*. 10:464. doi: 10.3389/fnhum.2016.00464

Mostapha, M., and Styner, M. (2019). Role of deep learning in infant brain MRI analysis. *Magnet. Reson. Imaging* 64, 171–189. doi: 10.1016/j.MRI.2019.06.009

Muñoz-Organero, M., Powell, L., Heller, B., Harpin, V., and Parker, J. (2018). Automatic extraction and detection of characteristic movement patterns in children with ADHD based on a convolutional neural network (CNN) and acceleration images. *Sensors* 18:3924. doi: 10.3390/s18113924

Muñoz-Organero, M., Powell, L., Heller, B., Harpin, V., and Parker, J. (2019). Using recurrent neural networks to compare movement patterns in ADHD and normally developing children based on acceleration signals from the wrist and ankle. *Sensors* 19:2935. doi: 10.3390/s19132935

Ogawa, S., Lee, T.-M., Nayak, A. S., and Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magnet. Reson. Med*. 14, 68–78. doi: 10.1002/mrm.1910140108

O'Shea, A., Lightbody, G., Boylan, G., and Temko, A. (2020). Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture. *Neural Netw*. 123, 12–25. doi: 10.1016/j.neunet.2019.11.023

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., and Nichols, T. E. (2011). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier.

Quon, J., Bala, W., Chen, L., Wright, J., Kim, L., Han, M., et al. (2020). Deep learning for pediatric posterior fossa tumor detection and classification: a multi-institutional study. *Am. J. Neuroradiol*. 41, 1718–1725. doi: 10.3174/ajnr.A6704

Raichle, M. E. (2009). A brief history of human brain mapping. *Trends Neurosci*. 32, 118–126. doi: 10.1016/j.tins.2008.11.001

Raschle, N., Zuk, J., Ortiz-Mantilla, S., Sliva, D. D., Franceschi, A., Grant, P. E., et al. (2012). Pediatric neuroimaging in early childhood and infancy: challenges and practical guidelines. *Ann. N. Y. Acad. Sci*. 1252:43. doi: 10.1111/j.1749-6632.2012.06457.x

Remedios, S., Roy, S., Blaber, J., Bermudez, C., Nath, V., Patel, M. B., et al. (2019). "Distributed deep learning for robust multi-site segmentation of CT imaging after traumatic brain injury," in *Medical Imaging 2019: Image Processing* (International Society for Optics and Photonics), 109490A. doi: 10.1117/12.2511997

Robinson, L. R., Bitsko, R. H., Thompson, R. A., Dworkin, P. H., McCabe, M. A., Peacock, G., et al. (2017). CDC grand rounds: addressing health disparities in early childhood. *Morbidity Mortality Weekly Report* 66:769. doi: 10.15585/mmwr.mm6629a1

Roe, A. W., Pallas, S. L., Kwon, Y. H., and Sur, M. (1992). Visual projections routed to the auditory pathway in ferrets: receptive fields of visual neurons in primary auditory cortex. *J. Neurosci*. 12, 3651–3664. doi: 10.1523/JNEUROSCI.12-09-03651.1992

Rosenzweig, M. R., and Bennett, E. L. (1996). Psychobiology of plasticity: effects of training and experience on brain and behavior. *Behav. Brain Res*. 78, 57–65. doi: 10.1016/0166-4328(95)00216-2

Sánchez Fernández, I., Yang, E., Calvachi, P., Amengual-Gual, M., Wu, J. Y., Krueger, D., et al. (2020). Deep learning in rare disease. Detection of tubers in tuberous sclerosis complex. *PLoS ONE* 15:e0232376. doi: 10.1371/journal.pone.0232376

Santosh, P. J. (2000). Neuroimaging in child and adolescent psychiatric disorders. *Arch. Dis. Childhood* 82, 412–419. doi: 10.1136/adc.82.5.412

Sargolzaei, S., Sargolzaei, A., Cabrerizo, M., Chen, G., Goryawala, M., Pinzon-Ardila, A., et al. (2015). Estimating intracranial volume in brain research: an evaluation of methods. *Neuroinformatics* 13, 427–441. doi: 10.1007/s12021-015-9266-5

Schlund, M. W., Cataldo, M. F., Siegle, G. J., Ladouceur, C. D., Silk, J. S., Forbes, E. E., et al. (2011). Pediatric functional magnetic resonance neuroimaging: tactics for encouraging task compliance. *Behav. Brain Funct*. 7, 1–10. doi: 10.1186/1744-9081-7-10

Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Netw*. 61, 85–117. doi: 10.1016/j.neunet.2014.09.003

Serag, A., Ion-Margineanu, A., Qureshi, H., McMillan, R., Saint Martin, M.-J., Diamond, J., et al. (2019). Translational AI and deep learning in diagnostic pathology. *Front. Med*. 6:185. doi: 10.3389/fmed.2019.00185

Simons, D. J., and Land, P. W. (1987). Early experience of tactile stimulation influences organization of somatic sensory cortex. *Nature* 326, 694–697. doi: 10.1038/326694a0

Stiles, J., and Jernigan, T. L. (2010). The basics of brain development. *Neuropsychol. Rev*. 20, 327–348. doi: 10.1007/s11065-010-9148-4

Sun, L., Zhang, D., Lian, C., Wang, L., Wu, Z., Shao, W., et al. (2019). Topological correction of infant white matter surfaces using anatomically constrained convolutional neural network. *Neuroimage* 198, 114–124. doi: 10.1016/j.neuroimage.2019.05.037

Swanson, L. W., Grant, G., Hökfelt, T., Jones, E. G., and Morrison, J. H. (2007). A century of neuroscience discovery: reflecting on the nobel prize awarded to golgi and cajal in 1906. *Brain Res. Rev*. 55, 191–192. doi: 10.1016/j.brainresrev.2007.07.001

Tang, B., Pan, Z., Yin, K., and Khateeb, A. (2019). Recent advances of deep learning in bioinformatics and computational biology. *Front. Genet*. 10:214. doi: 10.3389/fgene.2019.00214

Thieba, C., Frayne, A., Walton, M., Mah, A., Benischek, A., Dewey, D., et al. (2018). Factors associated with successful MRI scanning in unsedated young children. *Front. Pediatr*. 6:146. doi: 10.3389/fped.2018.00146

Trachtenberg, J. T., Chen, B. E., Knott, G. W., Feng, G., Sanes, J. R., Welker, E., et al. (2002). Long-term *in vivo* imaging of experience-dependent synaptic plasticity in adult cortex. *Nature* 420, 788–794. doi: 10.1038/nature01273

Valliani, A. A.-A., Ranti, D., and Oermann, E. K. (2019). Deep learning and neurology: a systematic review. *Neurol. Therapy* 8, 351–365. doi: 10.1007/s40120-019-00153-8

Wang, C., Xu, P., Zhang, L., Huang, J., Zhu, K., and Luo, C. (2018). Current strategies and applications for precision drug design. *Front. Pharmacol*. 9:787. doi: 10.3389/fphar.2018.00787

Wang, M., Zhang, Q., Lam, S., Cai, J., and Yang, R. (2020). A review on application of deep learning algorithms in external beam radiotherapy automated treatment planning. *Front. Oncol*. 10:580919. doi: 10.3389/fonc.2020.580919

Xu, H., Dong, M., Lee, M.-H., O'Hara, N., Asano, E., and Jeong, J.-W. (2019). Objective detection of eloquent axonal pathways to minimize postoperative deficits in pediatric epilepsy surgery using diffusion tractography and convolutional neural networks. *IEEE Trans. Med. Imaging* 38, 1910–1922. doi: 10.1109/TMI.2019.2902073

Zeng, L.-L., Wang, H., Hu, P., Yang, B., Pu, W., Shen, H., et al. (2018). Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI. *EBiomedicine* 30, 74–85. doi: 10.1016/j.ebiom.2018.03.017

Zhang, L., Wang, M., Liu, M., and Zhang, D. (2020). A survey on deep learning for neuroimaging-based brain disorder analysis. *Front. Neurosci*. 14:779. doi: 10.3389/fnins.2020.00779

Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., et al. (2015). Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *Neuroimage* 108, 214–224. doi: 10.1016/j.neuroimage.2014.12.061

Zhu, G., Jiang, B., Tong, L., Xie, Y., Zaharchuk, G., and Wintermark, M. (2019). Applications of deep learning to neuro-imaging techniques. *Front. Neurol*. 10:869. doi: 10.3389/fneur.2019.00869

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.

frontiers
in Neuroinformatics

ORIGINAL RESEARCH
published: 28 May 2021
doi: 10.3389/fninf.2021.663592

CoMB-Deep: Composite Deep Learning-Based Pipeline for Classifying Childhood Medulloblastoma and Its Classes

*Omneya Attallah**

Department of Electronics and Communications Engineering, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt

Edited by:
Fahimeh Mamashli,
Massachusetts General Hospital and
Harvard Medical School,
United States

Reviewed by:
Kaisu Lankinen,
Massachusetts General Hospital and
Harvard Medical School,
United States
Elaheh Hatamimajoumerd,
Amherst College, United States

***Correspondence:**
Omneya Attallah
o.attallah@aast.edu

Received: *08 February 2021*
Accepted: *26 April 2021*
Published: *28 May 2021*

Citation:
Attallah O (2021) CoMB-Deep:
Composite Deep Learning-Based
Pipeline for Classifying Childhood
Medulloblastoma and Its Classes.
Front. Neuroinform. 15:663592.
doi: 10.3389/fninf.2021.663592

Childhood medulloblastoma (MB) is a threatening malignant tumor affecting children all over the globe. It is believed to be the foremost common pediatric brain tumor causing death. Early and accurate classification of childhood MB and its classes are of great importance to help doctors choose the suitable treatment and observation plan, avoid tumor progression, and lower death rates. The current gold standard for diagnosing MB is the histopathology of biopsy samples. However, manual analysis of such images is complicated, costly, time-consuming, and highly dependent on the expertise and skills of pathologists, which might cause inaccurate results. This study aims to introduce a reliable computer-assisted pipeline called CoMB-Deep to automatically classify MB and its classes with high accuracy from histopathological images. This key challenge of the study is the lack of childhood MB datasets, especially its four categories (defined by the WHO) and the inadequate related studies. All relevant works were based on either deep learning (DL) or textural analysis feature extractions. Also, such studies employed distinct features to accomplish the classification procedure. Besides, most of them only extracted spatial features. Nevertheless, CoMB-Deep blends the advantages of textural analysis feature extraction techniques and DL approaches. The CoMB-Deep consists of a composite of DL techniques. Initially, it extracts deep spatial features from 10 convolutional neural networks (CNNs). It then performs a feature fusion step using discrete wavelet transform (DWT), a texture analysis method capable of reducing the dimension of fused features. Next, the CoMB-Deep explores the best combination of fused features, enhancing the performance of the classification process using two search strategies. Afterward, it employs two feature selection techniques on the fused feature sets selected in the previous step. A bi-directional long-short term memory (Bi-LSTM) network; a DL-based approach that is utilized for the classification phase. CoMB-Deep maintains two classification categories: binary category for distinguishing between the abnormal and normal cases and multi-class category to identify the subclasses of MB. The results of the CoMB-Deep for both classification categories prove that it is reliable. The results also indicate that the feature sets selected using both search strategies have enhanced the performance of Bi-LSTM compared to individual spatial deep features. CoMB-Deep is compared to related studies to verify its competitiveness, and this comparison confirmed its robustness and outperformance. Hence, CoMB-Deep can help pathologists perform accurate diagnoses, reduce misdiagnosis risks that could occur with manual diagnosis, accelerate the classification procedure, and decrease diagnosis costs.

Keywords: childhood medulloblastoma, histopathology, computer-aided diagnosis, convolutional neural network, long short term memory

INTRODUCTION

Childhood brain tumors are the most common cancerous tumors among children accounting for nearly 25% of all pediatric tumors (Pollack and Jakacki, 2011; Pollack et al., 2019). They are considered the second primary rationale of death among children under 15 (Ailion et al., 2017). Among pediatric brain tumors, childhood medulloblastoma (MB) is believed to be the leading cancerous brain tumor among children (Iv et al., 2019). MB accounts for around 15–20% of the central nervous system tumors that affect pediatrics worldwide (Arseni and Ciurea, 1981; Polednak and Flannery, 1995). It evolves within the cerebellum on the posterior area of the brain and progresses rapidly to other parts of the brain (Hovestadt et al., 2020). According to the classification of WHO, there are four MB classes (Pickles et al., 2018). Precise and early diagnosis of MB and its classes is crucial to decide the appropriate treatment and follow-up procedure. This procedure will correspondingly lead to higher survival rates (Davis et al., 1998; Furata et al., 1998), slower disease progression, and avoid acute side effects that could occur if not diagnosed and treated in early stages. These side effects would affect children's movements and synchronization and reduce their quality of life.

Amid imaging modalities, MRI is used in the diagnosis of pediatric brain tumors (Manias et al., 2017; Iqbal et al., 2018). Radiologists experience some challenges, especially when diagnosing pediatric brain tumors (Sachdeva et al., 2013; Ritzmann and Grundy, 2018), like childhood MB and classifying its four classes (Fan et al., 2019). The reason is that several MB subtypes cannot show apparent dissimilarities within the visual appearance of MRI scans (Fetit et al., 2018) which could lead to misdiagnosis (Zarinabad et al., 2018). Thus, other imaging techniques, such as histopathology can be more appropriate for diagnosing MB (Grist et al., 2020). Histopathological examination of biopsy samples is presently thought to be the gold standard to accurately diagnose MB and its classes (Szalontay and Khakoo, 2020). The four MB categories described by the WHO are the classic MB, nodular MB, desmoplastic MB, and anaplastic/large cell MB. Classifying these four classes is essential (Ellison, 2010); however, minimal studies have examined the four childhood MB classes (Dasgupta and Gupta, 2019). Analysis of histopathological images manually is challenging due to the complexity and similarity among childhood MB classes and the small cell size and shape, and the correlation and orientation discrepancy of the different MB tumor classes. Besides, manual analysis wastes time. It also depends mainly on the experience and skills of the diagnostician (Dong et al., 2020). Even expert pathologists could provide distinct diagnoses concerning the MB class (Zhang et al., 2019; Anwar et al., 2020). Another challenge in the manual analysis is the lack of pathologists, especially in the developed and developing countries (Robboy et al., 2013). These challenges have raised the essential need for automated computer-assisted-based pipelines to decrease the burden of the pathologists in classifying the classes of childhood MB and assisting them in achieving high accuracy rates of classification (Das et al., 2018a).

Recently, computer-assisted methods have been extensively used to resolve many related medical problems, and they successfully achieved superior classification results (Ragab et al., 2019; Yanase and Triantaphyllou, 2019; Fujita, 2020; Kumar et al., 2020). However, fewer research articles have been conducted to classify the four childhood MB classes from histopathological images *via* computer-assisted-based approaches because of the shortage of available datasets. The previously acquired datasets are private or publicly available for classifying anaplastic MB and non-anaplastic, which is a binary classification problem. Therefore, most of the previous related studies performed binary classification to differentiate between anaplastic MB and non-anaplastic. For example, in 2011, Lai et al. (2011) presented a pipeline based on Haar wavelets, Haralick Laws texture features, and random forest (RF) classifier and attained a 91% accuracy. Similarly, Galaro et al. (2011) proposed a computer-assisted system based on Haar and MR8 wavelets and achieved an accuracy of 80% using the k-nearest neighbor (k-NN) classifier. The following year, Cruz-Roa et al. (2012) examined the influence of the bag of features histograms method constructed from Haar wavelet transform and visual latent semantic feature extraction methods on the performance of the k-NN classifier. The authors found that the bag of features using Haar wavelet transform is better and achieved an accuracy of 87%. Later, Cruz-Roa et al. (2015) made a comparative study to compare the performance of several feature extraction methods, such as sparse auto-encoder, topographic independent component analysis (TICA), and 3-layered convolutional neural network (CNN). The uppermost accuracy attained was 97% using the TICA method. In the same year, Otálora et al. (2015) merged the TICA with wavelet transform and utilized a 2-layer CNN for classification, attaining an accuracy of 97%.

The previous techniques endured some limitations. Initially, they were all based on traditional feature extraction methods with several parameters to be chosen manually, which increased the time needed before classification. These conventional methods may also be prone to error and are not suitable for all classification problems as they depend on user experience and skill to select the appropriate technique and adjust its parameters

(Hssayeni et al., 2017; Basaia et al., 2019). For example, in the gray level co-occurrence matrix (GLCM) texture-based method, choosing the appropriate distance, *d* is essential. It should be of sufficient size to involve all texture patterns and at the same time, maintain the regional nature of the spatial dependence (Humeau-Heurtier, 2019).

Another example is the local energy pattern feature extraction technique, where the features extracted using such a method are invariant to the imaging settings (Zhang et al., 2012). Some of the previous approaches also rely on the textural features alone, which might fail to describe different dataset patterns (Hira and Gillies, 2015; Babu et al., 2017). For example, when images are noisy, the GLCM method cannot extract significant features from images distinguishing among patterns (or classes) (Humeau-Heurtier, 2019). Moreover, almost all these methods depend only on a single form of feature extraction to construct their classification model. These methods did not examine the impact of fusing several feature extraction methods on the accuracy of classification. Also, they are all constructed using a small dataset compromising on only 10 images. Lastly, they all perform binary classification to differentiate between non-anaplastic and analyptic (binary classification problem). On the other hand, multi-class category is essential to figure out the subclasses of childhood MB and decide the appropriate treatment and observation strategy according to each class.

Multi-class classification of the four classes of childhood MB is much more complicated than binary classification. Few research articles have investigated this multi-class classification problem. To the best of our knowledge, the first and only research group that investigated the classification of the four MB subtypes from histopathological images using machine and deep learning (DL) techniques were studies by Das et al. (2018b) who presented a diagnostic framework to diagnose the subclasses of such pediatric tumor. The authors initially segmented images using k-means clustering. Afterward, the authors pulled out morphological and color features. They extracted textural features, including GLCM, histogram of oriented gradients (HOG), and Tamura, in addition to the local binary pattern (LBP) and gray level run matrix (GLRM). Next, they reduced the feature space using principal component analysis (PCA) and used them to construct a support vector machine (SVM) classifier. They attained an accuracy of 84.9%. Later in 2020, the same authors utilized the same features but used multivariate analysis of variance (MANOVA) as a feature reduction method. They found out that the MANOVA method increased the classification accuracy to 65.2%, which is greater than that of 56.5% without MANOVA. In the same year, Das et al. (2020b) decided that instead of using individual sets of features (Das et al., 2018b), they produced various groups of fused features to study the influence of feature fusion and selected the best mixture of fused feature sets. They employed PCA and SVM and achieved an accuracy of 96.7% utilizing the four combined features. Later, in the same year, Das et al. (2020c) used two pre-trained CNN, including AlexNet and VGG-16, with transfer learning. They first used a soft-max classifier for classification and reached an accuracy of 79.3 and 65.4% for AlexNet and VGG-16 CNNs, respectively. They extracted deep features from these networks and employed an SVM classifier for classification, achieving an accuracy of 93.21 and 93.38% for AlexNet and VGG-16 features, respectively. Afterward, Attallah (2021) introduced a framework based on the deep features of three CNNs combined with PCA and discrete cosine transform and classified using four machine learning classifiers.

These previous methods have shown several drawbacks that can be summarized as follows: the approaches by Das et al. (2018b, 2020a) utilized traditional individual feature extraction methods that depend on textural analysis, color, or morphological operations. Drawbacks of conventional methods are discussed earlier. The disadvantages of using color features are their vulnerability to light situations and occlusion (Afifi and Ashour, 2012; Park et al., 2012). The limitations of morphological features like shape features are their dependence on the human experience, and they might not produce efficient results when used alone for classification (Liu and Shi, 2011). Furthermore, the pipeline presented by Das et al. (2020b) examined the combination of textural-based features alone to perform classification. Das et al. (2020c) utilized only DL features to build the classification model. In other words, the authors employed only individual DL features for classifying the MB subtypes, where every one of them was of enormous feature dimension. The authors did not examine the influence of fusing numerous DL features extracted from different DL approaches to take advantage of different DL architectures. Additionally, Das et al. (2020c) employed only two CNNs separately. The training time achieved utilizing these DL methods is enormous. Finally, some of them did not attain superior accuracy and cannot be considered as reliable systems. A reliable system means that it has a sensitivity that exceeds 80%, a specificity that exceeds 95%, and a precision that exceeds 95% (Ellis, 2010; Colquhoun, 2014). So, these systems did not achieve the criterion to be a reliable system.

The originality of the study may be listed in the following four contributions. First, a reliable pipeline called CoMB-Deep is built to classify the four classes of childhood MB with high accuracy. Second, CoMB-Deep is based on a composite of DL methods that are merged. Third, it blends the advantages of textural-based feature extraction and DL approaches using three stages. Initially, 10 CNN DL methods are utilized to mine deep spatial features. These deep spatial features are fused using discrete wavelet transform (DWT), a textural analysis-based approach capable of reducing the dimension of DL features. Afterward, the best combination of fused features is searched using two strategies to choose the appropriate set of fused features that influence the classification performance. Next, CoMB-Deep employs the bidirectional long short-term memory (Bi-LSTM) DL method, which illustrates the temporal information of the data, and not the spatial information like the CNNs used in the literature. The fourth contribution is the further reduction of the feature space dimension after the process of DWT-based fusion using two feature selection methods. We want to highlight that one of the core difficulties in classifying the childhood MB classes is the obtainability of the dataset.

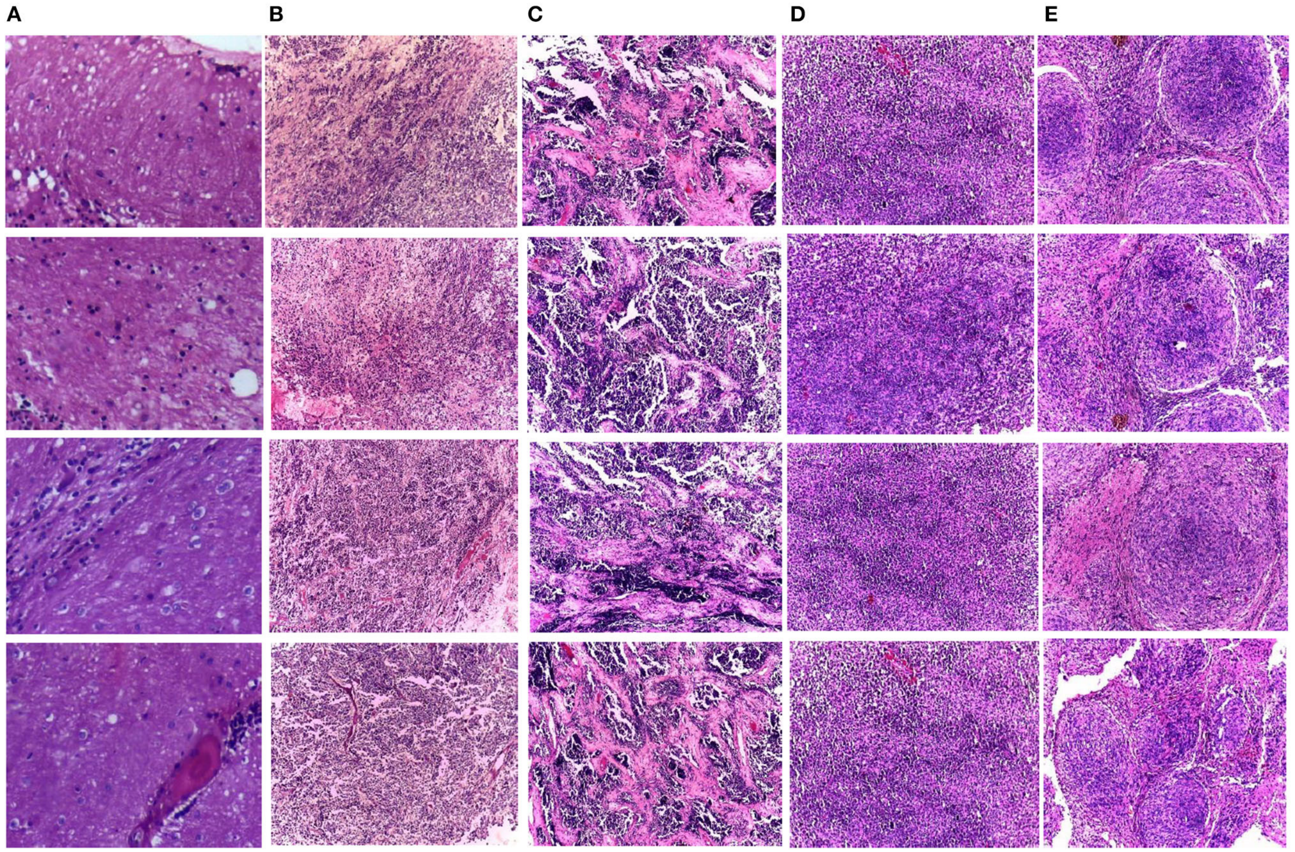

FIGURE 1 | Samples of the childhood pediatric MB images, **(A)** normal, **(B)** classic, **(C)** desmoplastic, **(D)** large cell, **(E)** nodular.

METHODS

Dataset Acquisition

Two medical centers for neurological research, including Guwahati Neurological Research Center (GNRC), and the Guwahati Medical College and Hospital (GMCH), involved collaborating medical institutes to acquire the dataset. The childhood MB dataset employed in CoMB-Deep was gathered from kids diagnosed with MB tumor. These kids were aged <15 years. Some information blocks were generated from those kids at the surgical department of the GMCH. The swabs were acquired through these tissue blocks. Next, Hematoxylin and eosin (HE) were employed to stain these sections of tissues at Ayursundra Pvt. Ltd., in which a local medical specialist supplied pathological help. Later, a certified specialist determined the region of interest for ground truth from such slide scans. Afterward, the images of these areas of interest taken with an amplification factor of 10x with a Leica 1CC50 HD microscope were stored in the Joint Photographic Experts Group (JPEG) format. The four subclasses of MB tumor were included in the dataset, which involved 204 images. Some of these images are displayed in **Figure 1**. The number of samples for normal, classic, desmoplastic, large cell, and nodule MB classes include 50, 59, 42, 30, and 23, respectively. Details of the dataset can be found in the study by Das et al. (2019).

Methods of Deep Learning

During the last decade, a new branch of machine learning techniques called DL has been extensively used in many areas due to its high capability of overcoming the drawbacks of traditional artificial neural networks (Litjens et al., 2017; Attallah et al., 2020b). There are several architectures of neural networks based on DL. Among these architectures is the recurrent neural networks (RNNs), such as long-short term memory which are used for sequential data, and the CNN utilized for both image/video classification (Liu et al., 2017, Ceschin et al., 2018; Alom et al., 2019; Attallah et al., 2020c; Ragab and Attallah, 2020). The CNN demonstrates only spatial information from images; therefore, this type of DL technique is combined with Bi-LSTM, which determines the temporal data from the data given as its input.

Convolution Neural Network Architectures

The fundamental components of CNNs involve convolutional layers, rectified linear unit (Relu) layers, pooling layers, dropouts, fully connected (FC) layers, softmax and output layers (Alom et al., 2019). The organization of these layers may form a new architecture. The convolutional layer convolves a portion of the image with a filter of a small size. The outputs of the filter after passing through the whole picture are known as the feature map. The Relu layer applies the Relu activation function, which alters the entire negative activations to zero. It also enhances the non-linear characteristics of the CNN without influencing the receiving fields of the convolutional layer. The pooling layer is responsible for downsizing the massive dimension of the feature map of the previous layer. The most common types are maximum and average pooling. The dropout layer is used to prevent the network from overfitting by randomly adjusting the output edges of hidden neurons to zero at every training iteration. The FC layers are the last few layers of a CNN where the entire inputs of the previous layer are connected to each neuron in the subsequent layer. The softmax layer is responsible for the classification procedure using the softmax function, assigning probabilities for each class. The output layer produces the result of the network after training (Pouyanfar et al., 2018). Some of the state-of-the-art CNNs are briefly discussed below.

ResNet-50 CNN was proposed by He et al. (2016). The fundamental building block of ResNet is the residual block which involves shortcuts (known as residuals) within the layers of a traditional CNN to step over specific convolution layers at a time. This structure can enhance the performance of the CNN and accelerate the convergence process of the CNN, regardless of many deep convolution layers. ResNet-50 consists of 50 layers. The sizes and names of these layers are illustrated in **Supplementary Table 1**. In 2016, Google proposed the Inception-V3 CNN architecture (Szegedy et al., 2016b). It is based on the inception module that combines several convolutional filters of different capacities into a new filter. This process correspondingly lowered the number of parameters used in the training and the computational time. It consists of 48 layers, including convolutional, pooling, convolutional padded, and fully connected layers. The sizes and names of these layers are illustrated in the **Supplementary Table 2**. DenseNet was proposed by Huang et al. (2017). The main building block of DenseNet is the Dense block that connects every layer to every subsequent layer in the feed-forward procedure. At every layer, the feature maps of the earlier layers' are considered inputs to the current layer. DenseNet-201 consists of 201 layers. Its structure comprises a convolutional layer followed by a Dense block and a transition layer (Huang et al., 2017; Li et al., 2021). The transition layer consists of a convolutional layer followed by a pooling layer. The sizes and names of these layers are illustrated in **Supplementary Table 3**.

MobileNet is a lightweight CNN proposed by Howard et al. (2017). Its structure depends on two layers: depth-wise layers and point-wise layers. The depth-wise layer has numerous convolutional layers of 3 × 3 kernels. The point-wise layer consists of several convolutional layers of 1 × 1 kernels, where it contains 53 deep layers. The dimensions and names of these layers are illustrated in **Supplementary Table 4**. Inception-ResNet-V2 was proposed by Szegedy et al. (2016a), who inserted residual shortcuts in the Inception block of the Inception CNN structure. This network consists of 164 layers. The design of Inception-ResNet-V2 and the output size of layers are shown in the **Supplementary Figure 1**. Xception was introduced in 2017 by Chollet (2017). Its main building block is the Xception module. The Xception module switched the standard Inception modules with depth-wise separable convolution layers. This module begins with two convolutional layers, followed by depth-wise separable convolution layers, four convolution layers, and a fully connected layer without average pooling. The non-overlapping sections of the output channels from these layers are concatenated. Xception CNN has 36 convolutional layers.

NasNetMobile was introduced by Zoph et al. (2018). The main building blocks of NasNetMobile are called cell convolutional, which have a similar structure but different in weight. They are optimized using reinforcement. This network searches for the best architecture for a given dataset. ShuffleNet was proposed in 2018 by Zhang et al. (2018). Its crucial element is the ShuffleNet which delivers two new procedures known as point-wise group convolution and channel shuffle. The point-wise process utilizes a 1 × 1 convolution to decrease the computation time while attaining an adequate accuracy, whereas the channel shuffle procedure helps in streaming information among feature channels. It consists of 50 layers. The architecture of ShuffleNet is shown in the **Supplementary Table 5**. SqueezeNet was proposed by Iandola et al. (2016). The basic building block of SqueezeNet is the fire module. This module has a squeeze convolutional layer (having only a 1 × 1 filter), supplying an expand layer that consists of a mixture of 1 × 1 and 3 × 3 convolutional filters. The layers and their sizes are shown in the **Supplementary Table 6**. DarkNet-53 was initially proposed in 2017 by Redmon and Farhadi (2017). DarkNet-53 primarily consists of a sequence of convolution layers sizes of 1 × 1 and 3 × 3. It usually pairs with the number of channels after each pooling phase. DarkNet-53 has a total number of 53 layers (involving the last fully connected layer but not including the residual layer).

Bidirectional Long Short-Term Memory

The long short-term memory DL architecture (Hochreiter and Schmidhuber, 1997) was inspired by analyzing the stream of error in the RNN (Angeline et al., 1994). The LSTM architecture consists of an input gate, a forget gate, and an output gate. These gates are responsible for recognizing the long-range temporal reliance. The fundamental concept of the Bi-LSTM (Schuster and Paliwal, 1997; Baldi et al., 1999) is to introduce every training cycle forward and backward to two distinct LSTMs, both of which are associated with the same output layer.

Proposed CoMB-Deep Pipeline

This study introduces a computer-assisted pipeline called CoMB-Deep to classify childhood MB subclasses from histopathological scans automatically. The CoMB-Deep consists of a combination of multiple DL methods to classify childhood MB and its subclasses. The classification is accomplished in two categories: binary and multi-class classification categories. The former category categorizes the microscopic scans into abnormal and normal (binary classification category). The latter category classifies the abnormal microscopic scans into four childhood MB tumors (multi-classification category).

CoMB-Deep involves six phases: image preprocessing, deep spatial feature extraction, feature fusion and reduction, fused feature set election, feature selection, and classification phases. During the first phase, the microscopic images are augmented

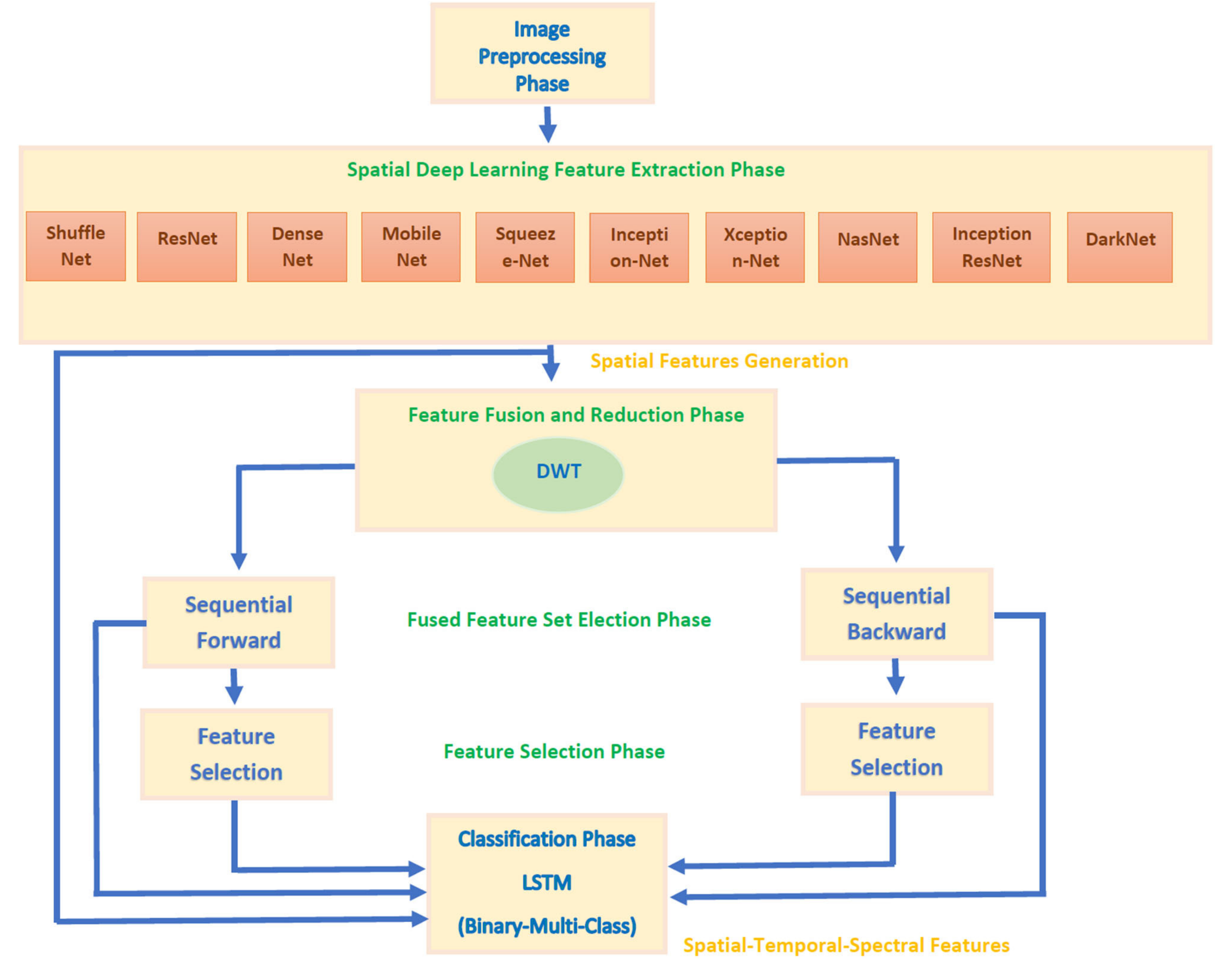

FIGURE 2 | The block diagram of CoMB-Deep pipeline.

and resized. Afterward, spatial DL variables are obtained from 10 pre-trained CNNs in the deep spatial feature extraction phase. Next, these features are fused using the DWT method. DWT is also used in this phase to decrease the enormous size of fused features. Then, in the combined feature set election phase, the feature sets generated from the 10 CNNs are searched using two strategies to detect the most acceptable mixture of fused feature sets that influence the performance of CoMB-Deep. In the feature selection phase, two feature selection approaches are performed on the best-fused sets of features selected in the prior phase to lessen the size of fused features further. Finally, a bidirectional LSTM DL method is used to accomplish the classification procedure of both the binary and multi-class categories. **Figure 2** illustrates the block diagram of the introduced CoMB-Deep.

Image Preprocessing Phase

During this phase, the microscopic images are resized to the input layer size of every CNN as shown in **Table 1**. Next, an augmentation step is made to enlarge the number of microscopic images available and avoid overfitting (Shorten and Khoshgoftaar, 2019). The techniques used for augmentation to create new microscopic scans from the training data in CoMB-Deep are translation (−30, 30), scaling (0.9, 1.1), flipping in x and y directions, shearing (0, 45) in x and y directions as observed by Ragab and Attallah (2020) and Attallah et al. (2020c).

Deep Spatial Feature Extraction Phase

During this phase, spatial DL features are extracted from 10 CNNs architectures using transfer learning. These CNN architectures include Inception V3, Xception, ResNet-50, Inception-ResNet-V2, DenseNet-201, NasNetMobile, ShuffleNet, SqueezeNet, MobileNet, and DarkNet-53 networks. The transfer learning process corresponds to the ability of the network to identify similarities between distinct data, which helps develop the training procedure for a new similar classification task (Thrun and Pratt, 1998; Raghu et al., 2019). In

TABLE 1 | The depth, input, and output sizes of the 10 CNNs along with the description of the layers from which features where extracted.

CNN structure	Depth	Size of input	Feature extraction layer name	Feature extraction layer description	Size of features
Shuffle	50	224 × 224 × 3	"node 200"	The last average pooling layer just before the fully connected layer	544
ResNet-50	50	224 × 224 × 3	"avg_pool"	The last average pooling layer just before the fully connected layer	2,048
NasNetMobile	913	224 × 224 × 3	"global_average_pooling2d_1"	The last average pooling layer just before the fully connected layer	1,056
MobileNet	19	224 × 224 × 3	"avg_pool"	The last average pooling layer just before the fully connected layer	1,280
SqueezeNet	18	227x227x3	"relu_conv10"	Rectified Linear unit layer just after the 10th convolutional layer	392 (Binary) 784(Multi-class)
DenseNet-201	201	224 × 224 × 3	"avg_pool"	The last average pooling layer just before the fully connected layer	1,920
Inception-V3	48	229 × 229 × 3	"avg_pool"	The last average pooling layer just before the fully connected layer	2,048
Xception	71	229 × 229 × 3	"avg_pool"	The last average pooling layer just before the fully connected layer	2,048
Inception-ResNet	164	229 × 229 × 3	"avg_pool"	The last average pooling layer just before the fully connected layer	1,536
DarkNet-53	53	256 × 256 ×3	"avg_1"	The last average pooling layer just before the fully connected layer	1,024

other words, transfer learning enables pre-trained CNN to learn demonstration from a huge number of images, such as those available in ImageNet and afterward used this knowledge in a similar classification problem which has smaller datasets (Pan and Yang, 2009; Han et al., 2018). In this paper, 10 pretrained CNNs that were previously trained on the ImageNet dataset are employed. Transfer learning is also employed to extract deep spatial features from a specific layer of a CNN. In this phase, after adapting the FC layers and some of the parameters (discussed later in the parameter setting section) of the 10 pretrained CNNs to the number of labels in the dataset used in this paper rather than the 1,000 labels of ImageNet, the 10 networks are trained. Afterward, features are extracted from the layers as mentioned in **Table 1**. **Table 1** illustrates the depth, input, and output sizes and mentions the layers where features are extracted.

Feature Fusion and Reduction Phase

Deep spatial features obtained in the earlier phase are fused using DWT. The reason for choosing DWT is that it is a well-known technique based on texture analysis and is commonly used in the medical area. DWT can reduce the vast dimension of the data and demonstrate temporal-frequency representations from any given data (Li and Meng, 2012; Ponnusamy and Sathiamoorthy, 2019). DWT utilizes a group of perpendicular basis functions called wavelets to analyze data (Antonini et al., 1992). In the case of 1-D data, like the deep spatial features mined from the 10 CNNs utilized in CoMB-Deep, the DWT process is performed by passing the data through low and high pass filters (Demirel et al., 2009). Afterward, a downsampling step is accomplished to reduce the data dimension (Hatamimajoumerd and Talebpour, 2019). Two sets of coefficients will be generated after this step: the approximation coefficients CA_1 and detail coefficients CD_1 (Attallah et al., 2019). This phase is executed to consider the privileges of DL and textural-based feature extraction approaches. The discrete Meyer wavelet (dmey) is employed as the mother wavelet. The CD_1 variables are only selected because these coefficients comprise information from most of the data. Furthermore, the large size of the features extracted in the former phase is decreased.

Fused Feature Set Election Phase

Deep spatial variables mined in the deep spatial feature extraction phase are ranked in descending order according to the classification accuracy obtained using the LSTM network. This ranking is used to generate several feature sets that are fused using DWT. In this phase, two searching strategies, including the sequential backward and forward schemes (Attallah, 2020), are done to select the merged feature sets which have the highest impact on the accuracy attained by CoMB-Deep. In the backward scheme, the feature set selection begins with fusing all features using DWT and eliminating the lowest ranking feature (lowest accuracy). Next, each set of features having a better ranking is deleted; if the excluded set of features improved the accuracy of CoMB-Deep, it is kept deleted; else, it is not deleted. Alternatively, in the forward scheme (Fauvel et al., 2015; Pohjalainen et al., 2015), the selection begins with the first feature set having the highest rank. Afterward, every successive set of features is included one by one; if the set of features included increased the accuracy of CoMB-Deep, it is retained; else, it is ignored. This procedure continues until all feature sets are investigated (Ververidis and Kotropoulos, 2005).

Feature Selection Phase

To further reduce the vast dimension of the fused feature sets chosen in the previous phase, feature selection is essential (Attallah et al., 2017b) because the massive size of the features raises the complexity of the classification phase and might decrease its performance (Li and Liu, 2017; Hatamimajoumerd et al., 2020). Feature selection is frequently utilized in medical frameworks to reduce the dimension of the feature set and delete unnecessary and irrelevant variables (Chandrashekar and Sahin, 2014; Attallah et al., 2017a, 2020a; Cai et al., 2018). In this phase, two feature selection approaches, including Relief-F and Information gain (IG) feature selection methods are used to select a reduced set of features.

Information gain is a standard filter feature selection method. The fundamental concept of filter methods is to utilize the overall characteristics of the data to choose the most significant attributes based on these characteristics before the construction process of the classifier. The main benefit of the filter feature selection approach is its fast and straightforward feature (Sánchez-Marono et al., 2007). IG selects relevant features according to the fundamental idea of the entropy by calculating the differences among the entropy of the whole training samples and the weighted total sum of the values of the subset of their partition (classes) available for a given attribute (Roobaert et al., 2006). IG method is considered to be one of the fastest and straightforward filter methods. Therefore, it is utilized in this paper.

Relief-F feature selection is a well-recognized filter feature selection approach commonly used in medical classification problems due to computation efficiency and straightforwardness. The Relief-F method was proposed by Kononenko (1994) to be used for multi-class, noisy, and incomplete problems of the datasets. Its basic idea is to calculate the significance of features based on their capabilities to differentiate among the same class instances that are close and distinct to others in a local neighborhood. This capability is measured by estimating the weight or score of each feature (Urbanowicz et al., 2018). Those features that have a higher ability to distinguish between different class instances and increase the distance between them are given higher scores than others with lower capacities.

Classification Phase

The Bi-LSTM DL-based technique is used in the two classification categories of CoMB-Deep (binary and multi-class). This process is accomplished using four schemes. In the first scheme, spatial DL features are taken out from the 10 CNNs to train several Bi-LSTM classifiers. Afterward, these spatial DL features are ranked according to the accuracy achieved by the Bi-LSTM. This ranking is used in the following two schemes. The order of the previous scheme is employed to explore the most acceptable combination of fused features during the second scheme, enhancing the accuracy of the LSTM classifier using a sequential forward strategy. Note that fusion is done using the DWT method. The third scheme is like the second scheme, but it uses a sequential backward strategy instead of a sequential forward strategy. In the last scheme, two feature selection approaches are used to select a reduced number of features from the chosen fused feature sets in the second and third schemes. **Figure 3** defines the four schemes of the proposed CoMB-Deep.

All CoMB-Deep schemes were done with Matlab 2020a. The Bi-LSTM classifier was constructed using the Weka data mining software (Hall et al., 2009). The processor used is Intel(R) Core (TM) i7-10750H with NVIDIA GeForce GTX 1660 video controller of 6 GB capacity, with the processor frequency of 2.6 GHz 64-bit operating system.

EXPERIMENTAL SETTINGS

Assessment Measures

To assess the performance of the presented CoMB-Deep pipeline, several assessment measures are utilized. These measures include the accuracy, Matthews correlation coefficient (MCC), sensitivity, precision, F1 score, and specificity. They are computed using the following rules (Attallah et al., 2020b) (Equations 1–6).

$$Accuracy = \frac{TP + TN}{TN + FP + FN + TP} \tag{1}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

$$F1 - Score = \frac{2 \times TP}{(2 \times TP) + FP + FN} \tag{6}$$

Where, the number of MB images that are correctly identified to belong to the MB class which they refer to is called true positive (TP) and true negative (TN) is the total number of MB images that does not refer to the identified MB class, and does not actually refer to. For each subtype of MB tumor, false positive (FP) is the sum of all classified images as this MB subtype, but they do not truly belong to. For each subtype of MB tumor, false negative (FN) is the total sum of images not classified as this MB subtype.

Parameters Set Up

Several parameters are modified for the 10 CNNs. The mini-batch dimension and validation frequency are set to 4 and 26 for multi-class and 17 for binary class. The epochs and the initial learning rate are found to be 20 and 3×10^{-4}, respectively, while other parameters of the 10 CNNs are set with their default values. To assess the performance of the classification process and avoid overfitting, a 10-fold cross-validation is applied. It is used in Bi-LSTM classification, and the announced accuracy is found to be average across the 10-cross validation folds. For the Bi-LSTM network, the number of epochs is set to 30, the batch size is set to 100, and the validation frequency is set to 10.

RESULTS

This section will illustrate the four scheme results of the CoMB-Deep pipeline for the binary and multi-class classification categories. Scheme I represents the deep spatial features obtained through the 10 CNNs distinctly as inputs to LSTM networks. The classification accuracies attained using the 10 DL features are utilized to rank them in the descending order. This sorting is then employed in the following two schemes. Scheme II explores the 10 spatial feature sets using a forward strategy to determine the best combined spatial feature sets that improve the performance of CoMB-Deep. Similarly, Scheme III searches for the most acceptable mixture of combined feature sets but using a backward strategy. The combination of the feature sets is made utilizing DWT to reduce the dimension of the fused feature sets. In Scheme IV, two feature selection methods are employed to further decrease the size of the merged feature sets and to select a reduced number of features.

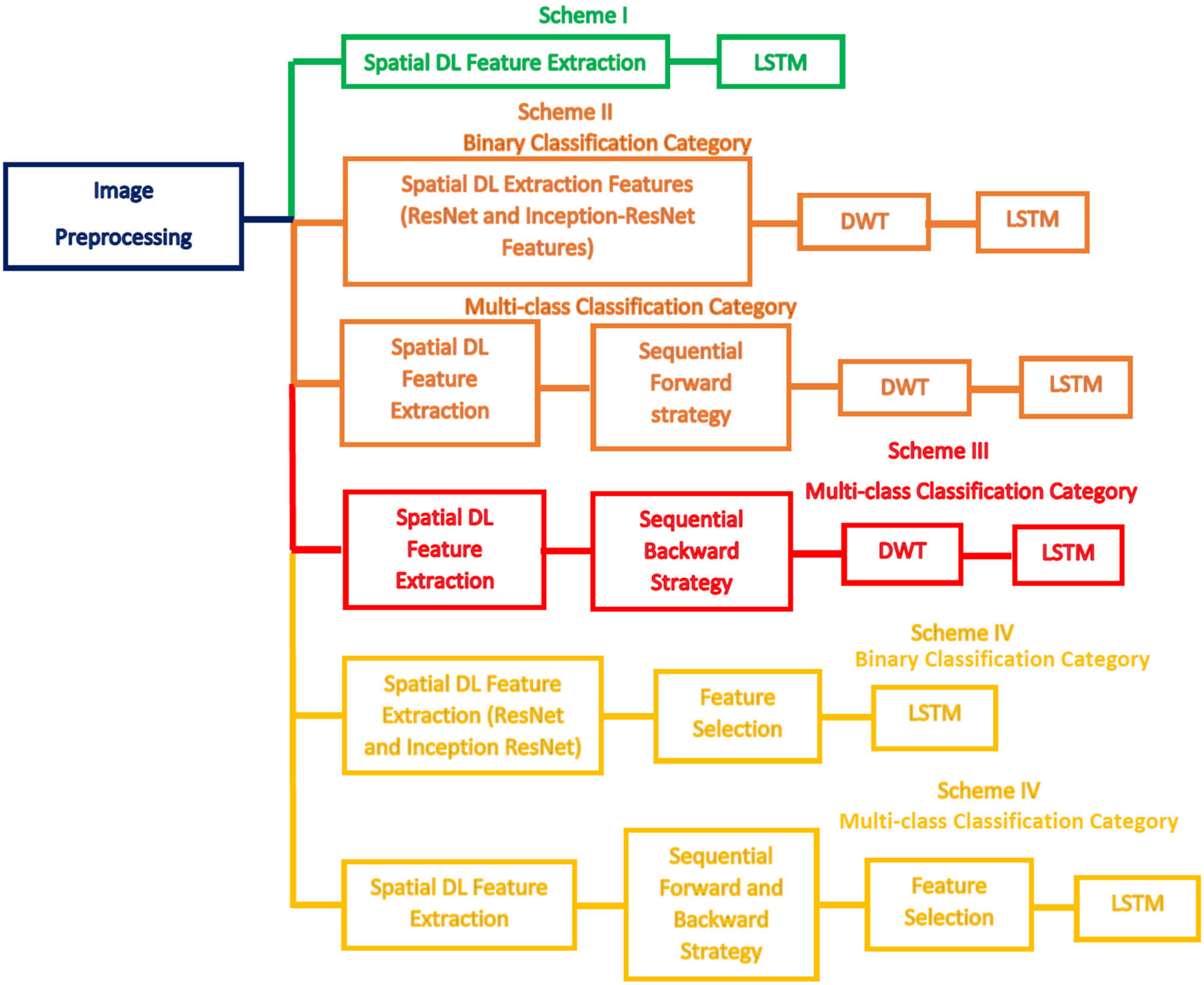
Scheme I
Spatial DL Feature Extraction
LSTM
Scheme II
Binary Classification Category
Image
Preprocessing
Spatial DL Extraction Features (ResNet and Inception-ResNet Features)
DWT
LSTM
Multi-class Classification Category
Spatial DL Feature Extraction
Sequential Forward strategy
DWT
LSTM
Scheme III
Multi-class Classification Category
Spatial DL Feature Extraction
Sequential Backward Strategy
DWT
LSTM
Scheme IV
Binary Classification Category
Spatial DL Feature Extraction (ResNet and Inception ResNet)
Feature Selection
LSTM
Scheme IV
Multi-class Classification Category
Spatial DL Feature Extraction
Sequential Forward and Backward Strategy
Feature Selection
LSTM

FIGURE 3 | The four schemes of CoMB-Deep.

Results of Scheme I

The results of Scheme I are shown in **Table 2**. For the binary classification category, the highest accuracy of 100% is attained using Inception-ResNet-V2 and ResNet-50. This is followed by DarkNet 53, Inception V3, MobileNet, and Squeeze CNNs to obtain an accuracy of 99%. Then, DenseNet-201 and ShuffleNet CNNs accomplish an accuracy of 98%, which is >96 and 92% accuracy obtained using Xception and NasNetMobile CNNs. On the other hand, for the multi-class classification category, the greatest accuracy of 96.1% is achieved using DenseNet-201 and Inception V3 spatial features. Next, ShuffleNet, SqueezeNet, and ResNet-50 features reach an accuracy of 95.45, 92.86, and 92.2%, respectively. Afterward, both Inception-ResNet-V2 and MobileNet features attain an accuracy of 90.91%. Subsequently, DarkNet-53, Xception,

TABLE 2 | The classification accuracies (%) achieved using LSTM network trained with individual DL features for binary and multi-class classification categories.

CNN	Binary classification	Multi-class classification
DarkNet53	99	88.31
DenseNet-201	98	96.1
Inception	99	96.1
InceptionResNet	100	90.91
MobileNet	99	90.91
NasNetMobile	92	75.97
ResNet-50	100	92.2
ShuffleNet	98	95.45
Xception	96	87.66
Squeeze	99	92.86

TABLE 3 | Binary classification accuracy (%) and number of features before and after DWT for ResNet-50 and Inception-ResNet CNNs.

Type of features	No of features	Accuracy
ResNet-50		
Spatial DL	2,048	100
DWT	1,074	100
Inception-ResNet		
Spatial DL	1,536	100
DWT	818	100

and NasNetMobile CNNs obtain accuracies of 88.31, 87.66, and 75.97%.

Results of Scheme II

This section discusses the results of the fusion of multiple DL features using the forward strategy and DWT. Note that in the binary classification category, it can be noticed that both the Inception-ResNet-V2 and ResNet-50 attain a 100% accuracy in Scheme I; thus, there is no need for feature fusion in the binary classification category done in Schemes II and III. The DWT in this category is used to reduce the dimension of features extracted from both Inception-ResNet-V2 and ResNet-50. **Table 3** shows the accuracy and size of features after and before the DWT for both Inception-ResNet-V2 and ResNet-50 CNNs. It is clear from **Table 3** that the DWT has reduced the number of features from 2,048 to 1,074 for the ResNet-50 CNN and from 1,536 to 818 features for the Inception-ResNet-V2 achieving the same accuracy of 100%.

For the multi-class classification category, the main target of Scheme II is to explore the different combinations of deep spatial features using the forward strategy. The search starts with the feature set 1 corresponding to only one feature type. It starts to add feature sets iteratively; if the feature set included increased the accuracy, then it is kept, else it is ignored. Note that the feature selection processing is performed on the training set separate from the testing set. The feature set selection results on the training set of the multi-class classification category of Scheme II are shown in **Figure 4**. This figure shows that feature set 3, which contains the fused features of DenseNet-201+ShuffleNet after DWT having a dimension of 1,282 features, has attained the highest accuracy of 97.22%. For this reason, it is selected in Scheme II. The results of feature set 3 are then evaluated on a separate testing set, and the results are shown in **Table 4**. **Table 4** indicates that feature set 3 selected using the forward strategy has an accuracy of 95.65%, a sensitivity of 0.957, a specificity of 0.992, a precision of 0.958, an F1-score of 0.946, and an MCC of 0.99.

Results of Scheme III

The results of the backward search strategy for the multi-class classification category are illustrated in this section. Feature set 1 corresponds to the fusion of all the 10 deep features; each feature set is eliminated iteratively; if the accuracy is improved, then it is removed; else, it is kept. The results of the backward strategy executed on the training set are shown in **Figure 5**. It can be noticed from **Figure 5** that feature sets, 3 and 5 achieve the highest accuracy of 97.22%. Feature set 3 has a dimension of 6,170 features representing the fusion of the features from all CNNs except those of Xception, whereas feature set 5 contains 5,402 features of the combined DenseNet-201 + Inception V3 + ResNet-50 + Darknet-53 + MobileNet + ShuffleNet + SqueezeNet + NasNetMobile CNNs. This dimension shows that feature set 5 has a lower size of features than the feature set 3; therefore, it is selected. The testing performance of feature 5 selected during the backward strategy is shown in **Table 4**. **Table 4** illustrates that feature set 5 chosen using backward strategy has an accuracy of 95.65%, a sensitivity of 0.957, a specificity of 0.975, a precision of 0.961, an F1-score of 0.955, and an MCC of 0.94.

Results of Scheme IV

This section illustrates the results of Relief-F and IG feature selection methods for both classification categories of CoMB-Deep. **Table 5** shows the number of features, classification accuracy, and other performance metrics after feature selection for the binary classification category. This table verifies that 100% accuracy is achieved using only 578 features selected using Relief-F and IG methods instead of 1,074 utilized in Scheme II (using DWT) for ResNet-50 CNN. On the other hand, the same accuracy is reached using 200 and 550 features obtained with Relief-F and IG approaches which are lower than the 818 features employed in Scheme II (using DWT) Inception-ResNet-V2 CNN. It is clear from the table that the sensitivities, specificities, precisions, F1-scores, and MCCs are equal to 1.

For the multi-class classification category, it can be noticed from **Table 6** that for the forward search strategy, both Relief-F and IG methods have reduced the number of features from 1,282 features of feature set 3 (Scheme II) to 448 and 738 features, respectively, while attaining the same accuracy of 98.05% which is higher than that obtained in Scheme II. Similarly, in the backward search strategy, the features selected using Relief-F and IG methods have attained an accuracy of 99.35%, which is better than that obtained in Scheme III, where the number of features has been decreased to 739 (for Relief-F), and 2,313 (for IG) which are lower than the 5,403 features of feature set 5 chosen in Scheme III. **Table 6** also indicates that both feature selection methods for the forward strategy achieve a sensitivity of 0.981, precisions of 0.981, and F1 scores of 0.981 and 0.972, respectively. The specificity and MCC attained using Relief-F, and IG methods are 0.993, 0.99, 0.972, and 1. For the backward approach, both feature selection methods achieved sensitivities of 0.994, specificities of 0.996, precisions of 0.984, F1 scores of 0.993, and MCCs of 0.991.

DISCUSSION

This study presented a computer-assisted pipeline called CoMB-Deep for the automatic classification of childhood MB and its classes from histopathological images. CoMB-Deep has two classification categories: binary and multi-class classification categories. CoMB-Deep discriminates between normal and MB images in the binary classification category, whereas the multi-class classification category classifies the four classes of childhood

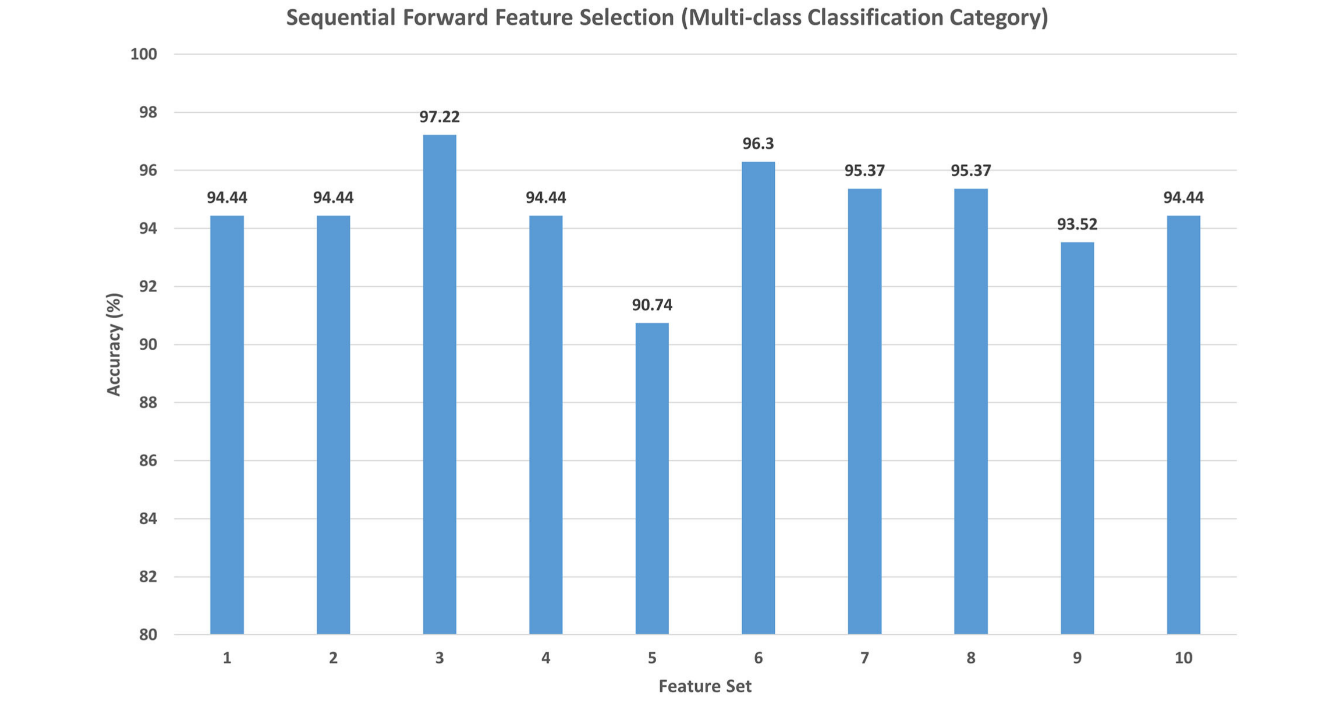

FIGURE 4 | The classification accuracy (%) attained for the different feature sets generated using forward strategy. Feature set 1: DenseNet-201 (1,010 Features); Feature set 2: DenseNet-201 + Inception V3(2,034 Features); Feature set 3: DenseNet-201 + Shuffle (1,282 Features); Feature set 4: DenseNet-201 + Shuffle + Squeeze (1,674 Features); Feature set 5: DenseNet-201 + Shuffle + ResNet-50 (2,306 Features); Feature set 6: DenseNet-201 + Shuffle + MobileNet (1,922 Features); Feature set 7: DenseNet-201 + Shuffle + Inception-ResNet-V2 (2,050 Features); Feature set 8: DenseNet-201 + Shuffle + DarkNet-53 (1,794 Features); Feature set 9: DenseNet-201 + Shuffle + Xception (2,306 Features); Feature set 10: DenseNet-201 + Shuffle + NasNetMobile (1,810 Features).

TABLE 4 | Testing performance metrics of the feature sets selected using forward and backward strategies.

Selected feature set	Accuracy (%)	Sensitivity	Specificity	Precision	F1-score	MCC
Forward search strategy						
Feature Set 3	95.65	0.957	0.992	0.966	0.958	0.946
Backward search strategy						
Feature Set 5	95.65	0.957	0.975	0.961	0.955	0.940

MB. The presented pipeline involved a mixture of deep learning techniques fused by the DWT method. CoMB-Deep searched for the most acceptable combination of combined deep learning feature sets using forward and backward strategies. It consists of six phases which include image preprocessing, deep spatial feature extraction, feature fusion and reduction, fused feature set election, feature selection, and classification phases. In the first phase, images were resized and augmented, and then spatial deep features were extracted from 10 CNNs. Afterward, a fusion process was performed using the DWT method, which lowered the dimension of the fused features. Next, sequential forward and backward search strategies were utilized to explore the best-reduced feature sets, which improved the accuracy of CoMB-Deep. Two feature selection approaches were then employed further to reduce the selected feature sets of the previous phase. Lastly, in the classification phase, a bidirectional LSTM classifier was utilized to accomplish classification.

CoMB-Deep consists of a cascaded composite of CNNs and Bi-LSTM. For the feature extraction, these composite of CNNs extracted the spatial sequence of features (N) from the images. Afterward, these features were fused using the DWT method, which presents the time-frequency representation of the features. Next, the Bi-LSTM (Hochreiter and Schmidhuber, 1997) was employed to learn the temporal information from the set of N features vectors previously extracted and fused using the CNNs and DWT. The Bi-LSTM is an extension of a bidirectional RNN (Schuster and Paliwal, 1997) that is capable of modeling the temporal dependent states, and this is done through a direct cyclic connection among its units that can save its intrinsic hidden status and enable the modeling of the dynamic temporal

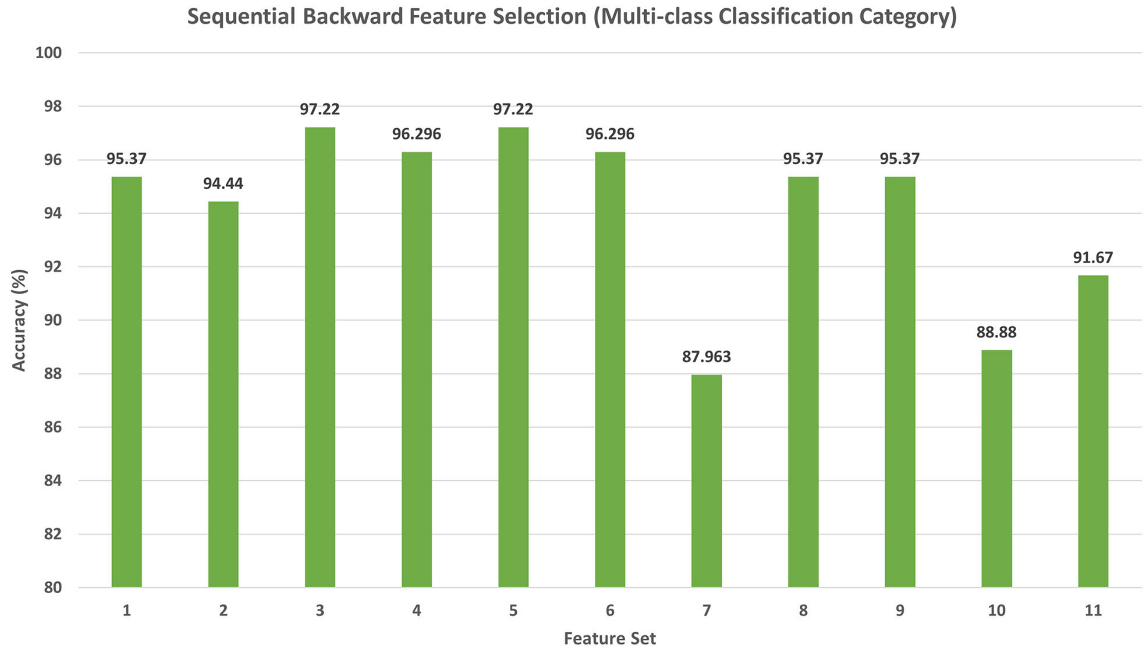

FIGURE 5 | The classification accuracy (%) attained for the different feature sets generated using backward strategy. Feature set 1: DenseNet-201 + Inception V3 + Shuffle + Squeeze + ResNet-50 + MobileNet + Inception-ResNet-V2 + DarkNet-53 + Xception + NasNetMobile (7,194 Features); Feature set 2: DenseNet-201 + Inception V3 +Shuffle + Squeeze + ResNet-50 +MobileNet + Inception-ResNet-V2 + DarkNet-53 +Xception (6,666 Features); Feature set 3: DenseNet-201 + Inception V3 + Shuffle + Squeeze + ResNet-50 + MobileNet + Inception-ResNet-V2 + DarkNet-53 + NasNetMobile (6,170 Features); Feature set 4: DenseNet-201 + Inception V3 + Shuffle + Squeeze + ResNet-50 + MobileNet + Inception-ResNet-V2 + NasNetMobile (5,658 Features); Feature set 5: DenseNet-201 + Inception V3 +Shuffle + Squeeze + ResNet-50 + MobileNet+DarkNet-53 + NasNetMobile (5,402 Features); Feature set 6: DenseNet-201 + Inception V3 + Shuffle + Squeeze + ResNet-50 + DarkNet-53 + NasNetMobile (4,762 Features); Feature set 7: DenseNet-201 + Inception V3 + Shuffle + Squeeze + MobileNet + DarkNet-53 + NasNetMobile (4,378 Features); Feature set 8: DenseNet-201 + Inception V3 +Shuffle + ResNet-50 + MobileNet+DarkNet-53 + NasNetMobile (5,010 Features); Feature set 9: DenseNet-201 +Inception V3 + Squeeze + ResNet-50 + MobileNet + DarkNet-53 + NasNetMobile (5,030 Features); Feature set 10: DenseNet-201 + Shuffle + Squeeze + ResNet-50 + MobileNet + DarkNet-53 + NasNetMobile (4,378 Features); Feature set 11: Inception V3 + Shuffle + Squeeze + ResNet-50 + MobileNet + DarkNet-53 + NasNetMobile (4,442 Features).

TABLE 5 | Binary classification accuracy (%), number of features, and other performance metrics before and after feature selection for ResNet-50 and Inception-ResNet CNNs.

Method	No of features	Accuracy (%)	Sensitivity	Specificity	Precision	F1-score	MCC
ResNet-50							
DWT	1,074	100	1	1	1	1	1
Relief-F	578	100	1	1	1	1	1
IG	578	100	1	1	1	1	1
Inception-ResNet							
DWT	818	100	1	1	1	1	1
Relief-F	200	100	1	1	1	1	1
IG	550	100	1	1	1	1	1

attitude. It consists of three gates called input, output, and forget gates, which help realize the prolonged temporal dependencies.

For the binary classification category of CoMB-Deep, it was noticed from the results of Scheme I that both Inception-ResNet-V2 and ResNet-50 had attained a 100% accuracy. Therefore, feature fusion using forward and backward strategies were not performed. The feature reduction process using DWT followed by the Relief-F and IG feature selection methods was only accomplished. **Figure 6** shows the number of features of Scheme I (spatial DL feature extraction), feature reduction (DWT), the Relief-F, and IG feature selection methods for both Inception-ResNet-V2 and ResNet-50 CNNs. The figure shows that the number of features of Scheme I is 2,048 and 1,536 for ResNet-50 and Inception-ResNet-V2, respectively. The DWT technique has reduced the number of features to 1,074 and 818 for ResNet-50, and Inception-ResNet-V2 DL features, respectively. These features were further reduced to 578 and 200 using the Relief-F approach, and 578 and 550 using the IG method for ResNet-50

TABLE 6 | Binary classification accuracy (%), number of features, and other performance metrics before and after feature selection for the feature set 5 selected using forward and backward strategies.

Method	No of features	Accuracy (%)	Sensitivity	Specificity	Precision	F1-score	MCC
Forward search							
Relief-F	448	98.05	0.981	0.993	0.981	0.981	0.972
IG	738	98.05	0.981	0.99	0.981	0.972	1
Backward search							
Relief-F	739	99.35	0.994	0.996	0.994	0.993	0.991
IG	2,313	99.35	0.994	0.996	0.994	0.993	0.991

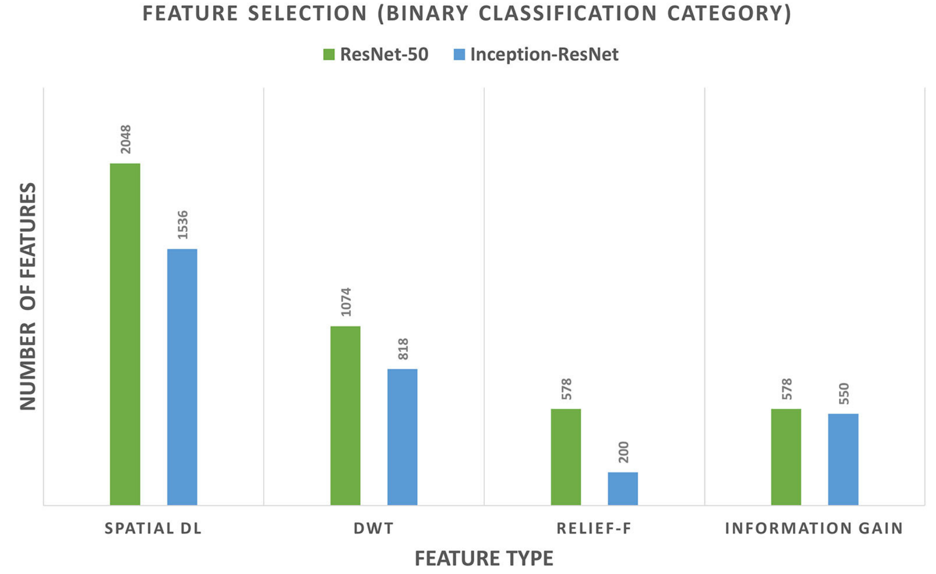

FIGURE 6 | The number of features of Scheme I (spatial DL feature extraction), feature reduction (DWT), and the Relief-F and information gain feature selection methods for both Inception-ResNet-V2 and ResNet-50 CNNs for the binary classification category.

and Inception-ResNet-V2 DL features. Note that the accuracy attained after all these processes is 100%. This means that the CoMB-Deep has successfully decreased the number of features used to construct the Bi-LSTM classifier while attaining the same accuracy of 100%.

On the other hand, for the multi-class classification, **Figure 7A** shows a comparison between the highest classification accuracies attained using the four schemes of CoMB-Deep using a 10-fold cross-validation. The figure indicates that the highest accuracy of 96.1% is obtained using ResNet-50 and DenseNet-201 features in Scheme I. For Scheme II, the same accuracy of 96.1% is obtained using feature set 3 (Dense-201 + Shuffle Features). The figure also verifies that Scheme III (Feature set 5 for backward search strategy without feature selection), including DenseNet-201 + Inception V3 + Shuffle + Squeeze + ResNet-50 + MobileNet + DarkNet-53 + NasNetMobile, and Scheme IV (Feature set 5 for backward search strategy with feature selection) have higher accuracies than the previous schemes. The highest accuracy attained using both Schemes III and IV is 99.35%. The peak accuracy in Scheme II (forward search strategy) was achieved using the feature set 3 corresponding to feature size of 1,282 features as shown in **Figure 7B**. The maximum accuracy in Scheme III was reached using feature set 5 (backward strategy), representing 5,402 features as illustrated in **Figure 7B**. In contrast, Scheme IV has lowered the feature dimension to 739 features using the Relief-F feature selection technique. **Figure 7B** compares the number of features obtained using the four schemes of CoMB-Deep.

Figure 8 shows a comparison between the classification accuracy of CoMB-Deep compared to the end-to-end DL classification of the 10 CNNs for the multi-class classification category. It is evident from the figure that CoMB-Deep has outperformed the 10 CNNs constructed based on an end-to-end classification procedure. This is because CoMB-Deep has attained an accuracy of 99.35% using 1,541 features selected using the IG method (Scheme IV) applied on the feature set 5 chosen using the forward search strategy (Scheme II). The outperformance of CoMB-Deep proves that it is better than using individual CNNs constructed by an end-to-end approach. This verifies that CoMB-Deep can help the pathologist classify the four

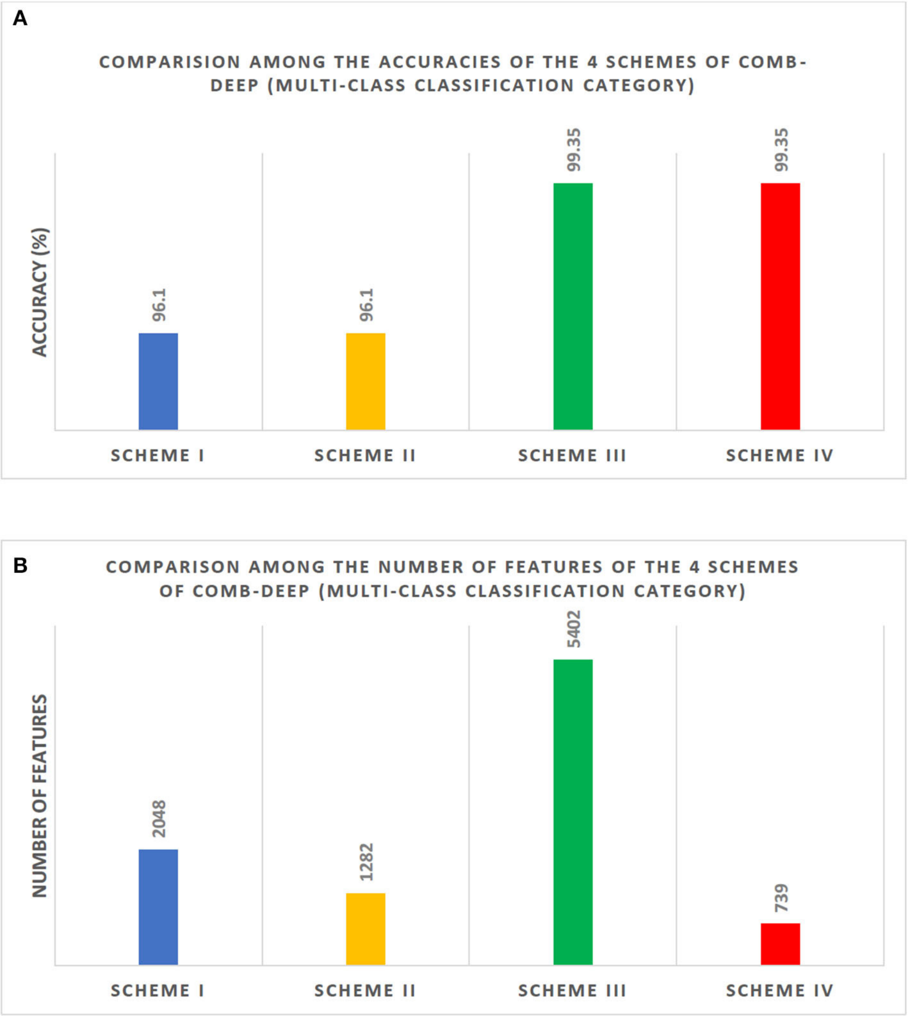

FIGURE 7 | Comparison between the four schemes of the multi-class classification category of CoMB-Deep. **(A)** The highest classification accuracy (%) attained using the four schemes. **(B)** The number of features utilized in the four schemes. Scheme I: ResNet-50 or DenseNet-201; Scheme II: DenseNet-201 + Shuffle; Scheme III: DenseNet-201 + Inception V3 + Shuffle + Squeeze + ResNet-50 + MobileNet + DarkNet-53 + NasNetMobile; Scheme IV: DenseNet-201 + Inception V3 + Shuffle + Squeeze + ResNet-50 + MobileNet + DarkNet-53 + NasNetMobile.

childhood MB classes, which can help them select the appropriate treatment and follow-up plan.

The performance of CoMB-Deep is compared with related studies based on the same dataset to prove its competence. **Table 7** displays this comparison. Das et al. (2018b) extracted textural features, comprising GLCM, HOG, Tamura, LBP, and GLCM. They fused all these features and used PCA to reduce them and SVM classifier to classify pediatric MB classes. The highest accuracy attained was 84.9%. Similarly, the same authors utilized the same textural features and fused them all using MANOVA. The authors employed SVM for classification, attaining an accuracy of 65.2%. It can be noticed from these results that fusing all these features does not always guarantee the best performance. Therefore, the authors (Das et al., 2020b) searched for the best combination of these feature extraction methods and found that fusing only four features sets (GLCM + Tamura + LB + GRLN) has the highest impact on the classification performance, obtaining an accuracy of 91.3% using SVM classifier and 96.7% using PCA with SVM classifier. This means that investigating a different combination of feature sets and selecting the most influential fused feature set can improve the accuracy of the classifier. In the same manner, CoMB-Deep pipeline explored different combinations of features extracted from several CNNs and fused using DWT to select the most fused feature sets that impact the performance of the classifier As mentioned before DL techniques are more favorable than the traditional feature extraction methods (Das et al., 2018b, 2020a,b). CoMB-Deep merges DL and textural analysis benefits, as it first extracts spatial features from 10 CNNs. It then searches for the most significant feature sets fused using DWT producing spatial-temporal-frequency features. DWT is a textural analysis

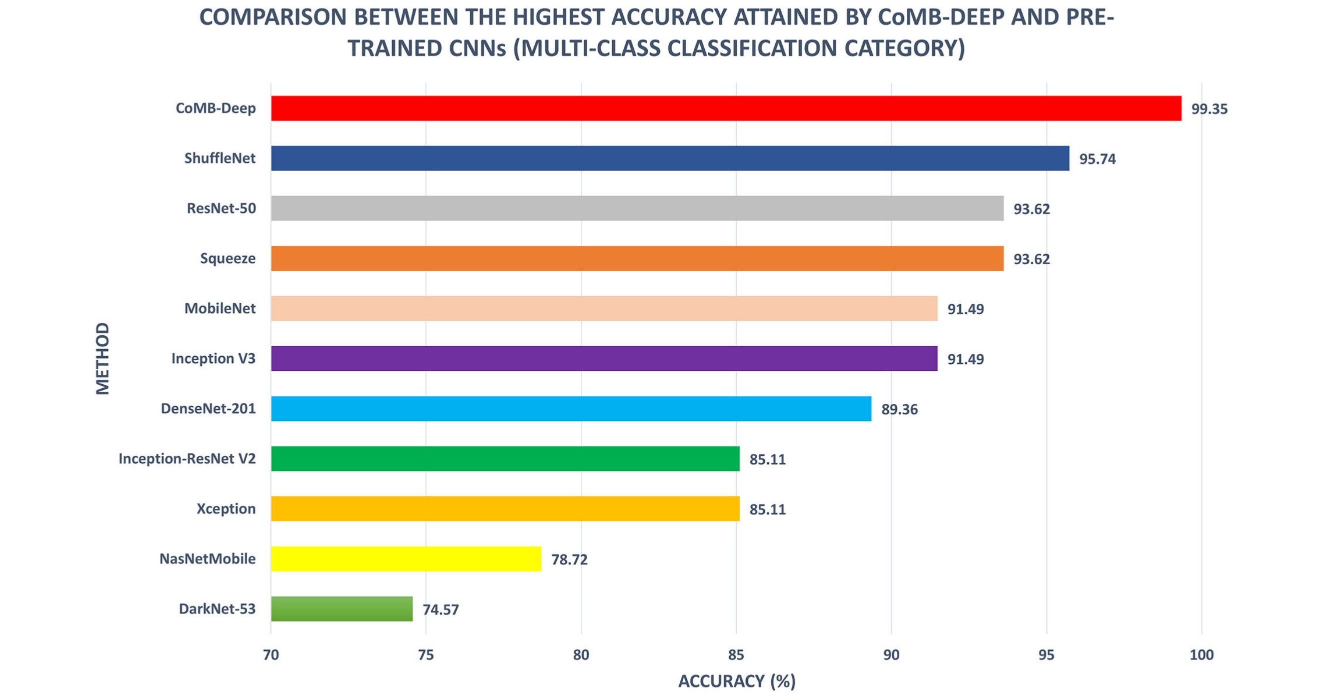

FIGURE 8 | Multi-class classification accuracy (%) of CoMB-Deep compared to end-to-end deep learning classification of 10 CNNs.

technique that illustrates the temporal-frequency information of the data while reducing the dimension of the features. Finally, the Bi-LSTM DL technique is used for classification. The accuracy attained by CoMB-Deep is 99.35% which proves that merging DL techniques is better than conventional feature extraction techniques as it is higher than those achieved by Das et al. (2018b, 2020a,b). Also, it verifies that combining DL and textural analysis feature extraction methods can enhance the classification accuracy. Furthermore, utilizing spatial-temporal-frequency features is better than using each one of them independently.

The results indicated in **Table 7** verify that CoMB-Deep is a competitive pipeline for both classification categories. This is obvious as CoMB-Deep accomplished an accuracy of 100% (for binary classification category) which is the same as observed by Das et al. (2018b, 2019, 2020a), but greater than that achieved by Das et al. (2020c). The competitiveness of CoMB-Deep is found in its ability to distinguish the subclasses of MB. This is because it reached an accuracy of 98.05%, a sensitivity of 0.981, a specificity of 0.993, and a precision of 0.981 using the forward strategy in Scheme IV. It also reached 99.35% accuracy, a sensitivity of 0.998, specificity, and precision of 0.994 using the backward strategy in Scheme IV. These performance metrics achieved in Scheme IV are greater than all related works. This outperformance confirms that CoMB-Deep is reliable; thus, it may be utilized to support clinicians and pathologists in performing diagnosis with high accuracy, decreasing the chances of misdiagnosis during manual diagnosis, and speeding up the classification process, and finally, lower the cost of diagnosis.

CONCLUSION

This study introduced a computer-assisted pipeline called CoMB-Deep, a composite of deep learning techniques for the automatic classification of MB and its classes. It has six phases: image preprocessing, spatial DL feature extraction, feature fusion and reduction, fused feature set election, feature selection, and classification phases. CoMB-Deep examined if deep feature fusion can enhance the accuracy of Bi-LSTM classifier compared to deep individual features. The fusion phase was accomplished using DWT, which lowers the dimension of fusion features, whereas deep feature sets are explored using forward and backward search methods. The results showed that deep feature fusion had improved the performance of the Bi-LSTM classifier. To further reduce the dimension of selected features chosen using forward and backward approaches, Relief-F and information gain feature selection methods were utilized. The results verified that both methods had successfully decreased the dimension of features while attaining the same peak accuracy of 100% and 99.35% for binary and multi-class classification categories, respectively. The performance of CoMB-Deep was compared with related works, which proved its competence; thus, it is reliable and can be used to help pathologists in accurately classifying childhood MB and its classes. Also, it may reduce the complications that pathologists face during manual diagnosis. It can accelerate the classification procedure while achieving high accuracy, decreasing the classification cost, lowering the possibility of tumor progression, and helping the pathologist choose suitable follow-up and treatment plans.

TABLE 7 | A comparison between CoMB-Deep and related studies using the same dataset.

	Classification category				
	Method	**Sensitivity (%)**	**Precision (%)**	**Specificity (%)**	**Accuracy (%)**
Das et al. (2019)	HOG, GLCM, Tamura, and LBP features + GRLN + SVM	100	100	100	100
Das et al. (2020c)	AlexNet + SVM VGG-16 + SVM	–	–	–	99.44 99.62
Das et al. (2018b)	(Shape + Color) features + PCA + SVM	100	100	100	100
Das et al. (2020c)	AlexNet sVGG-16	–	–	–	98.5 98.12
Das et al. (2020a)	HOG, GLCM, Tamura, and LBP features + GRLN + MANOVA + SVM	100	100	100	100
Proposed CoMB-Deep	Inception-ResNet + DWT + Information gain + LSTM	100	100	100	100
	Classification (multi-class category)				
		Sensitivity (%)	**Precision (%)**	**Specificity (%)**	**Accuracy (%)**
Das et al. (2018b)	(Shape + Color) features + PCA + SVM	–	–	–	84.9
Das et al. (2020a)	HOG, GLCM, Tamura, and LBP features + GRLN + MANOVA + SVM	72	66.6	–	65.21
Das et al. (2020c)	AlexNet VGG-16 s	–	–	–	79.33 65.4
Das et al. (2020b)	GLCM + Tamura + LBP + GRLN + SVM	91.3	91.3	97	91.3
Das et al. (2020b)	GLCM + Tamura + LBP + GRLN + PCA + SVM	–	–	–	96.7
Das et al. (2020c)	AlexNet + SVM VGG-16 + SVM	–	–	–	93.21 93.38
Proposed CoMB-Deep	Deep features of (DenseNet-201 + ShuffleNet) + Relief-F + Bi-LSTM	98.1	98.1	99.3	98.05
	Deep features of (DenseNet-201 + Inception + Resnet-50 + Darknet-53 + MobileNet + ShuffleNet + SqueezeNet + NasNetMobile) + Relief-F + Bi-LSTM	99.8	99.4	99.4	99.35

Future work will focus on merging handcrafted features used in the literature with DL techniques. Also, further analysis will be conducted on using other DL techniques to analyze childhood MB classes.

DATA AVAILABILITY STATEMENT

The dataset analyzed for this study can be found in the IEEE Dataport [https://ieee-dataport.org/open-access/childhood-medulloblastoma-microscopic-images].

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fninf.2021.663592/full#supplementary-material

REFERENCES

Afifi, A. J., and Ashour, W. M. (2012). Image retrieval based on content using color feature. *ISRN Comput. Graph*. 2012:248285. doi: 10.5402/2012/248285

Ailion, A. S., Hortman, K., and King, T. Z. (2017). Childhood brain tumors: a systematic review of the structural neuroimaging literature. *Neuropsychol. Rev.* 27, 220–244. doi: 10.1007/s11065-017-9352-6

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., et al. (2019). A state-of-the-art survey on deep learning theory and architectures. *Electronics* 8, 292–357. doi: 10.3390/electronics8030292

Angeline, P. J., Saunders, G. M., and Pollack, J. B. (1994). An evolutionary algorithm that constructs recurrent neural networks. *IEEE Trans. Neural Netw.* 5, 54–65. doi: 10.1109/72.265960

Antonini, M., Barlaud, M., Mathieu, P., and Daubechies, I. (1992). Image coding using wavelet transform. *IEEE Trans. Image Process.* 1, 205–220. doi: 10.1109/83.136597

Anwar, F., Attallah, O., Ghanem, N., and Ismail, M. A. (2020). "Automatic breast cancer classification from histopathological images," in *2019 International Conference on Advances in the Emerging Computing Technologies (AECT)* (Al Madinah Al Munawwarah: IEEE), 1–6. doi: 10.1109/AECT47998.2020.9194194

Arseni, C., and Ciurea, A. V. (1981). Statistical survey of 276 cases of medulloblastoma (1935–1978). *Acta Neurochir.* 57, 159–162. doi: 10.1007/BF01664834

Attallah, O. (2020). An effective mental stress state detection and evaluation system using minimum number of frontal brain electrodes. *Diagnostics* 10, 292–327. doi: 10.3390/diagnostics10050292

Attallah, O. (2021). MB-AI-His: histopathological diagnosis of pediatric medulloblastoma and its subtypes via AI. *Diagnostics* 11, 359–384. doi: 10.3390/diagnostics11020359

Attallah, O., Abougharbia, J., Tamazin, M., and Nasser, A. A. (2020a). A BCI system based on motor imagery for assisting people with motor deficiencies in the limbs. *Brain Sci.* 10, 864–888. doi: 10.3390/brainsci10110864

Attallah, O., Karthikesalingam, A., Holt, P. J., Thompson, M. M., Sayers, R., Bown, M. J., et al. (2017a). Feature selection through validation and un-censoring of endovascular repair survival data for predicting the risk of re-intervention. *BMC Med. Informatics Decision Making* 17, 115–133. doi: 10.1186/s12911-017-0508-3

Attallah, O., Karthikesalingam, A., Holt, P. J., Thompson, M. M., Sayers, R., Bown, M. J., et al. (2017b). Using multiple classifiers for predicting the risk of endovascular aortic aneurysm repair re-intervention through hybrid feature selection. *Proc. Inst. Mech. Eng. Part H J. Eng. Med.* 231, 1048–1063. doi: 10.1177/0954411917731592

Attallah, O., Ragab, D. A., and Sharkas, M. (2020b). MULTI-DEEP: a novel CAD system for coronavirus (COVID-19) diagnosis from CT images using multiple convolution neural networks. *PeerJ* 8:e10086. doi: 10.7717/peerj.10086

Attallah, O., Sharkas, M. A., and Gadelkarim, H. (2019). Fetal brain abnormality classification from MRI images of different gestational age. *Brain Sci.* 9, 231–252. doi: 10.3390/brainsci9090231

Attallah, O., Sharkas, M. A., and Gadelkarim, H. (2020c). Deep learning techniques for automatic detection of embryonic neurodevelopmental disorders. *Diagnostics* 10, 27–49. doi: 10.3390/diagnostics10010027

Babu, J., Rangu, S., and Manogna, P. (2017). A survey on different feature extraction and classification techniques used in image steganalysis. *J. Inf. Secur.* 8, 186–202. doi: 10.4236/jis.2017.83013

Baldi, P., Brunak, S., Frasconi, P., Soda, G., and Pollastri, G. (1999). Exploiting the past and the future in protein secondary structure prediction. *Bioinformatics* 15, 937–946. doi: 10.1093/bioinformatics/15.11.937

Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., et al. (2019). Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage Clin.* 21:101645. doi: 10.1016/j.nicl.2018.101645

Cai, J., Luo, J., Wang, S., and Yang, S. (2018). Feature selection in machine learning: a new perspective. *Neurocomputing* 300, 70–79. doi: 10.1016/j.neucom.2017.11.077

Ceschin, R., Zahner, A., Reynolds, W., Gaesser, J., Zuccoli, G., Lo, C. W., et al. (2018). A computational framework for the detection of subcortical brain dysmaturation in neonatal MRI using 3D convolutional neural networks. *NeuroImage* 178, 183–197. doi: 10.1016/j.neuroimage.2018.05.049

Chandrashekar, G., and Sahin, F. (2014). A survey on feature selection methods. *Comput. Electric. Eng.* 40, 16–28. doi: 10.1016/j.compeleceng.2013.11.024

Chollet, F. (2017). "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI: IEEE), 1251–1258. doi: 10.1109/CVPR.2017.195

Colquhoun, D. (2014). An investigation of the false discovery rate and the misinterpretation of p-values. *R. Soc. Open Sci.* 1:140216. doi: 10.1098/rsos.140216

Cruz-Roa, A., Arévalo, J., Basavanhally, A., Madabhushi, A., and González, F. (2015). "A comparative evaluation of supervised and unsupervised representation learning approaches for anaplastic medulloblastoma differentiation," in *10th International Symposium on Medical Information Processing and Analysis* (Cartagena de Indias: International Society for Optics and Photonics), 92870G. doi: 10.1117/12.2073849

Cruz-Roa, A., González, F., Galaro, J., Judkins, A. R., Ellison, D., Baccon, J., et al. (2012). "A visual latent semantic approach for automatic analysis and interpretation of anaplastic medulloblastoma virtual slides," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Nice: Springer), 157–164. doi: 10.1007/978-3-642-33415-3_20

Das, D., Mahanta, L. B., Ahmed, S., and Baishya, B. K. (2020a). A study on MANOVA as an effective feature reduction technique in classification of childhood medulloblastoma and its subtypes. *Netw. Model. Anal. Health Informatics Bioinformatics* 9, 1–15. doi: 10.1007/s13721-020-0221-5

Das, D., Mahanta, L. B., Ahmed, S., and Baishya, B. K. (2020b). Classification of childhood medulloblastoma into WHO-defined multiple subtypes based on textural analysis. *J. Microsc.* 279, 26–38. doi: 10.1111/jmi.12893

Das, D., Mahanta, L. B., Ahmed, S., Baishya, B. K., and Haque, I. (2018b). Study on contribution of biological interpretable and computer-aided features towards the classification of childhood medulloblastoma cells. *J. Med. Syst.* 42, 1–12. doi: 10.1007/s10916-018-1008-4

Das, D., Mahanta, L. B., Ahmed, S., Baishya, B. K., and Haque, I. (2019). Automated classification of childhood brain tumours based on texture feature. *Songklanakarin J. Sci. Technol.* 41, 1014–1020.

Das, D., Mahanta, L. B., Baishya, B. K., and Ahmed, S. (2020c). "Classification of childhood medulloblastoma and its subtypes using transfer learning features-a comparative study of deep convolutional neural networks," in *2020 International Conference on Computer, Electrical & Communication Engineering (ICCECE)* (Kolkata: IEEE), 1–5. doi: 10.1109/ICCECE48148.2020.9223104

Das, D., Mahanta, L. B., Baishya, B. K., Haque, I., and Ahmed, S. (2018a). Automated histopathological diagnosis of pediatric medulloblastoma–a review study. *Int. J. Appl. Eng. Res.* 13, 9909–9915.

Dasgupta, A., and Gupta, T. (2019). MRI-based prediction of molecular subgrouping in medulloblastoma: images speak louder than words. *Oncotarget* 10, 4805–4807. doi: 10.18632/oncotarget.27097

Davis, F. G., Freels, S., Grutsch, J., Barlas, S., and Brem, S. (1998). Survival rates in patients with primary malignant brain tumors stratified by patient age and tumor histological type: an analysis based on Surveillance, Epidemiology, and End Results (SEER) data, 1973–1991. *J. Neurosurg.* 88, 1–10. doi: 10.3171/jns.1998.88.1.0001

Demirel, H., Ozcinar, C., and Anbarjafari, G. (2009). Satellite image contrast enhancement using discrete wavelet transform and singular value decomposition. *IEEE Geosci. Remote Sens. Lett.* 7, 333–337. doi: 10.1109/LGRS.2009.2034873

Dong, J., Li, L., Liang, S., Zhao, S., Zhang, B., Meng, Y., et al. (2020). Differentiation between ependymoma and medulloblastoma in children with radiomics approach. *Acad. Radiol.* 28, 318–327. doi: 10.1016/j.acra.2020.02.012

Ellis, P. D. (2010). *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511761676

Ellison, D. W. (2010). Childhood medulloblastoma: novel approaches to the classification of a heterogeneous disease. *Acta Neuropathol.* 120, 305–316. doi: 10.1007/s00401-010-0726-6

Fan, Y., Feng, M., and Wang, R. (2019). Application of radiomics in central nervous system diseases: a systematic literature review. *Clin. Neurol. Neurosurg.* 187:105565. doi: 10.1016/j.clineuro.2019.105565

Fauvel, M., Dechesne, C., Zullo, A., and Ferraty, F. (2015). Fast forward feature selection of hyperspectral images for classification with Gaussian mixture models. *IEEE J. Select. Top. Appl. Earth Observ. Remote Sens.* 8, 2824–2831. doi: 10.1109/JSTARS.2015.2441771

Fetit, A. E., Novak, J., Rodriguez, D., Auer, D. P., Clark, C. A., Grundy, R. G., et al. (2018). Radiomics in paediatric neuro-oncology: a multicentre study on MRI texture analysis. *NMR Biomed.* 31:e3781. doi: 10.1002/nbm.3781

Fujita, H. (2020). AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. *Radiol. Phys. Technol.* 13, 1–14. doi: 10.1007/s12194-019-00552-4

Furata, T., Tabuchi, A., Adachi, Y., Mizumatsu, S., Tamesa, N., Ichikawa, T., et al. (1998). Primary brain tumors in children under age 3 years. *Brain Tumor Pathol.* 15, 7–12. doi: 10.1007/BF02482094

Galaro, J., Judkins, A. R., Ellison, D., Baccon, J., and Madabhushi, A. (2011). "An integrated texton and bag of words classifier for identifying anaplastic medulloblastomas," in *2011 Annual International Conference of the IEEE*

Engineering in Medicine and Biology Society (Boston, MA: IEEE), 3443–3446. doi: 10.1109/IEMBS.2011.6090931

Grist, J. T., Withey, S., MacPherson, L., Oates, A., Powell, S., Novak, J., et al. (2020). Distinguishing between paediatric brain tumour types using multi-parametric magnetic resonance imaging and machine learning: a multi-site study. *NeuroImage Clin.* 25:102172. doi: 10.1016/j.nicl.2020.102172

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explor. Newslett.* 11, 10–18. doi: 10.1145/1656274.1656278

Han, D., Liu, Q., and Fan, W. (2018). A new image classification method using CNN transfer learning and web data augmentation. *Expert Syst. Appl.* 95, 43–56. doi: 10.1016/j.eswa.2017.11.028

Hatamimajoumerd, E., and Talebpour, A. (2019). A temporal neural trace of wavelet coefficients in human object vision: an MEG study. *Front. Neural Circuits* 13:20. doi: 10.3389/fncir.2019.00020

Hatamimajoumerd, E., Talebpour, A., and Mohsenzadeh, Y. (2020). Enhancing multivariate pattern analysis for magnetoencephalography through relevant sensor selection. *Int. J. Imaging Syst. Technol.* 30, 473–494. doi: 10.1002/ima.22398

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

Hira, Z. M., and Gillies, D. F. (2015). A review of feature selection and feature extraction methods applied on microarray data. *Adv. Bioinformatics* 2015:198363. doi: 10.1155/2015/198363

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Hovestadt, V., Ayrault, O., Swartling, F. J., Robinson, G. W., Pfister, S. M., and Northcott, P. A. (2020). Medulloblastomics revisited: biological and clinical insights from thousands of patients. *Nat. Rev. Cancer* 20, 42–56. doi: 10.1038/s41568-019-0223-8

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: efficient convolutional neural networks for mobile vision applications. *arXiv [Preprint].* arXiv:1704.04861.

Hssayeni, M. D., Saxena, S., Ptucha, R., and Savakis, A. (2017). Distracted driver detection: deep learning vs handcrafted features. *Electronic Imaging* 2017, 20–26. doi: 10.2352/ISSN.2470-1173.2017.10.IMAWM-162

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

Humeau-Heurtier, A. (2019). Texture feature extraction methods: a survey. *IEEE Access* 7, 8975–9000. doi: 10.1109/ACCESS.2018.2890743

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., and Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. *arXiv [Preprint].* arXiv:1602.07360.

Iqbal, S., Khan, M. U. G., Saba, T., and Rehman, A. (2018). Computer-assisted brain tumor type discrimination using magnetic resonance imaging features. *Biomed. Engi. Lett.* 8, 5–28. doi: 10.1007/s13534-017-0050-3

Iv, M., Zhou, M., Shpanskaya, K., Perreault, S., Wang, Z., Tranvinh, E., et al. (2019). MR imaging–based radiomic signatures of distinct molecular subgroups of medulloblastoma. *Am. J. Neuroradiol.* 40, 154–161. doi: 10.3174/ajnr.A5899

Kononenko, I. (1994). "Estimating attributes: analysis and extensions of RELIEF," in *European Conference on Machine Learning* (Catania: Springer), 171–182. doi: 10.1007/3-540-57868-4_57

Kumar, A., Singh, S. K., Saxena, S., Lakshmanan, K., Sangaiah, A. K., Chauhan, H., et al. (2020). Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer. *Inf. Sci.* 508, 405–421. doi: 10.1016/j.ins.2019.08.072

Lai, Y., Viswanath, S., Baccon, J., Ellison, D., Judkins, A. R., and Madabhushi, A. (2011). "A texture-based classifier to discriminate anaplastic from non-anaplastic medulloblastoma," in *2011 IEEE 37th Annual Northeast Bioengineering Conference (NEBEC)* (Troy, NY: IEEE), 1–2.

Li, B., and Meng, M. Q.-H. (2012). Tumor recognition in wireless capsule endoscopy images using textural features and SVM-based feature selection. *IEEE Trans. Inf. Technol. Biomed.* 16, 323–329. doi: 10.1109/TITB.2012.2185807

Li, G., Zhang, M., Li, J., Lv, F., and Tong, G. (2021). Efficient densely connected convolutional neural networks. *Pattern Recogn.* 109:107610. doi: 10.1016/j.patcog.2020.107610

Li, J., and Liu, H. (2017). Challenges of feature selection for big data analytics. *IEEE Intelligent Syst.* 32, 9–15. doi: 10.1109/MIS.2017.38

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, J., and Shi, Y. (2011). "Image feature extraction method based on shape characteristics and its application in medical image analysis," in *International Conference on Applied Informatics and Communication* (Xi'ian: Springer), 172–178. doi: 10.1007/978-3-642-23214-5_24

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., and Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing* 234, 11–26. doi: 10.1016/j.neucom.2016.12.038

Manias, K. A., Gill, S. K., MacPherson, L., Foster, K., Oates, A., and Peet, A. C. (2017). Magnetic resonance imaging based functional imaging in paediatric oncology. *Eur. J. Cancer* 72, 251–265. doi: 10.1016/j.ejca.2016.10.037

Otálora, S., Cruz-Roa, A., Arevalo, J., Atzori, M., Madabhushi, A., Judkins, A. R., et al. (2015). "Combining unsupervised feature learning and riesz wavelets for histopathology image representation: application to identifying anaplastic medulloblastoma," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Munich: Springer), 581–588. doi: 10.1007/978-3-319-24553-9_71

Pan, S. J., and Yang, Q. (2009). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191

Park, S., Yu, S., Kim, J., Kim, S., and Lee, S. (2012). 3D hand tracking using Kalman filter in depth space. *EURASIP J. Adv. Signal Process.* 2012, 1–18. doi: 10.1186/1687-6180-2012-36

Pickles, J. C., Hawkins, C., Pietsch, T., and Jacques, T. S. (2018). CNS embryonal tumours: WHO 2016 and beyond. *Neuropathol. Appl. Neurobiol.* 44, 151–162. doi: 10.1111/nan.12443

Pohjalainen, J., Räsänen, O., and Kadioglu, S. (2015). Feature selection methods and their combinations in high-dimensional classification of speaker likability, intelligibility and personality traits. *Comput. Speech Lang.* 29, 145–171. doi: 10.1016/j.csl.2013.11.004

Polednak, A. P., and Flannery, J. T. (1995). Brain, other central nervous system, and eye cancer. *Cancer* 75, 330–337. doi: 10.1002/1097-0142(19950101)75:1+<330::AID-CNCR2820751315>3.0.CO;2-5

Pollack, I. F., Agnihotri, S., and Broniscer, A. (2019). Childhood brain tumors: current management, biological insights, and future directions: JNSPG 75th Anniversary Invited Review Article. *J. Neurosurg. Pediatrics* 23, 261–273. doi: 10.3171/2018.10.PEDS18377

Pollack, I. F., and Jakacki, R. I. (2011). Childhood brain tumors: epidemiology, current management and future directions. *Nat. Rev. Neurol.* 7:495. doi: 10.1038/nrneurol.2011.110

Ponnusamy, R., and Sathiamoorthy, S. (2019). "Bleeding and Z-line classification by DWT based SIFT using KNN and SVM," in *International Conference On Computational Vision and Bio Inspired Computing* (Coimbatore: Springer), 679–688. doi: 10.1007/978-3-030-37218-7_77

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., et al. (2018). A survey on deep learning: algorithms, techniques, and applications. *ACM Comput. Surv.* 51, 1–36. doi: 10.1145/3150226

Ragab, D. A., and Attallah, O. (2020). FUSI-CAD: coronavirus (COVID-19) diagnosis based on the fusion of CNNs and handcrafted features. *PeerJ Comput. Sci.* 6:e306. doi: 10.7717/peerj-cs.306

Ragab, D. A., Sharkas, M., and Attallah, O. (2019). Breast cancer diagnosis using an efficient CAD system based on multiple classifiers. *Diagnostics* 9, 165–191. doi: 10.3390/diagnostics9040165

Raghu, M., Zhang, C., Kleinberg, J., and Bengio, S. (2019). Transfusion: understanding transfer learning for medical imaging. *arXiv [Preprint].* arXiv:1902.07208.

Redmon, J., and Farhadi, A. (2017). "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI: IEEE), 7263–7271. doi: 10.1109/CVPR.2017.690

Ritzmann, T. A., and Grundy, R. G. (2018). Translating childhood brain tumour research into clinical practice: the experience of molecular classification and diagnostics. *Paediatrics Child Health* 28, 177–182. doi: 10.1016/j.paed.2018.01.006

Robboy, S. J., Weintraub, S., Horvath, A. E., Jensen, B. W., Alexander, C. B., Fody, E. P., et al. (2013). Pathologist workforce in the United States: I. Development of a predictive model to examine factors influencing supply. *Arch. Pathol. Lab. Med.* 137, 1723–1732. doi: 10.5858/arpa.2013-0200-OA

Roobaert, D., Karakoulas, G., and Chawla, N. V. (2006). "Information gain, correlation and support vector machines," in *Feature Extraction* (Berlin: Springer), 463–470. doi: 10.1007/978-3-540-35488-8_23

Sachdeva, J., Kumar, V., Gupta, I., Khandelwal, N., and Ahuja, C. K. (2013). Segmentation, feature extraction, and multiclass brain tumor classification. *J. Digital Imaging* 26, 1141–1150. doi: 10.1007/s10278-013-9600-0

Sánchez-Marono, N., Alonso-Betanzos, A., and Tombilla-Sanromán, M. (2007). "Filter methods for feature selection–a comparative study," in *International Conference on Intelligent Data Engineering and Automated Learning* (Birmingham: Springer), 178–187. doi: 10.1007/978-3-540-77226-2_19

Schuster, M., and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* 45, 2673–2681. doi: 10.1109/78.650093

Shorten, C., and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *J. Big Data* 6, 1–48. doi: 10.1186/s40537-019-0197-0

Szalontay, L., and Khakoo, Y. (2020). Medulloblastoma: an old diagnosis with new promises. *Curr. Oncol. Rep.* 22, 1–13. doi: 10.1007/s11912-020-00953-4

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. (2016a). Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv [Preprint].* arXiv:1602.07261.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016b). "Rethinking the inception architecture for computer vision." in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV: IEEE), 2818–2826. doi: 10.1109/CVPR.2016.308

Thrun, S., and Pratt, L. (1998). "Learning to Learn: Introduction and Overview," in *Learning to Learn*, eds S. Thrun and L. Pratt (Boston, MA: Springer US), 3–17. doi: 10.1007/978-1-4615-5529-2_1

Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., and Moore, J. H. (2018). Relief-based feature selection: introduction and review. *J. Biomed. Informatics* 85, 189–203. doi: 10.1016/j.jbi.2018.07.014

Ververidis, D., and Kotropoulos, C. (2005). "Sequential forward feature selection with low computational cost," in *2005 13th European Signal Processing Conference* (Antalya: IEEE), 1–4.

Yanase, J., and Triantaphyllou, E. (2019). A systematic survey of computer-aided diagnosis in medicine: past and present developments. *Expert Syst. Appl.* 138:112821. doi: 10.1016/j.eswa.2019.112821

Zarinabad, N., Abernethy, L. J., Avula, S., Davies, N. P., Rodriguez Gutierrez, D., Jaspan, T., et al. (2018). Application of pattern recognition techniques for classification of pediatric brain tumors by *in vivo* 3T 1H-MR spectroscopy—a multi-center study. *Magnetic Resonance Med.* 79, 2359–2366. doi: 10.1002/mrm.26837

Zhang, J., Liang, J., and Zhao, H. (2012). Local energy pattern for texture classification using self-adaptive quantization thresholds. *IEEE Trans. Image Process.* 22, 31–42. doi: 10.1109/TIP.2012.2214045

Zhang, X., Zhang, Y., Qian, B., Liu, X., Li, X., Wang, X., et al. (2019). "Classifying breast cancer histopathological images using a robust artificial neural network architecture," in *Bioinformatics and Biomedical Engineering Lecture Notes in Computer Science,* eds I. Rojas, O. Valenzuela, F. Rojas, and F. Ortuño (Cham: Springer International Publishing), 204–215.

Zhang, X., Zhou, X., Lin, M., and Sun, J. (2018). "Shufflenet: an extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 6848–6856. doi: 10.1109/CVPR.2018.00716

Zoph, B., Vasudevan, V., Shlens, J., and Le, Q. V. (2018). "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 8697–8710. doi: 10.1109/CVPR.2018.00907

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.

BRIEF RESEARCH REPORT
published: 21 June 2021
doi: 10.3389/fncom.2021.674028

Accurate Prediction of Children's ADHD Severity Using Family Burden Information: A Neural Lasso Approach

Juan C. Laria[1], *David Delgado-Gómez*[1,2]*, *Inmaculada Peñuelas-Calvo*[3], *Enrique Baca-García*[3,4] and *Rosa E. Lillo*[1,2]

[1] Department of Statistics, University Carlos III of Madrid, Madrid, Spain, [2] Santander Big Data Institute, Universidad Carlos III de Madrid, Madrid, Spain, [3] Department of Psychiatry, Fundación Jiménez Díaz Hospital, Madrid, Spain, [4] Department of Psychiatry, Nimes University Hospital, Nimes, France

The deep lasso algorithm (dlasso) is introduced as a neural version of the statistical linear lasso algorithm that holds benefits from both methodologies: feature selection and automatic optimization of the parameters (including the regularization parameter). This last property makes dlasso particularly attractive for feature selection on small samples. In the two first conducted experiments, it was observed that dlasso is capable of obtaining better performance than its non-neuronal version (traditional lasso), in terms of predictive error and correct variable selection. Once that dlasso performance has been assessed, it is used to determine whether it is possible to predict the severity of symptoms in children with ADHD from four scales that measure family burden, family functioning, parental satisfaction, and parental mental health. Results show that dlasso is able to predict parents' assessment of the severity of their children's inattention from only seven items from the previous scales. These items are related to parents' satisfaction and degree of parental burden.

Keywords: deep learning, lasso, feature selection, interpretability, ADHD

OPEN ACCESS

Edited by:
Yalda Mohsenzadeh,
Western University, Canada

Reviewed by:
Qing Yun Wang,
Beihang University, China
Rukiye Uzun Arslan,
Bülent Ecevit University, Turkey

Correspondence:
David Delgado-Gómez
ddelgado@est-econ.uc3m.es

Received: *28 February 2021*
Accepted: *20 May 2021*
Published: *21 June 2021*

Citation:
Laria JC, Delgado-Gómez D, Peñuelas-Calvo I, Baca-García E and Lillo RE (2021) Accurate Prediction of Children's ADHD Severity Using Family Burden Information: A Neural Lasso Approach. Front. Comput. Neurosci. 15:674028. doi: 10.3389/fncom.2021.674028

1. INTRODUCTION

Attention-deficit hyperactivity disorder (ADHD) is the most common chronic psychiatric disorder in childhood (Wender and Tomb, 2016). According to a recent systematic review, this neurodevelopmental disorder has an estimated prevalence in children and adolescents of 7.2% (Thomas et al., 2015). ADHD is characterized by inattention, excessive activity, and impulsive behavior. Children with ADHD have a higher risk of suffering from accidents, school failure or addiction problems (Harpin, 2005; Elkins et al., 2007). In addition, it has been observed that untreated children present low self-esteem and poor social functioning in the long term (Harpin et al., 2016). Fortunately, it has been observed that these negative consequences are reduced with an early and accurate diagnosis (Sonuga-Barke et al., 2011).

The diagnosis of ADHD is obtained through a clinical interview in which the clinician relies on the information provided by parents or teachers. However, several studies have shown that this information can be influenced by the characteristics of the informant. For example, it has been shown that female young teachers tend to provide more severe scores than older male teachers (Schultz and Evans, 2012). In another study, Chi and Hinshaw predicted the discrepancies between the reports provided by the mother and the teacher based on the mother's responses to the Beck

Depression Inventory (Chi and Hinshaw, 2002). They observed that the responses provided by mothers with depression were negatively biased. This result has been validated in other studies (Harvey et al., 2013; Madsen et al., 2020). In addition, it has been observed that parental stress is another factor that explains the discrepancy between parents and teachers (Yeguez and Sibley, 2016; Chen et al., 2017).

This article, which expands the above studies, focuses on determining whether it is possible to predict the severity of a child's inattention and hyperactivity/impulsivity reported by his/her parents based on their distress and family burden. Furthermore, it seeks to identify the factors that influence parents' assessments. To achieve these objectives, the deep lasso (dlasso) algorithm is developed. This algorithm combines recent advances in the fields of machine learning and statistics: deep learning and the least absolute shrinkage and selection operator (lasso).

Without a doubt, deep learning has become one of the greatest advances in recent years (Goodfellow et al., 2016). Improvements in hardware, the availability of larger databases, and algorithmic advances have made possible to accurately build neural networks with more than one hidden layer. These deep neural networks have managed to solve problems that were previously unattainable in the fields of computer vision (Liu et al., 2020), natural language processing (Young et al., 2018) or speech recognition (Nassif et al., 2019).

However, in mental health, getting accurate results is not enough. Knowing the factors that characterize a given disease is often as important as precisely detecting those patients who suffer the condition. Identifying the relevant factors allows for improved treatments and prevention measures. One scientific field that has put a lot of effort in creating explainable models is the area of statistics. Among the techniques developed in this field, the lasso algorithm is undoubtedly one of the most widely used (Tibshirani, 1996). The lasso algorithm performs variable selection by including a regularization term in the loss function of the linear regression. The importance of this technique can be observed in the several extensions that have been developed to deal with variable selection. These techniques, which modify the loss function, include Elastic-Net (Zou and Hastie, 2003, 2005; Witten et al., 2014), Group Lasso (Zhou and Zhu, 2010; Zhao et al., 2014) or recently Sparse Group Lasso (Simon et al., 2013; Vincent and Hansen, 2014; Rao et al., 2016; Laria et al., 2019). However, in order to use these techniques, a database with a moderate/large size is needed since the regularization parameter is estimated through crossvalidation. Having a database of this size is not always possible in the mental health field.

The proposed dlasso algorithm, a neuronal network with two hidden layers and a regularization term, combines the advantages of lasso and the neural networks. On the one hand, like the lasso linear model, dlasso performs variable selection and provides the weights associated with each selected variable. This makes the neural network explainable. On the other hand, the weights of the neural network are trained through the backpropagation algorithm which, unlike traditional lasso, makes to automatically find the optimal value of the regularization parameter possible. Therefore, dlasso proposes to be a bridge that connects the prominent area of neural networks with that of modern statistics to obtain mutual benefits.

The rest of this article is organized as follows. Next section introduces the dlasso technique. Section 3 compares the performance of dlasso with respect to the traditional non-neural lasso. Additionally, some technical details of our implementation are highlighted in this section. After showing dlasso's performance, also in this section, it is used to predict children's ADHD severity based on their parent's burden and to identify the factors that influence parents' assessment. Finally, section 4 includes a discussion of the implication of our findings for future research.

2. MODEL FORMULATION

Consider the usual linear lasso framework, where we have a data matrix $\boldsymbol{X} \in \mathbb{R}^{N\times p}$ containing N observations of dimension p, a response vector $\boldsymbol{y} \in \mathbb{R}^N$, and the objective is to find $\boldsymbol{\beta} \in \mathbb{R}^p$ that minimizes, for some λ, the objective function

$$\mathcal{L}(\boldsymbol{\beta},\lambda) = \sum_{i=1}^{N}\left(y_i - \sum_{j=1}^{p} X_{ij}\beta_j\right)^2 + \lambda||\boldsymbol{\beta}||_1. \tag{1}$$

where $||\boldsymbol{\beta}||_1 = \sum_{i=1}^{p} |\beta_i|$ is the L_1 norm.

The general approach chooses the value of λ that minimizes the quadratic error term in $\mathcal{L}$ on a separate dataset, using some type of cross-validation over a grid (see, for example, Friedman et al., 2010; Friedman, 2012). In order to develop our methodology, the following proposition provides an alternative definition of the lasso problem.

Proposition 1. The lasso problem Equation (1) is equivalent to,

$$\min_{\boldsymbol{w}} \left\{ \sum_{i=1}^{N}\left(y_i - \frac{\lambda_0}{\lambda}\sum_{j=1}^{p} X_{ij}w_j\right)^2 + \lambda_0||\boldsymbol{w}||_1 \right\}, \text{ with } \lambda, \lambda_0 > 0. \tag{2}$$

The proof of Proposition 1 is straightforward, taking $\omega = \lambda\boldsymbol{\beta}/\lambda_0$. Although there is also λ_0, the regularization hyper-parameter in Equation (2) is λ, because λ_0 is a fixed constant. The hyper-parameter λ_0 could be interpreted, to some extent, as an initial approximation of λ if we were solving the traditional lasso.

Based on Proposition 1, Problem Equation (2) can be formulated as the neural network of **Figure 1**, assuming that the weight $\gamma = \lambda_0/\lambda$ is constant. However, this neural representation is a more general approach than Equation (2), because the parameter γ is optimally selected as part of the training process, unlike previous methodologies that rely on some sort of cross validated set-up to select λ.

Regarding the weights' optimization, it is known that in the context of neural networks, L_1 regularization does not completely zero out the weights. This is because the neural network optimizer does not take into account the non-differentiability of the regularization term at $\omega_j = 0$. To carry out feature selection from the neuronal network perspective, a condition on the weights is imposed so that, after a given number of iterations,

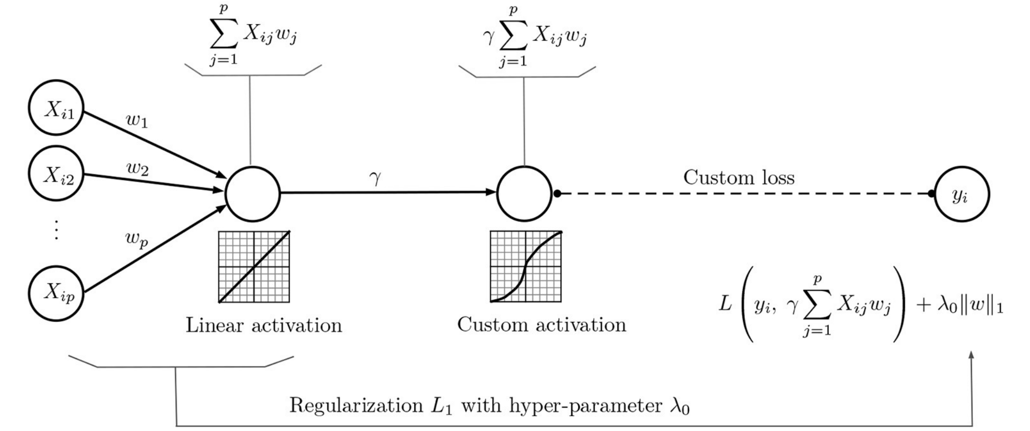
$\sum_{j=1}^{p} X_{ij} w_j$
$\gamma \sum_{j=1}^{p} X_{ij} w_j$
X_{i1}
X_{i2}
$\vdots$
X_{ip}
w_1
w_2
w_p
γ
Custom loss
y_i
Linear activation
Custom activation
$L\left(y_i,\ \gamma \sum_{j=1}^{p} X_{ij} w_j\right) + \lambda_0 \|w\|_1$
Regularization L_1 with hyper-parameter λ_0

FIGURE 1 | Neural network representation of the lasso problem Equation (2).

the weights that satisfy this condition are forced to be exactly 0. The mathematical derivation of the condition is presented from an optimization perspective, using subgradient conditions on problem Equation (2). General optimization problems where the objective function is the sum of a convex differentiable function (the squared error in this case) and a convex non-smooth part (the penalty) are discussed in Beck and Teboulle (2009a,b). In our case, it is simpler than that, since we are only interested in the condition that makes a particular $\omega_j = 0$. Notice that automatic feature selection in our context means a solution to Equation (2) where $\boldsymbol{w}$ has many components that are exactly zero.

Assume that we have some estimation of ω, γ, which is the result of optimizing the weights in the neural network of **Figure 1**, after a number of epochs. Focusing on ω, and letting γ fixed, we have,

$$\min_{\omega} F(\omega) := \left\| \boldsymbol{y} - \gamma \boldsymbol{X} \omega \right\|_2^2 + \lambda_0 \|\omega\|_1. \tag{3}$$

The optimality of any solution ω^* of Equation (3) is characterized by the subgradient conditions. That is, for every $j = 1, 2, \dots, p$,

$$0 = \partial_j F(\omega^*) = -2\gamma \boldsymbol{X}_j^T (\boldsymbol{y} - \gamma \boldsymbol{X} \omega^*) + \lambda_0 v_j, \tag{4}$$

where

$$v_j = \begin{cases} sign(\omega_j^*) & \omega_j^* \neq 0 \\ \in [-1, 1] & \omega_j^* = 0 \end{cases}.$$

In particular, $\omega_j^* = 0$ if

$$0 = -2\gamma \boldsymbol{X}_j^T \left(\boldsymbol{y} - \gamma \sum_{i=1, i \neq j}^{p} \boldsymbol{X}_i \omega_i^* \right) + \lambda_0 v_j, \quad |v_j| \leq 1,$$

which is equivalent to

$$\left| 2\boldsymbol{X}_j^\top \left(\boldsymbol{y} - \gamma \sum_{i=1, i \neq j}^{p} \boldsymbol{X}_i \omega_i^* \right) \right| \leq \left| \frac{\lambda_0}{\gamma} \right|. \tag{5}$$

Equation (5) provides a natural criterion to update the weights ω, trained after some number of epochs.

3. EXPERIMENTAL RESULTS

In this section, the performance of dlasso is evaluated in two different scenarios that have been previously used in the literature.

3.1. Experiment 1

This first experiment is based on the one conducted in the original lasso article (Tibshirani, 1996). The data is simulated from the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \epsilon$, where $\epsilon_i \sim N(0, 5)$ and

$$\boldsymbol{\beta} = [3\ 1.5\ 0\ 0\ 2\ \underbrace{0 \dots 0}_{p-5}].$$

The data matrix $\boldsymbol{X}$ is simulated so that the correlation between its columns $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ is given by $\rho_{ij} = 0.5^{|i-j|}$, for $1 \leq i < j \leq p$. To illustrate different configurations, the number of variables p varies in $\{20, 100, 200\}$. In addition, in order to obtain significant results, the simulation for each configuration is repeated 100 times.

The training data set was composed of 50 observations, whereas 950 observations were used to test the performance of lasso and dlasso. Unlike dlasso, which automatically tunes the regularization parameter along with the optimization of the coefficients, the $\boldsymbol{\beta}$ estimation provided by the lasso method depends on a user-supplied value of λ. This hyper-parameter was optimally selected using random search on a grid of size $1,000$, compared across 5 bootstrap repetitions of the training data. To fit the lasso model, and select the hyper-parameter λ, we used the following R libraries: `glmnet` (Friedman et al., 2010) for the model engine, `parsnip` (Kuhn and Vaughan, 2020) for the tidymodel interface, and `tune` (Kuhn, 2020) for tuning λ using random search.

Regarding the fit of the dlasso's parameters, the algorithm was trained for 1,000 epochs. In order to avoid initialization

TABLE 1 | Simulation results of the first experiment.

	Method	rmse	recall (β)	precision (β)
$p = 20$	dlasso	0.5 (0.08)	1 (0)	0.659 (0.18)**
	lasso	0.5 (0.08)	1 (0)	0.486 (0.19)
$p = 100$	dlasso	0.518 (0.08)**	0.99 (0.06)	0.348 (0.13)
	lasso	0.531 (0.09)	0.99 (0.06)	0.362 (0.22)
$p = 200$	dlasso	0.545 (0.11)**	1 (0)*	0.244 (0.08)
	lasso	0.561 (0.11)	0.99 (0.06)	0.313 (0.19)**

*Differences at the 0.05 and 0.01 significance levels are denoted by * and **, respectively.*

TABLE 2 | Simulation results of the second experiment.

	Method	rmse	recall (β)	precision (β)
			$\rho = 0.1$	
$p = 40$	dlasso	0.73 (0.1)	0.684 (0.1)	0.748 (0.1)**
	lasso	0.701 (0.11)**	0.804 (0.14)**	0.679 (0.11)
$p = 100$	dlasso	0.797 (0.1)**	0.612 (0.1)	0.526 (0.09)**
	lasso	0.838 (0.15)	0.624 (0.23)	0.469 (0.15)
$p = 400$	dlasso	0.875 (0.13)**	0.511 (0.12)**	0.275 (0.06)
	lasso	0.939 (0.16)	0.332 (0.24)	0.399 (0.2)**
			$\rho = 0.5$	
$p = 40$	dlasso	0.454 (0.07)	0.722 (0.09)	0.689 (0.08)**
	lasso	0.401 (0.05)**	0.784 (0.08)**	0.658 (0.07)
$p = 100$	dlasso	0.477 (0.08)	0.71 (0.11)	0.655 (0.09)**
	lasso	0.476 (0.14)	0.75 (0.12)**	0.463 (0.12)
$p = 400$	dlasso	0.494 (0.08)**	0.73 (0.12)	0.539 (0.1)**
	lasso	0.588 (0.27)	0.74 (0.17)	0.33 (0.17)
			$\rho = 0.8$	
$p = 40$	dlasso	0.344 (0.05)	0.772 (0.1)**	0.597 (0.07)
	lasso	0.302 (0.04)**	0.637 (0.09)	0.634 (0.08)**
$p = 100$	dlasso	0.356 (0.05)	0.747 (0.1)**	0.584 (0.08)**
	lasso	0.347 (0.1)	0.627 (0.13)	0.451 (0.13)
$p = 400$	dlasso	0.368 (0.06)**	0.782 (0.09)**	0.558 (0.06)**
	lasso	0.425 (0.22)	0.652 (0.18)	0.354 (0.18)

*Differences at the 0.05 and 0.01 significance levels are denoted by * and **, respectively.*

dependence, the training process was repeated 20 times and the network with the minimum training loss (mean squared error) was selected as the final model.

Table 1 summarizes the simulation results for Experiment 1, displaying the root mean squared error (rmse), the $\boldsymbol{\beta}$ recall (proportion of true non-zero coefficients correctly identified) and the $\boldsymbol{\beta}$ precision (proportion of zero coefficients correctly identified), of both lasso and dlasso. The values reported in **Table 1** are averaged over the 100 repetitions, with the corresponding standard deviations in parenthesis. We performed paired t-tests to evaluate the significance of the differences between both methods, denoting with $(**)$ those for which the p-value was lower than 0.01 and $(*)$ if the p-value was lower than 0.05. This table shows that the proposed dlasso method obtains a lower rmse than lasso in all the scenarios. Furthermore, this difference seems to be accentuated as the number of noise variables in the problem increases.

3.2. Experiment 2

The second experimental set-up is based on the simulation studies carried out by Witten et al. (2014). The data is simulated according to the linear model $\boldsymbol{y} = \boldsymbol{X\beta} + \epsilon$, with the number of features p, and ϵ_i i.i.d. from a $N(0, 10)$ distribution $(1 \leq i \leq n)$. The data matrix $\boldsymbol{X}$ is simulated from a multivariate $N(\mathbf{0}, \boldsymbol{\Sigma})$ distribution, where $\boldsymbol{\Sigma} \in R^{p\times p}$ is block diagonal, given by

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_\rho & 0 & 0 \\ 0 & \boldsymbol{\Sigma}_\rho & 0 \\ 0 & 0 & 0 \end{bmatrix}_{p\times p},$$

with $\boldsymbol{\Sigma}_\rho \in R^{20\times 20}$ such that

$$\boldsymbol{\Sigma}_\rho(i,j) = \begin{cases} 1 & i = j \\ \rho & i \neq j \end{cases},$$

with ρ denoting the correlation inside groups. The true coefficient vector $\boldsymbol{\beta} \in R^p$ is also random, given by,

$$\boldsymbol{\beta} = [\beta_1\ \beta_2\ \dots\ \beta_{10}\ \ \underbrace{0\ \dots\ 0}_{10}\ \ \beta_{21}\ \beta_{22}\ \dots\ \beta_{30}\ \ \underbrace{0\ \dots\ 0}_{p-30}],$$

where

$$\beta_j \sim \begin{cases} U[0.9, 1.1], & 1 \leq j \leq 10 \\ U[-1.1, -0.9], & 21 \leq j \leq 30 \end{cases}.$$

To explore different scenarios, the parameter $\rho \in \{0, 0.5, 0.8\}$, whereas $p \in \{40, 100, 400\}$, resulting in a total of 9 possible configurations. Similarly to the previous experiment, the data is composed of 50 and 950 observation for training and test, respectively, and the simulation for each configuration is repeated 100 times. The estimation of $\boldsymbol{\beta}$ and the selection of λ is carried out as described in the previous experiment.

Table 2 illustrates the results that were obtained by each methods in the different configurations. As before, the values reported in **Table 2** are averaged over the 100 repetitions, with the corresponding standard deviations in parenthesis, and significant differences at level 0.01 are denoted with $(**)$. It is observed that dlasso obtains better results as the number of variables increases, which suggests that dlasso might be a more suitable approach in the high-dimensional setting.

4. COMPUTATIONAL ISSUES

The dlasso algorithm has been implemented in R, using the `keras` library (Allaire and Chollet, 2019), and TensorFlow as backend (Abadi et al., 2016). Concerning the optimizer, we have chosen ADAM (Kingma and Ba, 2014), but our theoretical model formulation is not limited to a particular optimizer, and in this regard, different configurations may be tested.

An important computational issue is how to set the number of iterations to train the network, since a very large number can degrade its performance on future data. **Figure 2** shows how the rmse evolves in the test data over the different iterations for one of the simulations of the first experiment (**Figure 2A**)

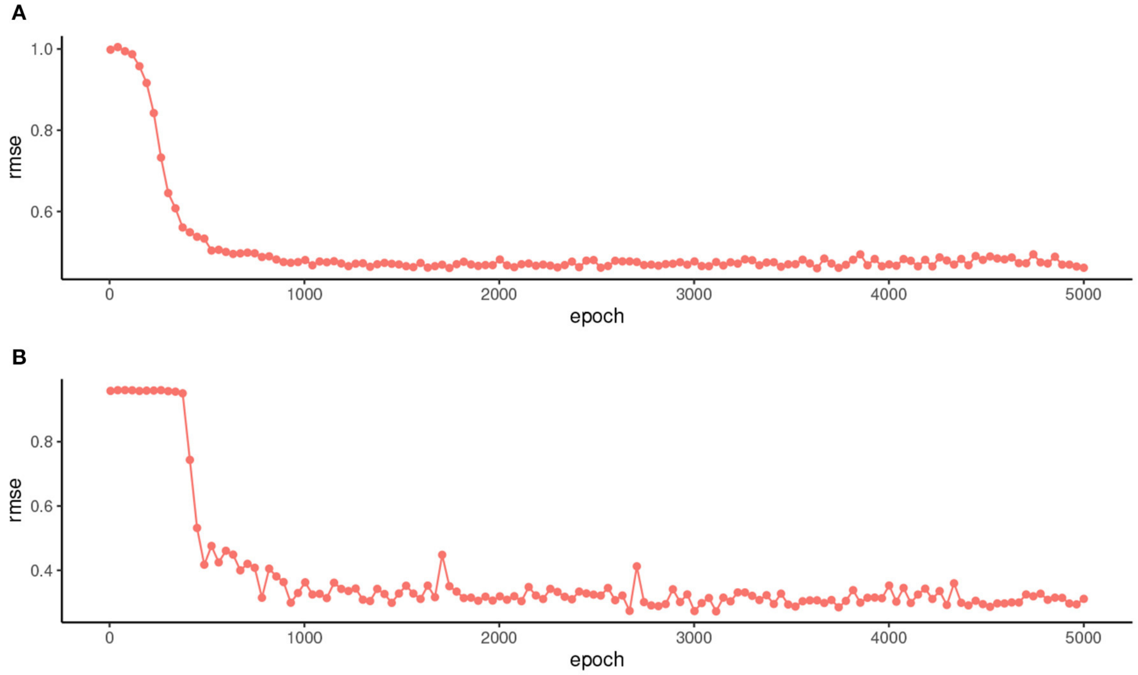

FIGURE 2 | Evolution of the rmse in the test data over 5,000 iterations for one of the simulations of the first experiment **(A)** and for one of the second experiment **(B)**.

and for one of the second experiment (**Figure 2B**). These two plots are representative of the different simulations. In both cases, no performance drop is observed in the test data as the number of iterations increases. This shows that the proposed method is quite robust to variations in the number of epochs and that, for example, 1,000 iterations is an acceptable value in both experiments.

4.1. Experiment 3

Once that the performance of the proposed dlasso technique has successfully been evaluated, this third experiment examines whether it is able to predict the assessments that parents make about the severity of their children's ADHD symptoms based on their distress and family burden.

To this end, the parents of 73 children diagnosed with ADHD by the medical professionals at the Fundacioń Jimeńez Díaz Hospital of Madrid participated in this study. The parents (54 mothers and 19 fathers) were required to sign an informed consent after the study was explained in detail to them. The mean age of their children was 12.4 years. The consent form and the study protocol were reviewed and approved by the Institutional Review Board of Fundacioń Jimeńez Díaz Hospital (reference: EO 77/2013_FJD_HIE_HRJC).

The participants assessed the severity of their children's symptoms through the Strengths and Weaknesses of ADHD-symptoms and Normal-behavior (SWAN) scale (Swanson et al., 2012). The SWAN scale is composed of 18 items, based on the DSM-5 criteria, for ADHD diagnosis which measure positive attention and impulse regulation behaviors in the normal population. The first nine items measure inattention while the remaining nine assess impulsivity/hyperactivity. The sum of the first nine items is used as an indicator of the severity of inattention, while the sum of the last nine items is used as an indicator of the severity of impulsivity/hyperactivity. The family burden and distress were assessed with the following scales:

- **The Zarit Burden Interview.** This 22-item self-report inventory examines the burden associated with functional/behavioral impairments and the home care situation (Zarit and Zarit, 1987; Schreiner et al., 2006). The responses ranged from never to always. It has been pointed out that this instrument has an excellent internal consistency (Bédard et al., 2001).
- **The General Health Questionnaire-12 (GHQ-12)**. This questionnaire, consisting on 12 items, measures general mental health (Anjara et al., 2020). The responses ranged from much worse than usual to better than usual. Sanchez-Lopez and Dresch showed that, in a Spanish sample of 1001 participants, the GHQ-12 exhibited an adequate reliability and external validity (Sánchez-López and Dresch, 2008). They also indicated that the GHQ-12 is an efficient technique to detect non-psychotic psychiatric problems.
- **The family Adaptability, Partnership, Growth, Affection, and Resolve (Apgar) scale.** This five-item scale is used to assess how family members perceive the level of functioning of the family unit (Smilkstein, 1978). Responses ranged from hardly ever to almost always.
- **Visual Analoge Scale (Vas) on life satisfaction.** This is an *ad hoc* questionnaire designed by the Department of

Translational Psychiatry of the Fundación Jiménez Díaz Hospital and which is part of the electronic questionnaires administered by a digital tool. Parents were asked to rate their own level of satisfaction in different life areas: themselves, family, friends, work and leisure activities. Parents scored these aspects on a scale from 0 to 10, where a higher number means more satisfaction.

Once the data were collected, a repeated validation analysis was conducted to test whether the dlasso algorithm was able to predict the severity of parent-reported child inattention, via the SWAN scale, based on the responses that they provided to the 44 previous predictors. The number of repetitions was set to 100. For each repetition, and similarly to the previous experiments, the training set contained 50 randomly selected observations. The remaining 23 observations were used to evaluate the performance of dlasso. The number of training epochs was 1,000. The performance measures were the rmse and the correlation between the total scores obtained with the SWAN inattention subscale and the values predicted by dlasso. The average correlation obtained by dlasso was 0.34 (std: 0.15) and the rmse was 1.99 (std: 2.87). These results indicate that dlasso was able to obtain good estimates of parents' assessment of their children's inattention. In addition, it was also observed that the correlation and rmse that would have been obtained if a multiple linear regression was applied were –0.01 (std: 0.23) and 4.14 (std: 14.12), respectively. These numbers reflect the importance of conducting feature election.

The previous repeated validation study was replicated, but using parent-reported child hyperactivity as the dependent variable. However, unlike the previous results, dlasso was not able to accurately estimate hyperactivity with these predictors. The average correlation obtained by dlasso was 0.03 (std: 0.19) and the rmse was 2.35 (std: 3.69). Similar results would have been obtained if all predictors were included in the multiple linear regression. Concisely, the average correlation would have been –0.05 (std: 0.20) and the rmse 4.34 (std: 17.6). These results show that, for our data, it is possible to estimate the degree of children's inattention through parental reported distress and family burden, but not the severity of children's hyperactivity.

After evaluating the performance of the proposed technique, a third analysis was carried out to determine which items dlasso used to estimate inattention. To do this, the dlasso was run on the whole sample and with all the predictor variables. The different predictors were standardized so that the selected items could be compared based on the absolute value of their weights. The variables selected were:

- GHQ-12, Item 1: Have you recently been able to concentrate on what you're doing?
- GHQ-12, Item 4: Have you recently felt capable of making decisions about things?
- Zarit, Item 20: Do you feel you should be doing more for your relative?
- Zarit, Item 22: How burdened do you feel in caring for your relative?
- Vas, Item 2: Satisfaction with Family
- Vas, Item 3: Satisfaction with Friends
- Vas, Item 5: Satisfaction with Leisure Activities

and the adjusted R^2 was 0.314. It is observed that only 7 of the 44 items were used to make the predictions. It is also noted that no items of the Apgar scale were selected. The values of the weights of the selected items are shown in **Figure 3**.

This figure shows that the most influential items are the second item of the Vas scale and item 20 of the Zarit scale. Both items are related to family satisfaction. It is also observed that three of the five items of the Vas scale, about the person's satisfaction with friends, pleasure activities and family, are selected.

5. CONCLUSIONS

In this article, the dlasso method has been proposed, implementing the well-known lasso feature selection technique using neural networks. The performance of the proposed dlasso has been assessed in two experiments previously referenced in the literature. In most of the conducted simulations, the proposed dlasso has attained a lower rmse, and significant higher precision and recall in the variable selection than the traditional lasso. Moreover, the simulation studies reveal that the gap between dlasso and its traditional counterpart widens as the number of variables increases.

In a third experiment, dlasso was used to predict the severity of symptoms in children with ADHD from the responses provided by their parents to four questionnaires aimed at measuring family burden, family functioning, parental satisfaction, and parental mental health. It was observed that dlasso was able to predict the severity of inattention using only seven items out of the 44 available. Interestingly, three of these seven items were obtained from the life satisfaction scale. Specifically, it was observed that higher parental satisfaction in essential domains such as family, friends and leisure activities are good predictors of inattentive symptomatology in children. The remaining four items are related to anxious symptomatology. Another noteworthy issue is that the algorithm did not select any of the items of the Apgar scale, which implies that family functioning is not taken into account in predicting inattention. This result complements the findings reported in the literature that show parental stress as one of the factors of disagreement among informants. Specifically, van der Oord et al. showed that parental stress explained 12% of the variance in the disagreement of parent and teacher ratings of inattention (van der Oord et al., 2006). Subsequently, Yeguez and Sibley showed that parental stress, maternal education, and maternal ADHD predicted high maternal grades relative to teacher-reported (Yeguez and Sibley, 2016). van der Veen-Mulders et al. also showed that differences in ratings between fathers and mothers were due to parental stress (van der Veen-Mulders et al., 2017).

The results obtained, together with those reported in the literature, raise the question of whether the stress reported by parents is mostly caused by their children's symptoms

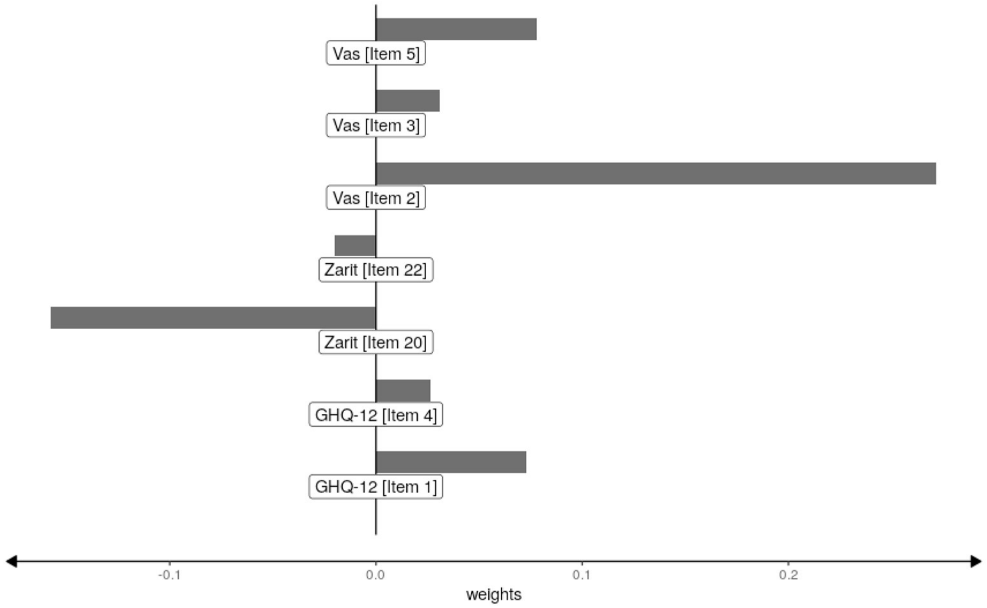

FIGURE 3 | Weights of the selected items.

or whether it is caused by external factors. Several studies have pointed to the first hypothesis. In this case, parental stress could be used as an excellent predictor of the severity of their children's ADHD symptoms. However, on the other hand, stress could also be caused mostly by external factors. Therefore, these results point to the need to establish mechanisms that identify the source of parental stress so that the relevance of the evaluation carried out by them can be assessed.

Regarding hyperactivity/impulsivity, dlasso was not able to obtain accurate estimates. Children diagnosed with ADHD inattentive subtype are usually diagnosed later than those with hyperactive/impulsive and/or combined subtypes (Milich et al., 2001). This may lead to significant family overload, psychological distress and poorer family functioning.

These results build a bridge between statistical and artificial intelligence approaches that allows tackling mental health conditions in which large samples are difficult to obtain.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the data are protected by the hospital as they refer to minors. Requests to access the datasets should be directed to ebacgar2@yahoo.es.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of Fundacioń Jimeńez Díaz as meta Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

JL and DD-G: conceptualization and formal analysis. JL, DD-G, and RL: methodology. JL: software. JL, DD-G, RL, and IP-C: writing–original draft preparation. RL and EB-G: supervision. DD-G: funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research has been partially supported by the Spanish National Project No. RTI2018-101857-B-I00. This research was also partly financed by the Community of Madrid in the framework of the multi-annual agreement with the University Carlos III Madrid in its line of action Excellence for University Teaching Staff. They are established within the framework of the V Regional Plan for Scientific Research and Technological Innovation 2016–2020.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). "Tensorflow: a system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (Savannah, GA), 265–283.

Allaire, J., and Chollet, F. (2019). *keras: R Interface to 'Keras'*. R package version 2.2.5.0.

Anjara, S., Bonetto, C., Van Bortel, T., and Brayne, C. (2020). Using the ghq-12 to screen for mental health problems among primary care patients: psychometrics and practical considerations. *Int. J. Mental Health Syst.* 14, 1–13. doi: 10.1186/s13033-020-00397-0

Beck, A., and Teboulle, M. (2009a). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* 2, 183–202. doi: 10.1137/080716542

Beck, A., and Teboulle, M. (2009b). "Gradient-based algorithms with applications to signal recovery," in *Convex Optimization in Signal Processing and Communications*, (Cambridge: Cambridge University Press), 42–88. doi: 10.1017/CBO9780511804458.003

Bédard, M., Molloy, D. W., Squire, L., Dubois, S., Lever, J. A., and O'Donnell, M. (2001). The zarit burden interview: a new short version and screening version. *Gerontologist* 41, 652–657. doi: 10.1093/geront/41.5.652

Chen, Y.-C., Hwang-Gu, S.-L., Ni, H.-C., Liang, S. H.-Y., Lin, H.-Y., Lin, C.-F., et al. (2017). Relationship between parenting stress and informant discrepancies on symptoms of adhd/odd and internalizing behaviors in preschool children. *PLoS ONE* 12:e0183467. doi: 10.1371/journal.pone.0183467

Chi, T. C., and Hinshaw, S. P. (2002). Mother–child relationships of children with adhd: The role of maternal depressive symptoms and depression-related distortions. *J. Abnormal Child Psychol.* 30, 387–400. doi: 10.1023/A:1015770025043

Elkins, I. J., McGue, M., and Iacono, W. G. (2007). Prospective effects of attention-deficit/hyperactivity disorder, conduct disorder, and sex on adolescent substance use and abuse. *Arch. General Psychiatry* 64, 1145–1152. doi: 10.1001/archpsyc.64.10.1145

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22. doi: 10.18637/jss.v033.i01

Friedman, J. H. (2012). Fast sparse regression and classification. *Int. J. Forecast.* 28, 722–738. doi: 10.1016/j.ijforecast.2012.05.001

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge: MIT press.

Harpin, V., Mazzone, L., Raynaud, J., Kahle, J., and Hodgkins, P. (2016). Long-term outcomes of adhd: a systematic review of self-esteem and social function. *J. Atten. Disord.* 20, 295–305. doi: 10.1177/1087054713486516

Harpin, V. A. (2005). The effect of adhd on the life of an individual, their family, and community from preschool to adult life. *Arch. Dis. Child.* 90(Suppl. 1), i2–i7. doi: 10.1136/adc.2004.059006

Harvey, E. A., Fischer, C., Weieneth, J. L., Hurwitz, S. D., and Sayer, A. G. (2013). Predictors of discrepancies between informants' ratings of preschool-aged children's behavior: an examination of ethnicity, child characteristics, and family functioning. *Early Child. Res. Q.* 28, 668–682. doi: 10.1016/j.ecresq.2013.05.002

Kingma, D. P. and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kuhn, M. (2020). *Tune: Tidy Tuning Tools*. R package version 0.0.1.

Kuhn, M., and Vaughan, D. (2020). *Parsnip: A Common API to Modeling and Analysis Functions*. R package version 0.0.5.

Laria, J. C., Carmen Aguilera-Morillo, M., and Lillo, R. E. (2019). An iterative sparse-group lasso. *J. Comput. Graph. Stat.* 28, 1–10. doi: 10.1080/10618600.2019.1573687

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., et al. (2020). Deep learning for generic object detection: A survey. *Int. J. Comput. Vis.* 128, 261–318. doi: 10.1007/s11263-019-01247-4

Madsen, K. B., Rask, C. U., Olsen, J., Niclasen, J., and Obel, C. (2020). Depression-related distortions in maternal reports of child behaviour problems. *Eur. Child Adolesc. Psychiatry* 29, 275–285. doi: 10.1007/s00787-019-01351-3

Milich, R., Amy C. Balentine, A., and Lynam, D. (2001). Adhd combined type and adhd predominantly inattentive type are distinct and unrelated disorders. *Clin. Psychol. Sci. Pract.* 8, 463–488. doi: 10.1093/clipsy.8.4.463

Nassif, A. B., Shahin, I., Attili, I., Azzeh, M., and Shaalan, K. (2019). Speech recognition using deep neural networks: a systematic review. *IEEE Access* 7, 19143–19165. doi: 10.1109/ACCESS.2019.2896880

Rao, N., Nowak, R., Cox, C., and Rogers, T. (2016). Classification with the sparse group lasso. *IEEE Trans. Signal Proc.* 64, 448–463. doi: 10.1109/TSP.2015.2488586

Sánchez-López, M. and Dresch, V. (2008). The 12-item general health questionnaire (ghq-12): reliability, external validity and factor structure in the spanish population. *Psicothema* 20, 839–843. Available online at: http://www.psicothema.com/psicothema.asp?id=3564, http://www.psicothema.com/pdf/3564.pdf

Schreiner, A. S., Morimoto, T., Arai, Y., and Zarit, S. (2006). Assessing family caregiver's mental health using a statistically derived cut-off score for the zarit burden interview. *Aging Mental Health* 10, 107–111. doi: 10.1080/13607860500312142

Schultz, B. K., and Evans, S. W. (2012). Sources of bias in teacher ratings of adolescents with adhd. *J. Educ. Dev. Psychol.* 2:151. doi: 10.5539/jedp.v2n1p151

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2013). A sparse-group lasso. *J. Comput. Graph. stat.* 22, 231–245. doi: 10.1080/10618600.2012.681250

Smilkstein, G. (1978). The family apgar: a proposal for a family function test and its use by physicians. *J. Fam Pract.* 6, 1231–1239.

Sonuga-Barke, E. J., Koerting, J., Smith, E., McCann, D. C., and Thompson, M. (2011). Early detection and intervention for attention-deficit/hyperactivity disorder. *Exp. Rev. Neurother.* 11, 557–563. doi: 10.1586/ern.11.39

Swanson, J. M., Schuck, S., Porter, M. M., Carlson, C., Hartman, C. A., Sergeant, J. A., et al. (2012). Categorical and dimensional definitions and evaluations of symptoms of adhd: history of the snap and the swan rating scales. *Int. J. Educ. Psychol. Assess.* 10:51. Available online at: https://sites.google.com/site/tijepa2012/articles/vol-10-1

Thomas, R., Sanders, S., Doust, J., Beller, E., and Glasziou, P. (2015). Prevalence of attention-deficit/hyperactivity disorder: a systematic review and meta-analysis. *Pediatrics* 135, e994–e1001. doi: 10.1542/peds.2014-3482

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B Methodol.* 58, 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x

van der Oord, S., Prins, P. J., Oosterlaan, J., and Emmelkamp, P. M. (2006). The association between parenting stress, depressed mood and informant agreement in adhd and odd. *Behav. Res. Therapy* 44, 1585–1595. doi: 10.1016/j.brat.2005.11.011

van der Veen-Mulders, L., Nauta, M. H., Timmerman, M. E., van den Hoofdakker, B. J., and Hoekstra, P. J. (2017). Predictors of discrepancies between fathers and mothers in rating behaviors of preschool children with and without adhd. *Eur. Child Adolesc. Psychiatry* 26, 365–376. doi: 10.1007/s00787-016-0897-3

Vincent, M., and Hansen, N. R. (2014). Sparse group lasso and high dimensional multinomial classification. *Comput. Stat. Data Anal.* 71, 771–786. doi: 10.1016/j.csda.2013.06.004

Wender, P. H., and Tomb, D. A. (2016). *ADHD: A Guide to Understanding Symptoms, Causes, Diagnosis, Treatment, and Changes Over Time in Children, Adolescents, and Adults*. Oxford: Oxford University Press.

Witten, D. M., Shojaie, A., and Zhang, F. (2014). The cluster elastic net for high-dimensional regression with unknown variable grouping. *Technometrics* 56, 112–122. doi: 10.1080/00401706.2013.810174

Yeguez, C. E., and Sibley, M. H. (2016). Predictors of informant discrepancies between mother and middle school teacher adhd ratings. *School Mental Health* 8, 452–460. doi: 10.1007/s12310-016-9192-1

Young, T., Hazarika, D., Poria, S., and Cambria, E. (2018). Recent trends in deep learning based natural language processing. *iEEE Comput. Intell. Mag.* 13, 55–75. doi: 10.1109/MCI.2018.2840738

Zarit, S., and Zarit, J. (1987). *Instructions for the Burden Interview*. University Park: Pennsylvania State University.

Zhao, W., Zhang, R., and Liu, J. (2014). Sparse group variable selection based on quantile hierarchical lasso. *J. Appl. Stat.* 41, 1658–1677. doi: 10.1080/02664763.2014.888541

Zhou, N., and Zhu, J. (2010). Group variable selection via a hierarchical lasso and its oracle property. *Stat. Its Interface* 3, 557–574. doi: 10.4310/SII.2010.v3.n4.a13

Zou, H., and Hastie, T. (2003). Regression shrinkage and selection via the elastic net, with applications to microarrays. *J. R. Stat. Soc. Ser. B* 67, 301–320.

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 67, 301–320. doi: 10.1111/j.1467-9868.2005.00503.x

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.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 09 September 2021
doi: 10.3389/fncom.2021.594659

Visual Explanation for Identification of the Brain Bases for Developmental Dyslexia on fMRI Data

Laura Tomaz Da Silva[1]*, *Nathalia Bianchini Esper*[2,3], *Duncan D. Ruiz*[1], *Felipe Meneguzzi*[1] *and Augusto Buchweitz*[3,4]

[1] School of Technology, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil, [2] Graduate School of Medicine, Neurosciences, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil, [3] BraIns, Brain Institute of Rio Grande do Sul, Porto Alegre, Brazil, [4] School of Health and Life Sciences, Psychology, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil

Problem: Brain imaging studies of mental health and neurodevelopmental disorders have recently included machine learning approaches to identify patients based solely on their brain activation. The goal is to identify brain-related features that generalize from smaller samples of data to larger ones; in the case of neurodevelopmental disorders, finding these patterns can help understand differences in brain function and development that underpin early signs of risk for developmental dyslexia. The success of machine learning classification algorithms on neurofunctional data has been limited to typically homogeneous data sets of few dozens of participants. More recently, larger brain imaging data sets have allowed for deep learning techniques to classify brain states and clinical groups solely from neurofunctional features. Indeed, deep learning techniques can provide helpful tools for classification in healthcare applications, including classification of structural 3D brain images. The adoption of deep learning approaches allows for incremental improvements in classification performance of larger functional brain imaging data sets, but still lacks diagnostic insights about the underlying brain mechanisms associated with disorders; moreover, a related challenge involves providing more clinically-relevant explanations from the neural features that inform classification.

Methods: We target this challenge by leveraging two network visualization techniques in convolutional neural network layers responsible for learning high-level features. Using such techniques, we are able to provide meaningful images for expert-backed insights into the condition being classified. We address this challenge using a dataset that includes children diagnosed with developmental dyslexia, and typical reader children.

Results: Our results show accurate classification of developmental dyslexia (94.8%) from the brain imaging alone, while providing automatic visualizations of the features involved that match contemporary neuroscientific knowledge (brain regions involved in the reading process for the dyslexic reader group and brain regions associated with strategic control and attention processes for the typical reader group).

Conclusions: Our visual explanations of deep learning models turn the accurate yet opaque conclusions from the models into evidence to the condition being studied.

Keywords: visual explanation, deep learning, dyslexia, neuroimaging, fMRI

OPEN ACCESS

Edited by:
Ali Fatemi,
Kennedy Krieger Institute,
United States

Reviewed by:
Keri Shiels Rosch,
Kennedy Krieger Institute,
United States
Saman Sargolzaei,
University of Tennessee at Martin,
United States

****Correspondence:***
Laura Tomaz Da Silva
laura.tomaz@edu.pucrs.br

Received: *14 August 2020*
Accepted: *08 June 2021*
Published: *09 September 2021*

Citation:
Tomaz Da Silva L, Esper NB, Ruiz DD, Meneguzzi F and Buchweitz A (2021) Visual Explanation for Identification of the Brain Bases for Developmental Dyslexia on fMRI Data. Front. Comput. Neurosci. 15:594659. doi: 10.3389/fncom.2021.594659

1. INTRODUCTION

Developmental dyslexia is a neurodevelopmental disorder that presents with persistent difficulty to read fluently and accurately; it is not related to intelligence, lack of educational opportunities or inadequate schooling and affects between 5 and 17% of children (American Psychiatric Association, 2013). Dyslexia is typically diagnosed after about 2–3 years of formal schooling (2nd or 3rd grade), after a child has failed to learn to read. In lower socioeconomic status (SES) countries, studies suggest an even older age at diagnosis among poor children (e.g., 10–11 years; Costa et al., 2015; Buchweitz et al., 2019). However, current neurobiological studies suggest that the functional and anatomical bases associated with dyslexia predate reading instruction (Gabrieli, 2009; Raschle et al., 2014; Ozernov-Palchik and Gaab, 2016; Centanni et al., 2019). In this sense, early identification of developmental dyslexia could help ameliorate the poor mental health and educational outcomes associated with the disorder (Sanfilippo et al., 2020).

There is an emerging consensus about the alterations in brain structure and function associated with dyslexia: functional MRI (fMRI) studies have shown left-lateralized, hypoactivation of posterior brain systems in the ventral occipitotemporal and temporoparietal regions; these regions are part of the brain's network that adapts to reading, in typical readers. The findings of hypoactivation in dyslexia in these posterior brain areas, relative to consistent activation in typical reading, have been replicated across fMRI studies in different languages (Paulesu et al., 2001; Seki et al., 2001; Kronbichler et al., 2006; Cattinelli et al., 2013; Cao et al., 2017; Buchweitz et al., 2019). On the other hand, a typical reader usually shows consistent activation of these occipitotemporal and parietotemporal posterior brain systems; these regions become functionally and morphologically integrated with the areas of the brain that are hardwired for spoken language as one learns to read (Pugh et al., 1996; Michael et al., 2001; Shaywitz et al., 2004; Buchweitz et al., 2009; Rueckl et al., 2015). The adaptations of these posterior brain regions represent brain markers of reading development, and their hypoactivation and altered function, markers of dyslexia. As markers of risk for dyslexia, understanding how these regions function and adapt can potentially inform earlier identification of risk for dyslexia and better understanding of reading treatment response (Gabrieli, 2009; Van Den Bunt et al., 2018).

Distinct brain imaging techniques such as structural MRI, fMRI, and diffusion-weighted imaging (DWI) are applied to investigate altered cortical tissue, structure and function associated with mental health and neurodevelopmental disorders (Atluri et al., 2013). These techniques allow for the identification of neural markers, which in turn may provide or inform a diagnosis based on image features (American Psychiatric Association, 2013).

Recent advances in deep learning have led researchers to employ machine learning to automate the analysis of medical imaging, including neurological images (Craddock et al., 2009; Froehlich et al., 2014; Tamboer et al., 2016). The most successful technique derived from deep learning for image classification consists of building neural network with convolutional layers, i.e., Convolutional Neural Networks (CNNs). The CNN specializes in processing multiple arrays, such as images (2D), audio and video or volumetric data (3D) (Bengio et al., 2015).

Brain imaging volumes have tens of thousands of voxels (3D-pixel) per image. Neurofunctional indices are mapped to these voxels, which makes feature selection a challenge for most machine learning approaches. Supervised approaches to machine learning relied on experts for feature selection (Bengio et al., 2015). Deep learning approaches obviate the dependence on supervision by automatically learning the features that better represent the problem domain (Bengio et al., 2015). Before deep learning methods were effectively applied to classification of brain imaging data, support vector machine (SVM) algorithms were the frequent choice for machine learning analyses of brain imaging (Cortes and Vapnik, 1995). SVM algorithms have the ability to generalize well in smaller fMRI datasets (Craddock et al., 2009; Buchweitz et al., 2012; Froehlich et al., 2014; Li et al., 2014; Tamboer et al., 2016; Just et al., 2017), which are typically in the dozens of participants due to the high costs of fMRI scans (Craddock et al., 2009; Froehlich et al., 2014). Moreover, SVM models trained with linear kernels offer relatively straightforward explanations. This SVM characteristic may be useful to break the "curse of dimensionality" by reducing the risk of overfitting the training data. The number of voxels used in feature selection should be reduced as much as possible.

Feature selection for brain imaging data is often performed on voxels in anatomically or functionally defined regions-of-interest (ROIs) based on the literature (Wolfers et al., 2015) or by data-driven methods that establish clusters of stable voxels (Shinkareva et al., 2008; Just et al., 2014). By contrast, deep learning models learn feature hierarchies at several levels of abstraction, which allows the system to learn complex functions independent of human-crafted features (Bengio et al., 2015). CNNs are applicable to a variety of medical image analysis problems, such as disorder classification (Heinsfeld et al., 2018), anatomy or tumor segmentation (Kamnitsas et al., 2017), lesion detection and classification (Ghafoorian et al., 2017), survival prediction (van der Burgh et al., 2017), and medical image construction (Li et al., 2014). Although these models can be accurate, their conclusions are opaque to human understanding and lack a straightforward explanation to help diagnosis. The provision of tools for healthcare practitioners to apply and trust the results of machine learning models of brain imaging to assist them in their clinical diagnoses is a challenge for brain imaging and machine learning research alike. Providing accurate visual representation of neural networks involved in deep learning classification may be a step in the direction of improving diagnostic application of classification using neurofunctional indices. For example, the prediction of brain states at slice level, and the subsequent generation of more fine-grained information about the features relevant for classification, can help improve interpretability (Ballester et al., 2021).

The goal of the present study is to integrate feature visualization techniques for CNNs. The key contribution is a visual representation of the regions involved in classifying whether children are dyslexic or not. This provides a better understanding of CNN behavior and may provide practitioners with a tool to glean neural alterations associated with a disorder from functional brain imaging scans.

2. MATERIALS AND METHODS

2.1. Data

The brain imaging data was collected as part of a research initiative to investigate the neural underpinnings of dyslexic children in Brazil. The participants were diagnosed with dyslexia following a multidisciplinary evaluation that included medical history, reading and writing tests (Costa et al., 2015; Toazza et al., 2017), and an IQ test (Wechsler Abbreviated Scale of Intelligence, Wechsler, 2012). The reading and other tests applied are described elsewhere (Costa et al., 2015; Buchweitz et al., 2019); in the interest of providing comprehensive information about the participants, see also the **Supplementary Materials** (supplementary information about participants and instruments).

2.1.1. Participants

The present study included 32 children who were divided into two groups: typical readers (TYP; $n = 16$) and dyslexic readers (DYS; $n = 16$) (Buchweitz et al., 2019). The participants were all monolingual speakers of Portuguese and right-handed. The two groups were matched for age, sex and IQ [age 8−12 (9 ± 1.39)]. The typical readers were scanned during the 2015 school year; the dyslexic children were scanned between 2014 and 2015 (Buchweitz et al., 2019). **Table 1** summarizes the complete demographics on this dataset. As indicate above, see also the **Supplementary Materials** for additional information on participants.

2.1.2. Word-Reading Task

Task based fMRI examines brain regions whose activity changes from baseline in response to the performance of a task or stimulus (Petersen and Dubis, 2012). The study was designed as a mixed event-related experiment using a word and pseudoword reading test validated for Brazilian children (Salles et al., 2013). The task consisted of 20 regular words, 20 irregular words, and 20 pseudowords. The 60 stimuli were divided into two 30-item runs to give the participants a break halfway into the task. Words and pseudowords were presented on the screen one at a time for 7 s each. A question was presented to participants along with each word ("Does the word exist?"), to which participants had to select "Yes" or "No" by pressing response buttons. After 10 trials (10 words) either a baseline condition or rest period was inserted in the experimental paradigm. The baseline condition consisted of presentation of a plus sign "+" in the middle of the screen for 30 s, during which participants were instructed to relax and clear their minds (Buchweitz et al., 2019).

2.1.3. Data Acquisition

Data was collected on a GE HDxT 3.0 T MRI scanner with an 8-channel head coil (Buchweitz et al., 2019). The following MRI sequences were acquired: a T1 structural scan (TR/TE = 6.16/2.18 ms, isotropic 1 mm^3 voxels); two task-related 5-min 26-s functional fMRI EPI sequences; and a 7-min resting state sequence. The task and the resting-state EPI sequences used the following parameters: TR = 2,000 ms, TE = 30 ms, 29 interleaved slices, slice thickness = 3.5 mm; slice gap = 0.1 mm; matrix size = 64 × 64, FOV = 220 × 220 mm, voxel size = 3.44 × 3.44 × 3.60 mm (Buchweitz et al., 2019).

2.1.4. Data Preprocessing

The preprocessing steps for the task-based (word-reading task) fMRI are described as follows. Word-reading task: preprocessing included slice-time and motion correction, smoothing with a 6 mm FWHM Gaussian kernel, and a nonlinear spatial normalization to 3.0 × 3.0 × 3.0 mm voxel template (HaskinsPedsNL template). TRs with motion outliers (>0.9 mm) were censored from the data. The criteria for exclusion due to head motion were: excessive motion in 20% of the TRs. The average head motion for each group for the participants included in the study, in the word-reading paradigm, was: DYS M = 0.16 ± 0.08, TYP M = 0.18 ± 0.15 (Buchweitz et al., 2019).

First level analysis included modeling regressor for the conditions for each of the three types of word (regular words, irregular words and pseudowords), and for the fixation condition. As a final preprocessing step, we averaged the words activation, and used this average as an input to the deep learning models. **T**-test analysis ($3dttest++$) were carried out to compare the distribution of activation between the two groups using a random-effects model and the contrast images for all the word types vs. fixation. Participant age was entered as a covariate in the analysis between groups to control for any effects due to the average 1-year difference in age between the groups. **Table 1** shows the demographics for the dataset used in this study. The accuracy and response time, during the MRI exam, were statistically significant between groups.

2.2. Classification Task

We trained a number of deep learning models for the classification task using two key recent techniques in learning for image classification: CNNs (LeCun et al., 1998) and data augmentation (Perez and Wang, 2017). For the CNNs, we evaluated both two-dimensional (2D) and three-dimensional (3D) CNNs.

First, regarding the CNNs, our choice of model focuses on 2D CNNs due to the size of our dataset. Specifically, 2D CNNs have a smaller number of parameters in comparison to 3D CNNs (Szegedy et al., 2015). Thus, training a 3D CNN necessitates substantially larger datasets in order to generalize well. Indeed, our experiments show that 2D CNNs achieve superior accuracy to 3D CNNs given the limitations of our dataset. A 2D CNN takes an input having three dimensions (a height h, a width w, and a number of color channels or a depth d). This input volume is then processed by k filters, which operate on the entire volume of feature maps that have been generated at a particular layer. 2D convolutions have a pseudo third dimension comprising the color channels in each image, such that a 2D CNN applies convolutions to each channel separately, combining the resulting activations. **Figure 1** illustrates each RGB channel in the input as a slice. A filter, which corresponds to weights in the convolutional layer, is

TABLE 1 | Demographic information of the study subjects.

	Typical readers	Dyslexic readers	*p*-value[a]
No. of subjects	16	16	
Age			
Mean ± STD	8.44 ± 0.51	9.63 ± 0.88	<0.001
Range	8–9	8–12	
Sex (Male/Female)	9 / 7	11 / 5	
IQ			
Mean ± STD	102.73 ± 15.37	107.85 ± 26.6	NS
Range	71–127	88–144	
Socioeconomic status (SES) (mean ± STD)	24.1 ± 4.9	26.4 ± 6.5	
Reading speed—words per minute (mean ± STD)	84.71 ± 31.89	13.07 ± 7.68	<0.001
Average motion—fMRI Task (mean ± STD)	0.17 ± 0.15	0.26 ± 0.08	NS
fMRI task—accuracy (mean ± STD)	54.68 ± 6.71	35.06 ± 14	<0.001
fMRI task—response time (mean ± STD)	2039.2 ± 423.56	2981.06 ± 954.82	NS

[a] *Independent samples t-test; STD, standard deviation; NS, not significant.*

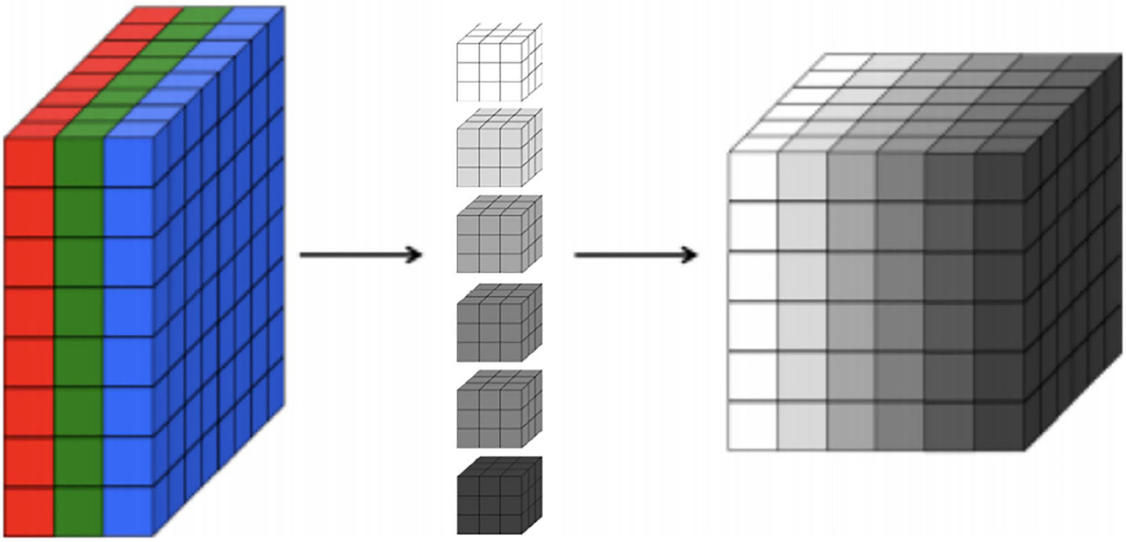

FIGURE 1 | A 3D representation of a convolutional layer, where each RGB channel in the input is a colored slice. We show six filters with the same depth of the input in the middle and, on the right, we show the output activations of combined convolution operations where each slice in the output corresponds to each filter.

then multiplied with a local portion of the input to produce a neuron in the next volumetric layer of neurons. In the **Figure 1**, the middle part represents filters, the depth of the filter corresponds to the depth of the input. The last cube in the figure represents the output activations of the combined convolution operations for each channel. The depth of the output volume of a convolutional layer is equivalent to the number of filters in that layer, that is, each filter produces its own slice. This can be viewed as using a 3D convolution for each output channel, which happens to have the same depth as the input (Buduma and Locascio, 2017). For this reason, it is possible to use volumetric images as inputs to a 2D CNN. In effect, this means that a 2D CNN processes the 3D volume of brain scan activations slice-by-slice.

Second, we avoid overfitting in our small dataset by employing data augmentation. Data augmentation is a technique (Perez and Wang, 2017) that provides the model with more data to increase the model's ability to generalize from it. Such techniques are already employed in several image problems in deep learning models, but are still incipient in fMRI data (Mikołajczyk and Grochowski, 2018).

We adopted two approaches to build the 2D CNN architectures: (i) use genetic programming, more specifically grammar-based genetic programming (GGP) fitted to our problem; and (ii) employ a modified version of the LeNet-5 (LeCun et al., 1998) classification model. We then trained the resulting architecture using our dataset, and compared the effectiveness of 3D convolutions by converting the generated 2D

CNNs into 3D ones by swapping the 2D convolutional layers to appropriately-sized 3D convolutions.

2.3. Visual Explanations Task

While many application areas for machine learning focus simply on model performance, recent work has highlighted the need for explanations for the decisions of trained models. Most users of machine learning often want to understand the trained models in order to gain confidence in the predictions. This is especially true for machine learning models used in medical applications, where the consequences of each decision must be carefully explained to patients and other stakeholders (Yang et al., 2018; Jin et al., 2019). Besides the explainability aspect required of direct medical applications, our key motivation is to allow neuroimaging specialists to derive new insights on underpinnings of specific learning disorders such as dyslexia. Indeed, clinical diagnosis of dyslexia is reliable and costs less than using fMRI scans to validate such diagnostics (Torgesen, 1998; Ramus et al., 2003). However, researchers of dyslexia are interested in further understanding of the disorder and its neural underpinnings *in-vivo* (Shaywitz et al., 2001; Hoeft et al., 2011). For this reason, building data-driven diagnostics models via machine learning and generating explanations for such models can be an invaluable tool for dyslexia research.

Recently researchers developed several methods for understanding and visualizing CNNs, in part as a response to criticism that the learned features in a neural network are not interpretable to humans (Zeiler and Fergus, 2013; Szegedy et al., 2014; Zhou et al., 2016). A category of techniques that aim to help understand which parts of an image a CNN model uses to infer class labels is called Class Activation Mapping (CAM) (Zhou et al., 2016). CAM produces heatmaps of class activations over input images. A class activation heatmap is a 2D grid of scores associated with a particular output class, computed for every location for an input image, indicating how important each location is with respect to that output class (Zhou et al., 2016). CAM can be used by a restricted class of image classification CNNs, precluding the model from containing any fully-connected layers and employing global average pooling (GAP).

A recent approach to visualize features learned by a CNN is GRAD-CAM (Selvaraju et al., 2017). GRAD-CAM is a generalization of CAM and can be applied to a broader range of CNN models without the need to change their architecture. Instead of trying to propagate back the gradients, GRAD-CAM infers a downsampled relevance heatmap of the input pixels from the activation heatmaps of the final convolutional layer. The downsampled heatmap is upsampled to obtain a coarse relevance heatmap. This approach has two key advantages: first, it can be applied to any CNN architecture; and second, it requires no re-training or change in the existing neural network architecture.

Figure 2 illustrates the GRAD-CAM approach. Given an image and a class of interest (in the example, "dyslexic reader") as input, GRAD-CAM first forward propagates the image through the CNN part of the model and then through task-specific computations to obtain a raw score for the category. Next, GRAD-CAM sets all the gradients that do not belong to the desired class (dyslexic reader), which are originally set to one, are set to zero. GRAD-CAM then backpropagates this signal to the rectified convolutional feature maps of interest, which it combines to compute the coarse GRAD-CAM localization (the bottom heatmap in the figure) representing where the model has to look to make the particular decision. Finally, GRAD-CAM pointwise multiplies the heatmap with guided backpropagation to get Guided GRAD-CAM visualizations which are both high-resolution and concept-specific (Selvaraju et al., 2017).

3. EXPERIMENTS AND RESULTS

3.1. Classification

The deep learning classification model was implemented using the Keras open source library (Chollet et al., 2015) and trained with an Nvidia Geforce GTX 1080 Ti graphical processing unit (GPU) with 12 GB of memory. In our GGP approach, we generated a population of CNN architectures, such that each CNN architecture was an individual in a population, and was evaluated to produce a fitness value. Network topology for all CNNs generated was based on a specific grammar for our problem and a set of different hyperparameters.

We introduced four key modifications in our version of the LeNet-5 architecture. First, we added batch normalization layers in the convolutional layers to improve convergence and generalization (Ioffe and Szegedy, 2015). Second, we used RELU activations in the convolutional layers instead of tanh. Third, we changed the average pooling to max pooling in the subsampling layers. Finally, we used a dropout rate of 0.5 in the fully connected layer. **Figure 3** illustrates our modified version of LeNet-5. Our model architecture contains ∼175 K parameters, a small amount in comparison to deeper architectures, such as VGG-16 (Simonyan and Zisserman, 2014), which contains over 138 million parameters.

Our 3D CNN was developed based on our 2D CNN model. We made the changes necessary to adapt 2D convolutions, 2D pooling layers to a 3D model. In order to fit our data to a 3D CNN model, we expanded our data adding one channel for gray images resulting in a 4-dimensional array as input to the network. The resulting architecture has over 3 million parameters.

We compared our induced deep learning models with the SVM (Cortes and Vapnik, 1995) technique, which has been used in a substantial number of previous neuroimaging studies (Froehlich et al., 2014; Tamboer et al., 2016). Specifically, this technique is popular for fMRI applications because datasets typically have many features (voxels), but only a relatively small set of subjects.

We trained all models to classify the participants between dyslexic readers and typical readers using the Adam optimizer (Kingma and Ba, 2014). We improved the performance of our classifier by employing two data augmentations to our dataset: (i) we added Gaussian noise to fMRI images to generalize to noisy images; and (ii) we added a random Gaussian offset, or contrast, to increase differences between images. The input of our machine and deep learning models was the whole brain volume (60 × 73 × 60 voxels) and a binary mask filling the brain volume to retrieve data from all brain regions. All of our deep learning

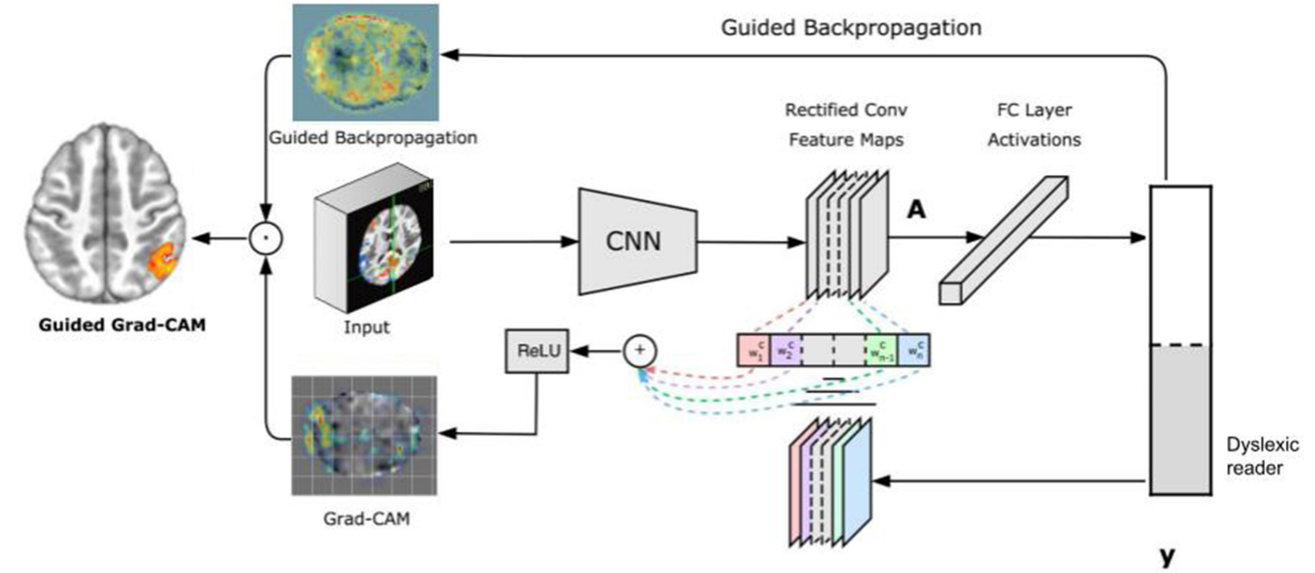

FIGURE 2 | GRAD-CAM overview.

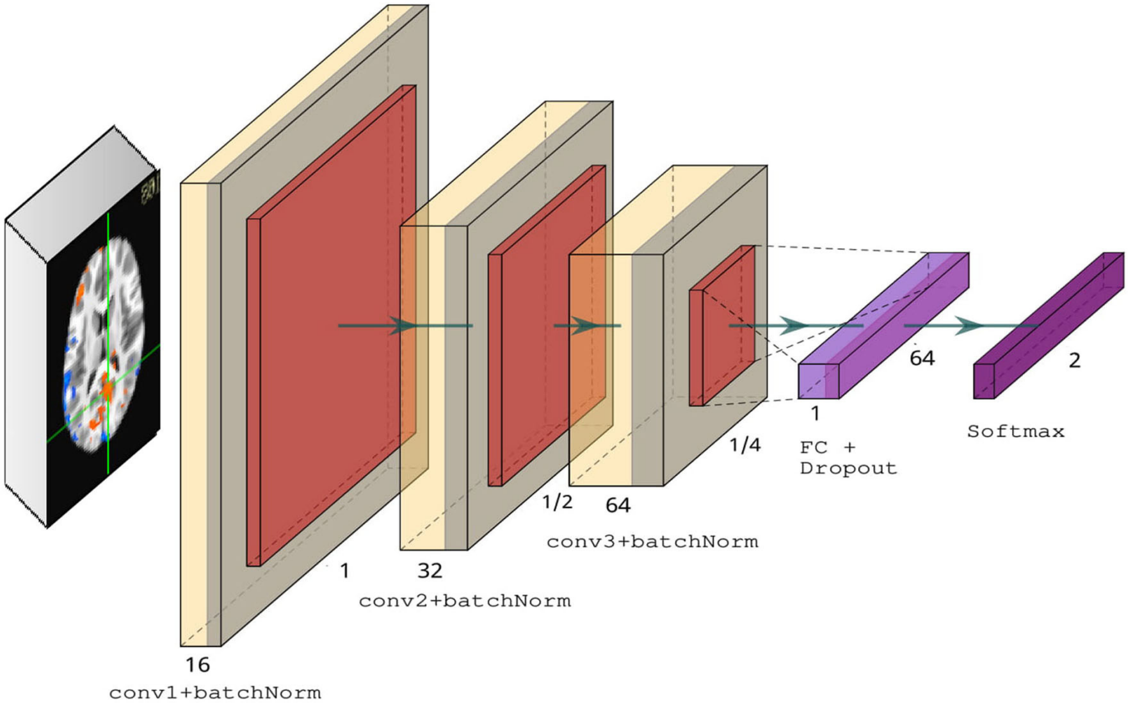

FIGURE 3 | Modified LeNet-5 overview containing three convolutional layers with ReLU activations, followed by a fully connected layer and dropout, and finally a softmax classifier.

models followed the same split, i.e., 80% train, 10% validation, and 10% test sets. The parameter values including learning rate, dropout rate, batch size, and epoch size were optimized using the ranges summarized in **Table 2**. Note that we optimized the batch size to use the maximum available GPU memory.

All hyperparameters were optimized for both the 2D and 3D CNN models. For our SVM models, first, we applied an exhaustive search over specified parameters values for our SVM estimator. Second, we evaluated different methods of cross-validation. We report the results from splitting the data into

TABLE 2 | CNN hyperparameters used to generate our GGP population of CNN architectures.

Hyperparameters	Values
Kernel size	Ranging from 1 to 5
No. of filters	Starts with 16; duplicates after every convolution
Stride	Ranging from 1 to 3
Learning rate	Logarithmic range of [1, 0.1, 0.01, 0.001, 0.0001, 0.00001]
Dropout rate	Tuned in the range of [0.1, 0.5, 1]
Batch size	16
No. of epochs	Tuned in the range of [10, 50, 100]
No. of Neurons FC layer	Tuned in the range of [32, 64, 128, 256, 512]

TABLE 3 | Summary of dyslexia classification results, including our Modified LeNet-5 architecture (2D and 3D) and our best GGP CNN.

Technique	Accuracy (%)
Best GGP 2D CNN	**94.83**
Modified LeNet-5	85.71
Best GGP 3D CNN	78.57
Modified LeNet-5 3D	71.43
SVM (80% train, 20% test)	70

train, validation, and test for Linear SVM implemented using scikit-learn (Pedregosa et al., 2011) library in Python.

Our modified version of LeNet-5 2D CNN network achieved 85.71% accuracy on subject classification. Our best GGP 2D CNN model achieved an accuracy of 94.83% on subject classification. In comparison to the 2D CNN architecture, the 3D CNN, from both the modified LeNet-5 and GGP approach, had an inferior accuracy on subject classification. The 3D CNN was also more prone to overfitting in the first few epochs of training. By contrast, the SVM approach achieved much lower classification accuracy, regardless of the training dataset composition. **Table 3** summarizes the results from all our classification approaches.

3.2. Visual Explanations

After training the 2D CNN model, we loaded the model with the best accuracy to visualize the learned gradients using GRAD-CAM technique (Selvaraju et al., 2017). The class activation generated by GRAD-CAM shows which regions were more instrumental to the classification.

To employ GRAD-CAM visualization to identify key differences between subjects and controls, we chose a pair of subjects as input, i.e., a control (non-dyslexic) subject and a dyslexic reader subject to generate the class activation mappings. **Figures 4A,B** show GRAD-CAM generated images of control and dyslexic readers subjects, with respect to the gradients learned by the network model. Both images depict the central slice from the axial view of the brain volume. Areas with lower class activation mappings are colored in gray, whereas areas with higher class activation mappings are color-coded from yellow (instrumental) to red (more instrumental). The color coding thus represents the brain regions impact on the model classification of subjects.

The visualization for the dyslexic readers group (**Figure 5**) showed frontal and temporal brain regions that are traditionally associated with reading processes, and also temporoparietal and dorsolateral prefrontal regions that are associated with increased working memory load, including during reading (Pugh et al., 1996; Chein and Schneider, 2005; Buchweitz et al., 2009, 2014; Waldie et al., 2013). Additional axial visualizations of brain regions can be found in **Supplementary Figure 2** showed bilateral inferior frontal gyrus (**Supplementary Figures 2A,C,D**), the parietal lobe (**Supplementary Figure 2F**) and the right temporal lobe (**Supplementary Figure 2E**) were some of the regions that presented high classification mapping in the group analysis. In addition to the frontal regions, the group analysis (**Figure 6**) showed that the left precuneus (**Supplementary Figure 3B**) and the right insula (**Supplementary Figure 3D**) were also among the regions with higher classification mapping for typical readers relative to dyslexic readers (Oh et al., 2014). **Tables 4**, **5** show the voxel count per brain region for visualization of the dyslexic readers group and for the typical readers group, respectively. For group-level analyses of brain activation differences between dyslexic readers and typical readers, please see Buchweitz et al. (2019), which included the same participants.

4. DISCUSSION AND RELATED WORK

To the best of our knowledge, there is little work on visual explanations and brain imaging; for instance, a recent study used these explanations for Alzheimer's disease (AD) and structural MRI (sMRI) (Jin et al., 2019). However, few approaches employed a visualization technique for MRI data, and there are none for fMRI data. The lack of approaches using brain imaging data of Dyslexia led us to search for related work employing deep learning to process any type of MRI data. **Table 6** summarizes previous work that employed deep learning (Sarraf and Tofighi, 2016; Heinsfeld et al., 2018; Jin et al., 2019) for subject classification, and approaches that applied machine learning to identify participants with dyslexia (Cui et al., 2016; Tamboer et al., 2016; Płoński et al., 2017).

The machine learning techniques we use in this article allow us to divide the related work into two types: (i) work that aimed to identify participants with dyslexia using traditional machine learning algorithms (e.g., SVM); and (ii) work that used Deep Neural Networks (DNNs) in brain imaging data for disease classification, as follows. Sarraf and Tofighi (2016) employed the LeNet-5 architecture to classify patients with Alzheimer's disease. Heinsfeld et al. (2018) used two stacked denoising autoencoders for the unsupervised pre-training stage to extract a lower-dimensional version of the ABIDE (Autism Brain Imaging Data Exchange) data. Jin et al. (2019) employed an attention-based 3D residual network based on the 3D ResNet to classify Alzheimer's Disease classification and to identify important regions in their visual explanation task. The remaining work applied machine learning techniques to classify dyslexic

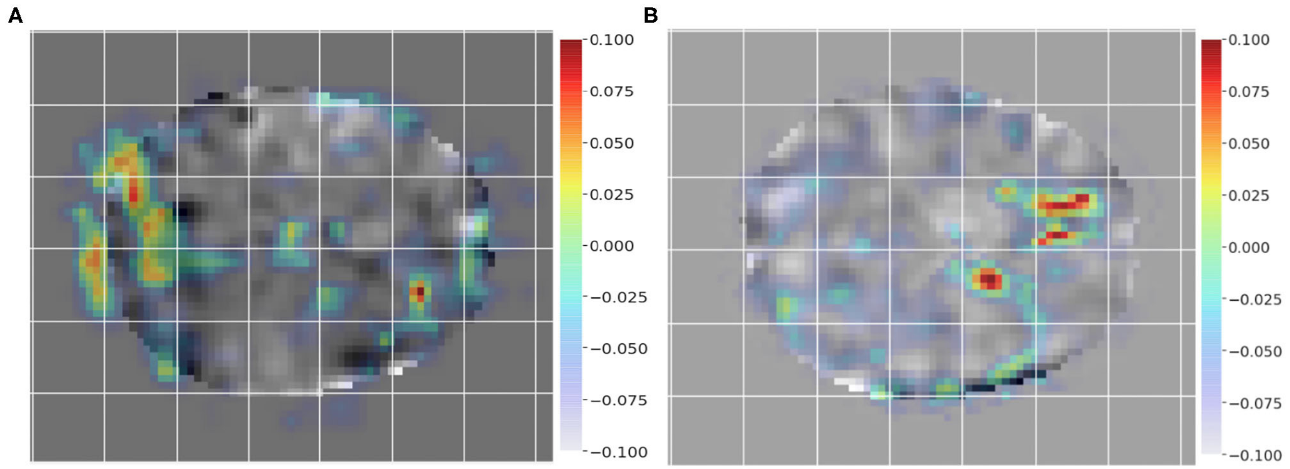

FIGURE 4 | Class activation mapping for a single dyslexic reader **(A)** and single typical reader **(B)** subject classification from GRAD-CAM technique. The visualization highlights areas with lower class activation colored from gray to light blue, whereas areas with higher class activation are colored from yellow to red.

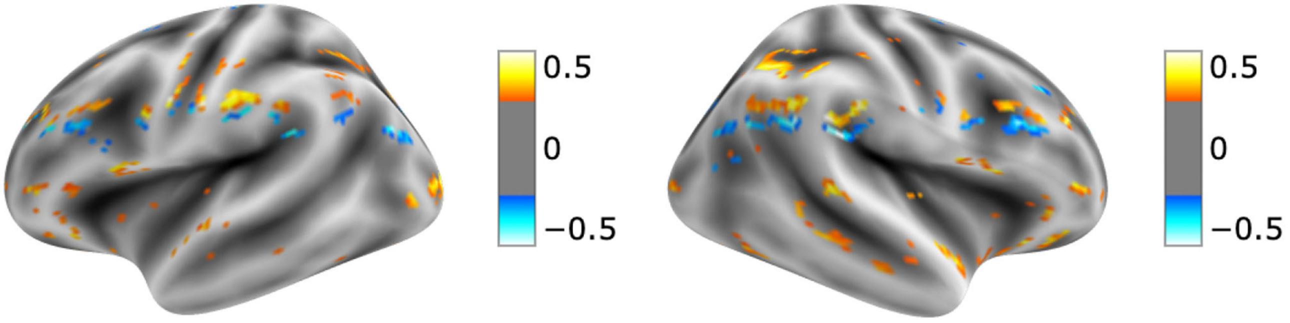

FIGURE 5 | Visual explanation for Dyslexic readers subjects. **Supplementary Material** contains axial images, instrumental brain regions for dyslexic readers identification summarized in **Table 4**. The left side of the images represents the left side of the brain. Surface images for left and right side of the brain showing the visual explanations at cortical level. AFNI (Cox, 1996) images showing brain activation from GRAD-CAM.

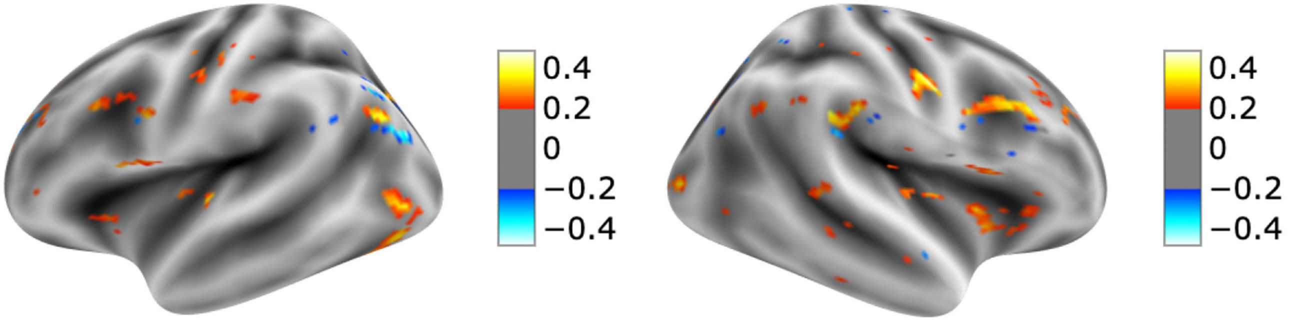

FIGURE 6 | Visual explanation for Typical readers subjects. **Supplementary Material** contains axial images, instrumental brain regions for Typical readers identification summarized in **Table 5**. The left side of the images represents the left side of the brain. Surface images for left and right side of the brain showing the visual explanations at cortical level. AFNI (Cox, 1996) images showing brain activation from GRAD-CAM.

readers and typical readers subjects. Tamboer et al. (2016) and Cui et al. (2016) used SVM. Płoński et al. (2017) on top of using SVM, also used logistic regression (LR), and random forest (RF).

Approaches that adopt deep learning models (Sarraf and Tofighi, 2016; Heinsfeld et al., 2018; Jin et al., 2019) show that DNN approaches can achieve competitive results using MRI

TABLE 4 | Voxel count per brain region of dyslexic readers for **Supplementary Figure 2**.

	Dyslexic readers		Peak of activation coordinates		
	Brain regions	**No. of voxels**	**x**	**y**	**z**
Supplementary Figure 2A	**Left inferior frontal gyrus**	**167**	**−40**	**27**	**27**
	Left rostral middle frontal	52			
	Left IFG[a] (pars opercularis)	20			
	Left postcentral	18			
	Left precentral	18			
	Left superior frontal	13			
	Left supramarginal	5			
	Left caudal middle frontal	5			
	White matter	36			
Supplementary Figure 2B	**Left superior frontal gyrus**	**123**	**−11**	**48**	**24**
	Left rostral middle frontal	43			
	Left IFG (pars opercularis)	23			
	Left superior frontal	19			
	Right superior frontal	9			
	Left precentral	4			
	Left caudal middle frontal	3			
	Right rostral middle frontal	3			
	Left caudal anterior cingulate	1			
	White matter	18			
Supplementary Figure 2C	**Right IFG (pars opercularis)**	**116**	**52**	**9**	**24**
	Right IFG (pars opercularis)	40			
	Right rostral middle frontal	21			
	Right precentral	14			
	Right postcentral	4			
	Right caudal anterior cingulate	2			
	Right IFG (pars triangularis)	1			
	White matter	34			
Supplementary Figure 2D	**Right IFG (pars triangularis)**	**98**	**49**	**24**	**30**
	Right rostral middle frontal	33			
	Right IFG (pars opercularis)	14			
	Right caudal middle frontal	12			
	Right precentral	11			
	Right postcentral	2			
	White matter	26			
Supplementary Figure 2E	**Right middle temporal**	**77**	**61**	**-8**	**-15**
	Right inferior temporal	30			
	Right middle temporal	22			
	Right superior parietal	11			
	White matter	14			
Supplementary Figure 2F	**Right angular**	**65**	**46**	**−57**	**45**
	Right inferior parietal	59			
	Right supramarginal	4			
	White matter	2			

Brain regions instrumental for dyslexic readers identification with Grad-CAM *(Selvaraju et al., 2017). Region labels follow Haskins pediatric atlas (Molfese et al., 2020).*
[a] *IFG, inferior frontal gyrus.*

and fMRI data. Heinsfeld et al. (2018) achieved state-of-the-art results with 70% accuracy in identification of ASD vs. control patients in the dataset. The authors that used classic machine learning techniques (Cui et al., 2016; Tamboer et al., 2016; Płoński et al., 2017) achieved 80, 83.6, and 65% accuracy respectively on dyslexia prediction from anatomical scans. Performance

TABLE 5 | Voxel count per brain region of typical readers for **Supplementary Figure 3**.

	Typical readers		**Peak of activation coordinates**		
	Brain regions	**No. of voxels**	**x**	**y**	**z**
Supplementary Figure 3A	**Right postcentral**	**201**	**43**	**–11**	**30**
	Right supramarginal	43			
	Right IFG[a] *(pars opercularis)*	29			
	Right caudal middle frontal	25			
	Right postcentral	17			
	Right precentral	14			
	Right supramarginal	8			
	Right inferior parietal	7			
	White matter	58			
Supplementary Figure 3B	**Left precuneus**	**89**	**–1**	**–68**	**39**
	Left precuneus	35			
	Left inferior parietal	25			
	Right precuneus	10			
	Left superior parietal	7			
	Left cuneus	2			
	Right cuneus	1			
	White matter	9			
Supplementary Figure 3C	**Left superior occipital**	**82**	**–16**	**-89**	**24**
	Left inferior parietal	51			
	Left lateral occipital	17			
	Left cuneus	7			
	White matter	7			
Supplementary Figure 3D	**Right insula**	**64**	**31**	**–23**	**-24**
	Right supramarginal	19			
	Right postcentral	9			
	Right insula	2			
	Right caudate	2			
	White matter	32			

Brain regions instrumental for typical readers identification with Grad-CAM *(Selvaraju et al., 2017). Region labels follow Haskins pediatric atlas (Molfese et al., 2020).*
[a] *IFG, inferior frontal gyrus.*

TABLE 6 | Comparison with the classification scores of related work.

Study references	**Modality**	**Dataset**	**Classifier**	**Task**	**Accuracy (%)**
Proposed method	Task based fMRI	ACERTA project	2D CNN	Subject classification for Dyslexia	94.83
Sarraf and Tofighi (2016)	rs-fMRI	ADNI[a]	LeNet-5	Subject classification for Alzheimer	96.86
Jin et al. (2019)	sMRI	ADNI[a]	Attention-based 3D ResNet	Subject classification for Alzheimer's Disease	92.1%
Cui et al. (2016)	sMRI	Private dataset	SVM	Subject classification for Dyslexia	83.6
Tamboer et al. (2016)	sMRI	Non-disclosed dataset	SVM	Subject classification for Dyslexia	80
Heinsfeld et al. (2018)	rs-fMRI	ABIDE	Denoising Autoencoder	Subject classification for Autism Spectrum Disorder	Above 70
Płoński et al. (2017)	sMRI	Private dataset	SVM, LR, RF	Subject classification for Dyslexia	65

[a] *adni.loni.usc.edu.*

of our deep learning models was consistent with other deep learning approaches for classification of neurological conditions. By contrast, our SVM results did not generalize as well as others (Cui et al., 2016; Tamboer et al., 2016; Płoński et al., 2017), but still outperformed another application of SVM for dyslexia classification (Płoński et al., 2017). Given the difference in datasets, accuracies obtained by our two approaches are not comparable to other ones.

Jin et al. (2019) visual explanations consisted of an attention map (much like a heatmap in visual representation) that indicated the significance of brain regions for AD classification. The authors compared their explanations to those generated by 3D-CAM and 3D-GRAD-CAM (Yang et al., 2018) methods. Jin et al. (2019) observed that these two 3D methods led to a substantial drop in model performance when classifying subjects for Alzheimer's Disease (AD) by the extra calculations needed to generate the heatmaps. By introducing the attention method, the authors obtained a 3D attention map for each testing sample and were able to identify the significance of brain regions related to changes in gray matter for AD classification. Our visualization technique may not be comparable to Jin et al. (2019), but the application of visualization techniques to medical imaging holds promise for making deep learning models interpretable.

5. CONCLUSION

We introduce a novel approach for the investigation of neural patterns in task-based fMRI that allow for the classification of dyslexic readers and typical readers. While deep learning classifiers provide accurate identification of dyslexic readers vs. typical readers based solely on their brain activation, such models are often hard to interpret. In this context, our main contribution is a visualization technique of the features that lead to specific classifications, which allows neuroscience domain experts to interpret the resulting models. Visual explanations of deep learning models allows us to compare regions instrumental to the classification with the latest neuroscientific evidence about dyslexia and the brain. The left occipital and inferior parietal regions that discriminated among groups are part of brain networks associated with phonological and lexical (word-level) processes in reading in different languages (Paulesu et al., 2001). Other regions reported in our visualization are also associated with reading and reading disorders (i.e., **Tables 4**, **5**). More activation of anterior right-hemisphere prefrontal regions (e.g., right pars triangularis) are associated with dyslexia and possible compensatory mechanisms (Vellutino et al., 2004; Shaywitz and Shaywitz, 2005).

Feature visualization techniques and visual explanations for deep learning models are a novel research area, and applying these techniques to neuroimaging data has the potential to help research in neuroscience. Our work offers encouraging results, since the brain areas identified by the visual explanations are consistent with neuroscientific knowledge about the neural correlates of dyslexia. There are a number of ways in which we could extend the present work. The deep learning classification models can be applied to publicly available, large fMRI or MRI datasets to investigate the areas that are instrumental for identification of, for example, autism spectrum disorder. Moreover, other visualization techniques can be applied to provide a qualitative comparison among techniques when used to illustrate machine learning and deep learning studies of brain imaging.

DATA AVAILABILITY STATEMENT

The code to reproduce our experiments is available at GitHub (https://github.com/lauratomaz/VisualExplanations). Data was provided by BraIns (Brain Institute of Rio Grande do Sul - Brazil) initiative to establish a brain database of dyslexic readers of Brazilian Portuguese. This data can be found here: https://inscer.pucrs.br/br/neuroimagem-da-cognicao-humana.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by PUCRS Ethics Committee. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

LT designed the study, implemented the framework, and ran experiments. FM and DR supervised the implementation and engineering of the work. AB and NE helped interpreting the findings and provided neuroimaging-related insights. All authors contributed to writing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fncom.2021.594659/full#supplementary-material

REFERENCES

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental disorders (DSM-5®)*. American Psychiatric Association.

Atluri, G., Padmanabhan, K., Fang, G., Steinbach, M., Petrella, J. R., Lim, K., et al. (2013). Complex biomarker discovery in neuroimaging data: finding a needle in a haystack. *NeuroImage* 3, 123–131. doi: 10.1016/j.nicl.2013.07.004

Ballester, P. L., da Silva, L. T., Marcon, M., Esper, N. B., Frey, B. N., Buchweitz, A., et al. (2021). Predicting brain age at slice level: convolutional neural networks and consequences for interpretability. *Front. Psychiatry* 12:518. doi: 10.3389/fpsyt.2021.598518

Bengio, Y., Goodfellow, I. J., and Courville, A. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Buchweitz, A., Costa, A. C., Toazza, R., de Moraes, A. B., Cara, V. M., Esper, N. B., et al. (2019). Decoupling of the occipitotemporal cortex and the brain's default-mode network in dyslexia and a role for the cingulate cortex in good readers: a brain imaging study of brazilian children. *Dev. Neuropsychol.* 44, 146–157. doi: 10.1080/87565641.2017.1292516

Buchweitz, A., Mason, R. A., Meschyan, G., Keller, T. A., and Just, M. A. (2014). Modulation of cortical activity during comprehension of familiar and unfamiliar text topics in speed reading and speed listening. *Brain Lang.* 139, 49–57. doi: 10.1016/j.bandl.2014.09.010

Buchweitz, A., Mason, R. A., Tomitch, L., and Just, M. A. (2009). Brain activation for reading and listening comprehension: an fMRI study of modality effects and individual differences in language comprehension. *Psychol. Neurosci.* 2, 111–123. doi: 10.3922/j.psns.2009.2.003

Buchweitz, A., Shinkareva, S. V., Mason, R. A., Mitchell, T. M., and Just, M. A. (2012). Identifying bilingual semantic neural representations across languages. *Brain Lang.* 120, 282–289. doi: 10.1016/j.bandl.2011.09.003

Buduma, N., and Locascio, N. (2017). *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*. Newton, MA: O'Reilly Media, Inc.

Cao, F., Yan, X., Wang, Z., Liu, Y., Wang, J., Spray, G. J., et al. (2017). Neural signatures of phonological deficits in chinese developmental dyslexia. *Neuroimage* 146, 301–311. doi: 10.1016/j.neuroimage.2016.11.051

Cattinelli, I., Borghese, N. A., Gallucci, M., and Paulesu, E. (2013). Reading the reading brain: a new meta-analysis of functional imaging data on reading. *J. Neurolinguist.* 26, 214–238. doi: 10.1016/j.jneuroling.2012.08.001

Centanni, T. M., Norton, E. S., Ozernov-Palchik, O., Park, A., Beach, S. D., Halverson, K., et al. (2019). Disrupted left fusiform response to print in beginning kindergartners is associated with subsequent reading. *NeuroImage* 22:101715. doi: 10.1016/j.nicl.2019.101715

Chein, J. M., and Schneider, W. (2005). Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Brain Res.* 25, 607–623. doi: 10.1016/j.cogbrainres.2005.08.013

Chollet, F., et al. (2015). *Keras.* Available online at: https://keras.io

Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018

Costa, A. C., Toazza, R., Bassoa, A., Portuguez, M. W., and Buchweitz, A. (2015). "Ambulatório de aprendizagem do projeto ACERTA (avaliação de crianças em risco de transtorno de aprendizagem): métodos e resultados em dois anos," in *Neuropsicologia do Desenvolvimento: Infância e Adolescência*, eds J. F. Salles, V. G. Haase, and L. Malloy-Diniz (Porto Alegre: Artmed), 151–158.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173. doi: 10.1006/cbmr.1996.0014

Craddock, R. C., Holtzheimer, P. E. III.„ Hu, X. P., and Mayberg, H. S. (2009). Disease state prediction from resting state functional connectivity. *Magn. Reson. Med.* 62, 1619–1628. doi: 10.1002/mrm.22159

Cui, Z., Xia, Z., Su, M., Shu, H., and Gong, G. (2016). Disrupted white matter connectivity underlying developmental dyslexia: a machine learning approach. *Hum. Brain Mapp.* 37, 1443–1458. doi: 10.1002/hbm.23112

Froehlich, C., Meneguzzi, F., Franco, A., and Buchweitz, A. (2014). "Classifying brain states for cognitive tasks: a functional mri study in children with reading impairments," in *Proceedings of the XXIV Brazilian Congress on, Biomedical Engineering* (Uberlandia), 2476–2479.

Gabrieli, J. D. (2009). Dyslexia: a new synergy between education and cognitive neuroscience. *Science* 325, 280–283. doi: 10.1126/science.1171999

Ghafoorian, M., Karssemeijer, N., Heskes, T., Bergkamp, M., Wissink, J., Obels, J., et al. (2017). Deep multi-scale location-aware 3D convolutional neural networks for automated detection of lacunes of presumed vascular origin. *NeuroImage* 14, 391–399. doi: 10.1016/j.nicl.2017.01.033

Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., and Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the abide dataset. *NeuroImage* 17, 16–23. doi: 10.1016/j.nicl.2017.08.017

Hoeft, F., McCandliss, B. D., Black, J. M., Gantman, A., Zakerani, N., Hulme, C., et al. (2011). Neural systems predicting long-term outcome in dyslexia. *Proc. Natl. Acad. Sci. U.S.A.* 108, 361–366. doi: 10.1073/pnas.1008950108

Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015*, Vol. 37, 448–456.

Jin, D., Xu, J., Zhao, K., Hu, F., Yang, Z., Liu, B., et al. (2019). "Attention-based 3d convolutional network for alzheimer's disease diagnosis and biomarkers exploration," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (Venice), 1047–1051.

Just, M. A., Cherkassky, V. L., Buchweitz, A., Keller, T. A., and Mitchell, T. M. (2014). Identifying autism from neural representations of social interactions: neurocognitive markers of autism. *PLoS ONE* 9:e113879. doi: 10.1371/journal.pone.0113879

Just, M. A., Pan, L., Cherkassky, V. L., McMakin, D. L., Cha, C., Nock, M. K., et al. (2017). Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. *Nat. Hum. Behav.* 1:911. doi: 10.1038/s41562-017-0234-y

Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., et al. (2017). Efficient multi-scale 3D CNN with fully connected crf for accurate brain lesion segmentation. *Med. Image Anal.* 36, 61–78. doi: 10.1016/j.media.2016.10.004

Kingma, D. P., and Ba, J. (2014). "Adam: a method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR*" .

Kronbichler, M., Hutzler, F., Staffen, W., Mair, A., Ladurner, G., and Wimmer, H. (2006). Evidence for a dysfunction of left posterior reading areas in german dyslexic readers. *Neuropsychologia* 44, 1822–1832. doi: 10.1016/j.neuropsychologia.2006.03.010

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791

Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., and Ji, S. (2014). "Deep learning based imaging data completion for improved brain disease diagnosis," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, eds P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe (Cham: Springer International Publishing), 305–312.

Michael, E. B., Keller, T. A., Carpenter, P. A., and Just, M. A. (2001). fmri investigation of sentence comprehension by eye and by ear: modality fingerprints on cognitive processes. *Hum. Brain Mapp.* 13, 239–252. doi: 10.1002/hbm.1036

Mikołajczyk, A., and Grochowski, M. (2018). "Data augmentation for improving deep learning in image classification problem," in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, (Świnouście) 117–122.

Molfese, P. J., Glen, D., Mesite, L., Cox, R. W., Hoeft, F., Frost, S. J., et al. (2020). The haskins pediatric atlas: a magnetic-resonance-imaging-based pediatric template and atlas. *Pediatr. Radiol.* 51, 628–639. doi: 10.1007/s00247-020-04875-y

Oh, A., Duerden, E. G., and Pang, E. W. (2014). The role of the insula in speech and language processing. *Brain Lang.* 135, 96–103. doi: 10.1016/j.bandl.2014.06.003

Ozernov-Palchik, O., and Gaab, N. (2016). Tackling the "dyslexia paradox:" reading brain and behavior for early markers of developmental dyslexia. *Wiley Interdisc. Rev.* 7, 156–176. doi: 10.1002/wcs.1383

Paulesu, E., Démonet, J.-F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., et al. (2001). Dyslexia: cultural diversity and biological unity. *Science* 291, 2165–2167. doi: 10.1126/science.1057179

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Perez, L., and Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv [preprint]. arXiv:1712.04621*.

Petersen, S. E., and Dubis, J. W. (2012). The mixed block/event-related design. *Neuroimage* 62, 1177–1184. doi: 10.1016/j.neuroimage.2011.09.084

Płoński, P., Gradkowski, W., Altarelli, I., Monzalvo, K., van Ermingen-Marbach, M., Grande, M., et al. (2017). Multi-parameter machine learning approach to the neuroanatomical basis of developmental dyslexia. *Hum. Brain Mapp.* 38, 900–908. doi: 10.1002/hbm.23426

Pugh, K. R., Shaywitz, B. A., Shaywitz, S. E., Constable, R. T., Skudlarski, P., Fulbright, R. K., et al. (1996). Cerebral organization of component processes in reading. *Brain* 119, 1221–1238. doi: 10.1093/brain/119.4.1221

Ramus, F., Pidgeon, E., and Frith, U. (2003). The relationship between motor control and phonology in dyslexic children. *J. Child Psychol. Psychiatry* 44, 712–722. doi: 10.1111/1469-7610.00157

Raschle, N. M., Stering, P. L., Meissner, S. N., and Gaab, N. (2014). Altered neuronal response during rapid auditory processing and its relation to phonological processing in prereading children at familial risk for dyslexia. *Cereb. Cortex* 24, 2489–2501. doi: 10.1093/cercor/bht104

Rueckl, J. G., Paz-Alonso, P. M., Molfese, P. J., Kuo, W.-J., Bick, A., Frost, S. J., et al. (2015). Universal brain signature of proficient reading: evidence from four contrasting languages. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15510–15515. doi: 10.1073/pnas.1509321112

Salles, J. F. d., Piccolo, L. d. R., Zamo, R. d. S., and Toazza, R. (2013). Normas de desempenho em tarefa de leitura de palavras/pseudopalavras isoladas (lpi) para crianças de 1º ano a 7º ano. *Est. Pesquis. Psicol.* 13, 397–419. doi: 10.12957/epp.2013.8416

Sanfilippo, J., Ness, M., Petscher, Y., Rappaport, L., Zuckerman, B., and Gaab, N. (2020). Reintroducing dyslexia: early identification and implications for pediatric practice. *Pediatrics* 146:e20193046. doi: 10.1542/peds.2019-3046

Sarraf, S., and Tofighi, G. (2016). Classification of alzheimer's disease using fMRI data and deep learning convolutional neural networks. *arXiv[Preprint].arXiv:1603.08631*.

Seki, A., Koeda, T., Sugihara, S., Kamba, M., Hirata, Y., Ogawa, T., et al. (2001). A functional magnetic resonance imaging study during sentence reading in japanese dyslexic children. *Brain Dev.* 23, 312–316. doi: 10.1016/S0387-7604(01)00228-5

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). "Grad-cam: visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision* (Venice), 618–626.

Shaywitz, B. A., Shaywitz, S. E., Blachman, B. A., Pugh, K. R., Fulbright, R. K., Skudlarski, P., et al. (2004). Development of left occipitotemporal systems for skilled reading in children after a phonologically-based intervention. *Biol. Psychiatry* 55, 926–933. doi: 10.1016/j.biopsych.2003.12.019

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Fulbright, R. K., Mencl, W. E., Constable, R. T., et al. (2001). The neurobiology of dyslexia. *Clin. Neurosci. Res.* 1, 291–299. doi: 10.1016/S1566-2772(01)00015-9

Shaywitz, S. E., and Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biol. Psychiatry* 57, 1301–1309. doi: 10.1016/j.biopsych.2005.01.043

Shinkareva, S. V., Mason, R. A., Malave, V. L., Wang, W., Mitchell, T. M., and Just, M. A. (2008). Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS ONE* 3:e1394. doi: 10.1371/journal.pone.0001394

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv [preprint]. arXiv:1409.1556*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA), 1–9.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., et al. (2014). "Intriguing properties of neural networks,, in 2nd International Conference on Learning Representations, ICLR, eds Y. Bengio and Y. LeCun.

Tamboer, P., Vorst, H., Ghebreab, S., and Scholte, H. (2016). Machine learning and dyslexia: Classification of individual structural neuro-imaging scans of students with and without dyslexia. *NeuroImage* 11, 508–514. doi: 10.1016/j.nicl.2016.03.014

Toazza, R., Costa, A., Bassôa, A., Portuguez, M. W., and Buchweitz, A. (2017). "Avaliação de dislexia do desenvolvimento no ambulatório de aprendizagem do projeto ACERTA," in *Guia de Boas Práticas: Do Diagnóstico à Intervenção de Pessoas Com Transtornos Específicos de Aprendizagem*, eds A. Navas, C. S. Azoni, D. G. Oliveira, J. P. Borges, and R. Mousinho (São Paulo: Instituto ABCD), 26–33.

Torgesen, J. (1998). *Catch Them Before They Fall: Identification and Assessment to Prevent Reading Failure in Young Children (On-Line).* National Institute of Child Health and Human Development, 1–15. Available online at: https://www.readingrockets.org/article/catch-them-they-fall-identification-and-assessment-prevent-reading-failure-young-children

Van Den Bunt, M., Groen, M. A., van der Kleij, S., Noordenbos, M., Segers, E., Pugh, K. R., et al. (2018). Deficient response to altered auditory feedback in dyslexia. *Dev. Neuropsychol.* 43, 622–641. doi: 10.1080/87565641.2018.1495723

van der Burgh, H. K., Schmidt, R., Westeneng, H.-J., de Reus, M. A., van den Berg, L. H., and van den Heuvel, M. P. (2017). Deep learning predictions of survival based on mri in amyotrophic lateral sclerosis. *NeuroImage* 13, 361–369. doi: 10.1016/j.nicl.2016.10.008

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., and Scanlon, D. M. (2004). Specific reading disability (dyslexia): what have we learned in the past four decades? *J. Child Psychol. Psychiatry Allied Discip.* 45, 2–40. doi: 10.1046/j.0021-9630.2003.00305.x

Waldie, K. E., Haigh, C. E., Badzakova-Trajkov, G., Buckley, J., and Kirk, I. J. (2013). Reading the wrong way with the right hemisphere. *Brain Sci.* 3, 1060–1075. doi: 10.3390/brainsci3031060

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence–4th Edn.* San Antonio, TX: The Psychological Corporation.

Wolfers, T., Buitelaar, J. K., Beckmann, C. F., Franke, B., and Marquand, A. F. (2015). From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neurosci. Biobehav. Rev.* 57, 328–349. doi: 10.1016/j.neubiorev.2015.08.001

Yang, C., Rangarajan, A., and Ranka, S. (2018). "Visual explanations from deep 3d convolutional neural networks for Alzheimer's disease classification," in *AMIA Annual Symposium Proceedings, Vol. 2018* (San Francisco, CA: American Medical Informatics Association), 1571.

Zeiler, M. D., and Fergus, R. (2013). "Visualizing and understanding convolutional networks," in *Computer Vision - ECCV 2014 - 13th European Conference*, Vol. 8689 (Springer), 818–833. doi: 10.1007/978-3-319-10590-1_53

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). "Learning deep features for discriminative localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 2921–2929.

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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Computational Neuroscience

ORIGINAL RESEARCH
published: 20 January 2022
doi: 10.3389/fncom.2021.789998

Deep Learning for Autism Diagnosis and Facial Analysis in Children

*Mohammad-Parsa Hosseini[1,2], Madison Beary[1], Alex Hadsell[1], Ryan Messersmith[1] and Hamid Soltanian-Zadeh[3]**

[1] Department of Bioengineering, Santa Clara University, Santa Clara, CA, United States, [2] AI Research, San Jose, CA, United States, [3] CIPCE, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

In this paper, we introduce a deep learning model to classify children as either healthy or potentially having autism with 94.6% accuracy using Deep Learning. Patients with autism struggle with social skills, repetitive behaviors, and communication, both verbal and non-verbal. Although the disease is considered to be genetic, the highest rates of accurate diagnosis occur when the child is tested on behavioral characteristics and facial features. Patients have a common pattern of distinct facial deformities, allowing researchers to analyze only an image of the child to determine if the child has the disease. While there are other techniques and models used for facial analysis and autism classification on their own, our proposal bridges these two ideas allowing classification in a cheaper, more efficient method. Our deep learning model uses MobileNet and two dense layers to perform feature extraction and image classification. The model is trained and tested using 3,014 images, evenly split between children with autism and children without it; 90% of the data is used for training and 10% is used for testing. Based on our accuracy, we propose that the diagnosis of autism can be done effectively using only a picture. Additionally, there may be other diseases that are similarly diagnosable.

Keywords: deep learning, autism, children, diagnosis, facial image analysis

OPEN ACCESS

Edited by:
Arman Sargolzaei,
Tennessee Technological University,
United States

Reviewed by:
Khaled Saad,
Assiut University Hospital, Egypt
Parthasakha Das,
Indian Institute of Engineering Science and Technology, India
Antonio Marcos Batista,
Universidade Estadual de Ponta Grossa, Brazil

****Correspondence:***
Hamid Soltanian-Zadeh
hszadeh@ut.ac.ir

Received: *05 October 2021*
Accepted: *26 November 2021*
Published: *20 January 2022*

Citation:
Hosseini M-P, Beary M, Hadsell A, Messersmith R and Soltanian-Zadeh H (2022) Deep Learning for Autism Diagnosis and Facial Analysis in Children. Front. Comput. Neurosci. 15:789998. doi: 10.3389/fncom.2021.789998

INTRODUCTION

Motivation

Autism is primarily a genetic disorder, though there are some environmental factors, that cause challenges with social skills, repetitive behaviors, speech, and non-verbal communication (Alexander et al., 2007). In 2018, the Centers for Disease Control and Prevention (CDC) claimed that about 1 in 59 children will be diagnosed with some form of autism. Because there are so many forms of autism, it is technically called autism spectrum disorder (ASD) (Austimspeaks, 2019). A child can be diagnosed with ASD as early as 18 months old. Interestingly, while ASD is believed to be a genetic disorder, it is mainly diagnosed through behavioral attributes: "the ways in which children diagnosed with ASD think, learn, and problem-solve can range from highly skilled to severely challenged." Early detection and diagnosis are crucial for any patient with ASD as this may significantly help them with their disorder.

We believe that facial recognition is the best possible way to diagnose a patient because of their distinct attributes. Scientists at the University of Missouri found that children diagnosed with autism share common facial feature distinctions from children who are not diagnosed with the disease (Aldridge et al., 2011; CBS News, 2017). The study found that children with autism have an unusually broad upper face, including wide-set eyes. They also have a shorter middle region of the face, including the cheeks and nose. **Figure 1** shows some of these differences. Because of this,

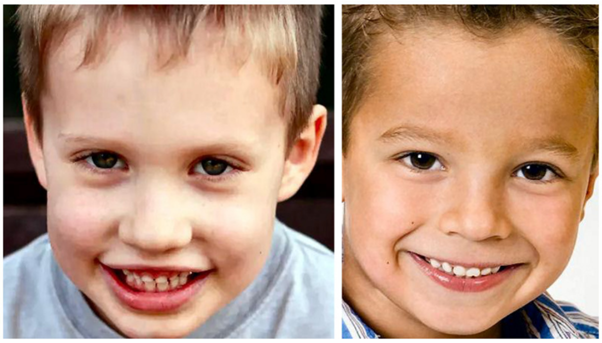

FIGURE 1 | On the **(Left)** is a child with autism, and on the **(Right)** is a child without autism, in order to compare some facial features.

conducting facial recognition binary classification on images of children with autism and children who are labeled as healthy could allow us to diagnose the disease earlier and in a cheaper way.

Data

In this work, we have used a dataset found on Kaggle, which consists of over three thousand images of both children with and without autism. This dataset is slightly unusual, as the publisher only had access to websites to gather all the images. When downloaded, the data is provided in two ways: split into training, testing, and validation vs. consolidated. If we decide to create our own machine learning model, the provided split of training, testing, and validating subgroups will be useful. The validation component will be important for determining the quality of the model we use, which means we do not have to strictly rely on the accuracy of the model to determine its quality. The training, testing, and validation subcategories are further split into autistic and non-autistic folders. The autistic training group consists of 1,327 images of facial images, and the non-autistic training directory consists of the same number of images. The autistic and non-autistic testing directories both have 140 images, for a total of 280 images. Lastly, the validation category has a total of 80 images: 40 facial images without autism and 40 with. If we can use a model already available, then using the consolidated images would be best, because that will allow us to control the amount used for training and testing (Gerry, 2020).

STATE-OF-THE-ART

There have been several studies conducted using neural networks for facial, behavioral, and speech analysis (Eni et al., 2020; Liang et al., 2021). Most of these studies have focused on determining the age and gender of the individual in question (Iga et al., 2003). Additionally, there have been a few studies done focusing on the classification of autism using brain imaging modalities. Our work has taken the techniques available for facial analysis and applied these to the classification of autism.

Facial Analysis

Wen-Bing Horng and associates (Horng et al., 2001) worked to classify facial images into one of four categories: babies, young adults, middle-aged adults, and old adults. Their study used two back-propagation neural networks for classification. The first focuses on geometric features, while the second focuses on wrinkle features. Their study achieved a 90.52% identification rate for the training images and 81.58% for the test images, which they noted is similar to a human's subjective justification for the same set of images. One of the complications noted by the researchers, which likely contributed to their seemingly low rates of success in comparison with other classification studies, was the fact that the age cutoffs for varying levels of "adults" do not typically have hard divisions, but for the sake of the study, this is necessary. For example, the researchers established the cutoff between young and middle adults at 39 years old ($\leq$39 for young, $>$39 for middle). This creates issues when individuals are right at the boundary of two age groups. To prevent similar issues with our experiment, we decided to simply classify the images as "Autistic" and "Non-Autistic" rather than trying to additionally classify the levels of autism.

In a study by Shan (2012), researchers used Local Binary Patterns (LBP) to describe faces. Through the application of support vector machines (SVM), they were able to achieve a 94.81% success rate in determining the gender of the subject. The main breakthrough of this study was its ability to use only real-life images in their classification. Up to this point, many of the proven studies used ideal images, most of which were frontal, occlusion-free, with a clean background, consistent lighting, and limited facial expressions. Similar to this study, our facial images are derived from real-life environments and the dataset was constructed organically.

Classification of Autism

A study conducted by El-Baz et al. (2007) focused on analyzing images of cerebral white matter (CWM) in individuals with autism to determine if classification could be achieved based only on the analysis of brain images. The CWM is first segmented from the proton density MRI and then the CWM gyrification is extracted and quantified. This approach used the cumulative distribution function of the distance map of the CWM gyrification to distinguish between the two classes: autistic and normal. While this study did yield successful results, the images were only taken from deceased individuals, so its success rate in classifying living individuals is still unknown. Our proposed classification system can achieve similar levels of accuracy (94.64%) while using significantly more subjects and only requiring an image of the individual rather than intensive, costly, brain scans, and subsequent detailed analysis.

MobileNet

There are many different convolutional neural networks (CNN) available for image analysis (Hosseini, 2018; Hosseini et al., 2020b). Some of the more well-known models include GoogleNet, VGG-16, SqueezeNet, and AlexNet. Each of these distinct models offers different advantages, but MobileNet has been proven to be similarly effective while greatly reducing computation time and costs. MobileNet has shown that making their models thinner and wider has resulted in similar accuracy while greatly reducing multi-adds and parameters required for analysis (**Tables 1–3**).

TABLE 1 | Depthwise separable vs. full convolution MobileNet (Jeatrakul and Wong, 2009; Howard et al., 2017).

Model	ImageNet accuracy %	Million mult-adds	Million parameters
Conv. MobileNet	71.7	4,866	29.3
MobileNet	70.6	569	4.2

TABLE 2 | Narrow vs. shallow MobileNet and MobileNet width multiplier (Jeatrakul and Wong, 2009; Howard et al., 2017; Hosseini et al., 2020a).

Model	ImageNet accuracy %	Million mult-adds	Million parameters
Shallow MobileNet	65.3	307	2.9
1.0 MobileNet-224	70.6	569	4.2
0.75 MobileNet	68.4	325	2.6
0.5 MobileNet-224	63.7	149	1.3
0.25 MobileNet-224	50.6	41	0.5

TABLE 3 | MobileNet resolution and MobileNet comparison to popular models.

Model	ImageNet accuracy %	Million mult-adds	Million parameters
1.0 MobileNet-224	70.6	569	4.2
1.0 MobileNet-192	69.1	418	4.2
1.0 MobileNet-160	67.2	290	4.2
1.0 MobileNet-128	64.4	186	4.2
GoogleNet	69.8	1,550	6.8
VGG 16	71.5	15,300	138

In comparison with the previously mentioned models, MobileNet has been shown to be just as accurate while significantly reducing the computing power necessary to run the model (**Tables 2, 3**). Using this knowledge, we have decided that MobileNet is a sufficient model to use for our analysis.

METHODS

In previous studies, children with autism have been found to have unusually wide faces and wide-set eyes. The cheeks and the nose are also shorter on their faces (Aldridge et al., 2011). In this study, deep learning has been used to train and learn about autism from facial analyses of children based on features that make them stand out from other children.

Our data set was obtained from Kaggle and consists of 3,014 children's facial images. Of these images, 1,507 images are presumed to have autism, and the remaining 1,507 are presumed to be healthy. **Figure 2** shows a sample of images used for the training step. Images were obtained online, both through Facebook groups and through Google Image searches. Independent research was not conducted to determine if the individual in a picture was truly healthy or autistic. Once all the images were gathered, they were subsequently cropped so

TABLE 4 | Dataset breakdown.

Data set	Composition	Overall data composition %
Train	1,327 autistic 1,327 healthy	88
Validation	80 autistic 80 healthy	5.3
Test	140 autistic 140 healthy	9.3
Total	1,507 autistic 1,507 healthy	100

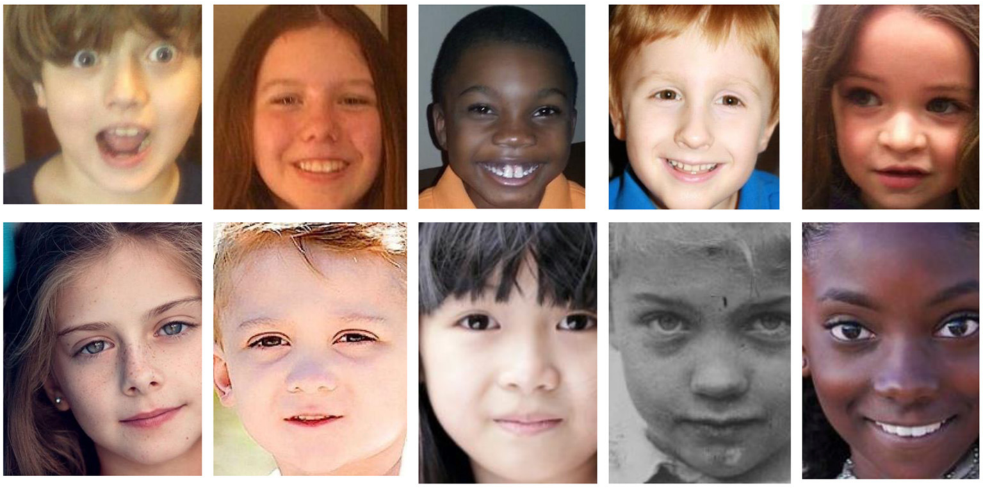

FIGURE 2 | Some images used in the deep learning training step. **(Top)** Children who have autism. **(Bottom)** Children who do not have autism.

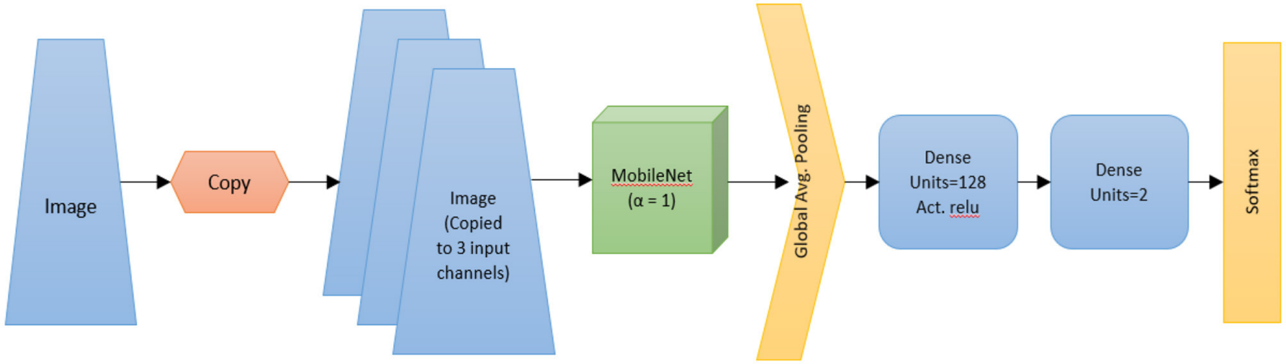

FIGURE 3 | The algorithm architecture of the proposed model, illustrating the use of MobileNet, followed by two dense layers to perform image recognition. MobileNet uses CNN to predict what is the shape of the object present and what is matched with it from the images.

that the faces occupied most of the image. Before training, the images were split into three categories: train, validation, and test (**Table 4**). Images that were placed into each category must be put there manually. Therefore, repeatedly running the algorithm will generally produce the same results, assuming that the neural network ends up with the same weights. It is also worth noting that, currently, the global dataset has multiple repetitions, some of which are shared between the training, test, and validation datasets (Faris et al., 2016; Hosseini et al., 2017). It is, therefore, essential that these duplicates be cleaned out of the datasets before running the algorithm. For this case study, the duplicates have not yet been removed, which is likely improving overall accuracy.

Deep learning is broken down into three subcategories: CNN, pretrained unsupervised networks, and recurrent and recursive networks (Hosseini et al., 2016b). For this data set, we decided to use a CNN model. CNN can intake an image, assign importance to various objects within the image, and then differentiate objects within the image from one another. Additionally, CNNs are advantageous because the preprocessing involved is minimal compared to other methods. In this case, the input is the many images from the dataset to give an output variable: autistic or non-autistic. When looking at CNN, there are various kinds of methods to apply: LeNet, GoogLeNet, AlexNet, VGGNet, ResNet, and so forth. When trying to decide which CNN to use, it is crucial to consider what kind of data is in use, and the size of data being applied. For this instance, MobileNet is used because of the dataset: MobileNet is able to compute outputs much faster, as it can reduce computation and model size.

To perform deep learning on the dataset, MobileNet was utilized followed by two dense layers as shown in **Figure 3**. The first layer is dedicated to the distribution and allows customization of weights to input into the second dense layer. Thus, the second dense layer allows for classification. The architecture of MobileNet can be reviewed in **Table 5**.

For our MobileNet, an alpha of 1 and depth multiplier of 1 were utilized, thus we use the most baseline version of MobileNet. To make binary predictions from MobileNet, two fully connected layers are appended to the end of the model. The first is a dense layer with 128 neurons (L2 regularization = 0.015, ReLu activation) which is then connected to the prediction layer which only has two outputs (softmax activation). A dropout of 0.4 is applied to the first layer to prevent overfitting. The final output is a binary classification of either "autistic" or "non-autistic."

The algorithm was run on an ASUS laptop (Beitou District, Taipei, Taiwan) with an Intel Core i7-6700HQ CPU at 2.60 GHz and 12 GB of RAM. The data was broken into batch sizes of 80. Upon completion of training and initial testing, the user can request additional training epochs.

RESULTS

The training was completed after ∼15 epochs, yielding a test accuracy of 94.64%. **Figure 4** shows how the loss of the training and test set changed with the continual addition of one epoch at a time. **Figure 5** shows how accuracy (Hosseini et al., 2016a) changed for training, validation, and testing data with the continual addition of one epoch.

During the training, the weights that gave the validation set the highest accuracy were always stored. Therefore, if the accuracy decreased during a training set, there would be no ultimate loss of accuracy on the test set (**Figures 1**, **2**). Similarly, if accuracy on the validation set decreased, the learning rate would also decrease during the next training session. Each epoch required ∼10 min to run.

These preliminary results are very promising. Currently, there are many issues with the dataset that was used including duplicate images, improper age ranges, and lack of validation about the

TABLE 5 | Body architecture of the developed MobileNet (Howard et al., 2017) for autism image recognition.

Input size	Filter	Layer	Stride
224 * 224 * 3	3 * 3 * 3 * 32	Convolution	S2
112 * 112 * 32	Depth-Wise 3 * 3 * 32	Depth-Wise Convolution	S1
112 * 112 * 32	1 * 1 * 32 * 64	Convolution	S1
112 * 112 * 64	Depth-Wise 3 * 3 * 64	Depth-Wise Convolution	S2
56 * 56 * 64	1 * 1 * 64 * 128	Convolution	S1
56 * 56 * 128	Depth-Wise 3 * 3 * 128	Depth-Wise Convolution	S1
56 * 56 * 128	1 * 1 * 128 * 128	Convolution	S1
56 * 56 * 128	Depth-wise 3 * 3 * 128	Depth-Wise Convolution	S2
28 * 28 * 128	1 * 1 * 128 * 256	Convolution	S1
28 * 28 * 256	Depth-Wise 3 * 3 * 256	Depth-Wise Convolution	S1
28 * 28 * 256	1 * 1 * 256 * 256	Convolution	S1
28 * 28 * 256	Depth-Wise 3 * 3 * 256	Depth-Wise Convolution	S2
14 * 14 * 256	1 * 1 * 256 * 512	Convolution	S1
14 * 14 * 512	Depth-Wise 3 * 3 * 512	Depth-Wise Convolution	S1
14 * 14 * 512	1 * 1 * 512 * 512	Convolution	S1
14 * 14 * 512	Depth-Wise 3 * 3 * 512	Depth-Wise Convolution	S1
14 * 14 * 512	1 * 1 * 512 * 512	Convolution	S1
14 * 14 * 512	Depth-Wise 3 * 3 * 512	Depth-Wise Convolution	S1
14 * 14 * 512	1 * 1 * 512 * 512	Convolution	S1
14 * 14 * 512	Depth-Wise 3 * 3 * 512	Depth-Wise Convolution	S1
14 * 14 * 512	1 * 1 * 512 * 512	Convolution	S1
14 * 14 * 512	Depth-Wise 3 * 3 * 512	Depth-Wise Convolution	S1
14 * 14 * 512	1 * 1 * 512 * 512	Convolution	S1
14 * 14 * 512	Depth-Wise 3 * 3 * 512	Depth-Wise Convolution	S2
7 * 7 * 512	1 * 1 * 512 * 1,024	Convolution	S1
7 * 7 * 1,024	Depth-Wise 3 * 3 * 1,024	Depth-Wise Convolution	S2
7 * 7 * 1,024	Depth-Wise 1 * 1 * 1,024	Convolution	S1
7 * 7 * 1,024	Pool 7 * 7	Average pooling	S1
1 * 1 * 1,024	1,024 * 1,000	Fully connected	S1
1 * 1 * 1,000	Classifier	Softmax	S1

conditions of the individuals in each photo. Improving the data set could result in better results.

CONCLUSION

While the statistics on how many children are diagnosed with autism are somewhat low, it is extremely important to diagnose as early as possible to provide the correct care for the patient. Additionally, the statistics on diagnosed children may be low because the method to accurately diagnose a child is somewhat ineffective. Thus, our classifier could prove to be very useful in diagnosing more children. Our results show that we have successfully achieved a high accuracy of 94.64%, meaning that it was able to identify a child with or without autism correctly about 95% of the time. To improve accuracy, cleaning the dataset would certainly help. Duplicates may falsely increase our test accuracy if an image is also in the training category.

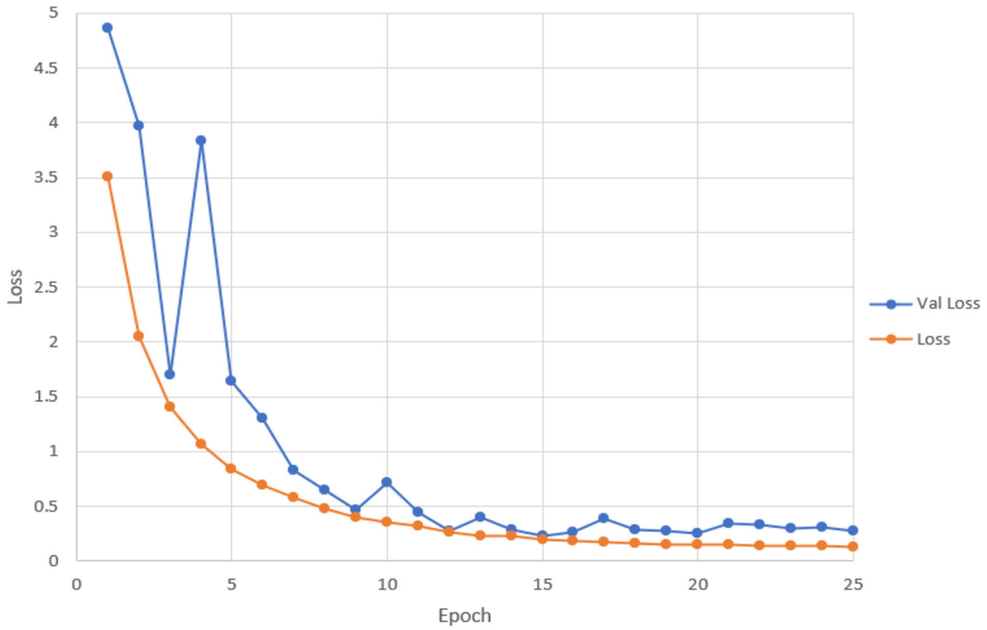

FIGURE 4 | By adding one epoch at a time, this figure shows how the loss changes.

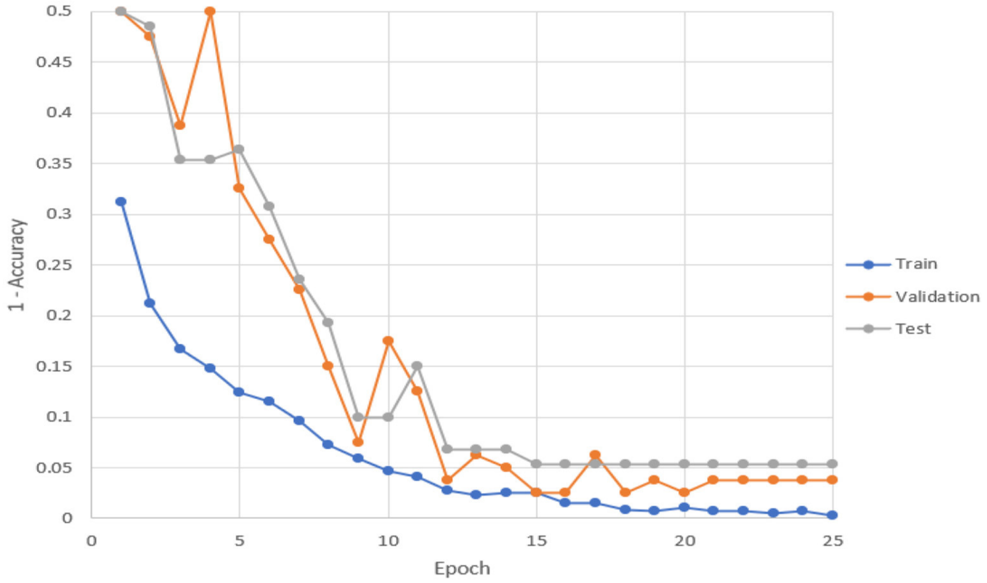

FIGURE 5 | With the addition of one epoch, this figure shows how accuracy changed for training, validation, and testing data.

With more information about the individuals in the pictures, we could also ensure that age distributions are similar between the two populations. Currently, autism is rarely diagnosed in young children, so we would also ensure that no pictures of young children are in our dataset. Similarly, we could ensure that each category is "pure," preventing false positives and false negatives. With these improvements, we would hope to get an accuracy >95%.

The success of this algorithm may also imply that other diseases can be diagnosed using only a picture, saving valuable time and resources in diagnosing other diseases and conditions. Down's Syndrome, for example, is another disease that markedly alters the facial features of those it afflicts. It is possible that, given sufficient and good data, our algorithm could distinguish between individuals with the disease and individuals without it.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: https://www.kaggle.com/cihan063/autism-image-data.

AUTHOR CONTRIBUTIONS

M-PH defined and lead the project and defined the model and structures. MB, AH, and RM worked on the model. HS-Z proofread and organized the study for publication. All authors approved the final version of the manuscript for submission.

REFERENCES

Aldridge, K., George, I. D., Cole, K. K., Austin, J. R., Takahashi, T. N., Duan, Y., et al. (2011). Facial phenotypes in subgroups of prepubertal boys with autism spectrum disorders are correlated with clinical phenotypes. *Mol. Autism* 2, 1–12. doi: 10.1186/2040-2392-2-15

Alexander, A. L., Lee, J. E., Lazar, M., Boudos, R., DuBray, M. B., Oakes, T. R., et al. (2007). Diffusion tensor imaging of the corpus callosum in autism. *Neuroimage* 34, 61–73. doi: 10.1016/j.neuroimage.2006.08.032

Austimspeaks (2019). *What Is Autism?* Austimspeaks. Available online at: https://www.autismspeaks.org/what-autism (accessed March 11, 2020).

CBS News (2017). *Is it Autism? Facial Features That Show Disorder*. CBS News. Available online at: https://www.cbsnews.com/pictures/is-it-autism-facial-features-that-show-disorder/8/ (accessed March 12, 2020).

El-Baz, A., Casanova, M. F., Gimel'farb, G., Mott, M., and Switwala, A. E. (2007). "A new image analysis approach for automatic classification of autistic brains," in *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (Arlington, VA: IEEE), 352–355. doi: 10.1109/ISBI.2007.356861

Eni, M., Dinstein, I., Ilan, M., Menashe, I., Meiri, G., and Zigel, Y. (2020). Estimating autism severity in young children from speech signals using a deep neural network. *IEEE Access* 8, 139489–139500. doi: 10.1109/ACCESS.2020.3012532

Faris, H., Aljarah, I., and Mirjalili, S. (2016). Training feedforward neural networks using multi-verse optimizer for binary classification problems. *Appl. Intell.* 45, 322–332. doi: 10.1007/s10489-016-0767-1

Gerry (2020). *Autistic Children Data Set*. Kaggle.

Horng, W. B., Lee, C. P., and Chen, C. W. (2001). Classification of age groups based on facial features. *Tamkang Instit. Technol.* 4, 183–192. doi: 10.6180/jase.2001.4.3.05

Hosseini, M. P. (2018). *Brain-Computer Interface for Analyzing Epileptic Big Data.* (Doctoral dissertation), New Brunswick, NJ: Rutgers University-School of Graduate Studies.

Hosseini, M. P., Hosseini, A., and Ahi, K. (2020a). A Review on machine learning for EEG Signal processing in bioengineering. *IEEE Rev. Biomed. Eng.* 14, 204–218. doi: 10.1109/RBME.2020.2969915

Hosseini, M. P., Lu, S., Kamaraj, K., Slowikowski, A., and Venkatesh, H. C. (2020b). "Deep learning architectures," in *Deep Learning: Concepts and Architectures* (Cham: Springer), 1–24. doi: 10.1007/978-3-030-31756-0_1

Hosseini, M. P., Nazem-Zadeh, M. R., Pompili, D., Jafari-Khouzani, K., Elisevich, K., and Soltanian-Zadeh, H. (2016a). Comparative performance evaluation of automated segmentation methods of hippocampus from magnetic resonance images of temporal lobe epilepsy patients. *Med. Phys.* 43, 538–553. doi: 10.1118/1.4938411

Hosseini, M. P., Pompili, D., Elisevich, K., and Soltanian-Zadeh, H. (2017). Optimized deep learning for EEG big data and seizure prediction BCI via internet of things. *IEEE Trans. Big Data* 3, 392–404. doi: 10.1109/TBDATA.2017.2769670

Hosseini, M. P., Soltanian-Zadeh, H., Elisevich, K., and Pompili, D. (2016b). "Cloud-based deep learning of big EEG data for epileptic seizure prediction," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (Washington, DC: IEEE), 1151–1155. doi: 10.1109/GlobalSIP.2016.7906022

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv [Preprint]* arXiv:1704.04861.

Iga, R., Izumi, K., Hayashi, H., Fukano, G., and Ohtani, T. (2003). "A gender and age estimation system from face images," in *SICE 2003 Annual Conference (IEEE Cat. No. 03TH8734), Vol. 1* (Fukui: IEEE), 756–761.

Jeatrakul, P., and Wong, K. W. (2009). "Comparing the performance of different neural networks for binary classification problems," in *2009 Eighth International Symposium on Natural Language Processing* (Bangkok: IEEE), 111–115. doi: 10.1109/SNLP.2009.5340935

Liang, S., Sabri, A. Q. M., Alnajjar, F., and Loo, C. K. (2021). Autism spectrum self-stimulatory behaviors classification using explainable temporal coherency deep features and SVM classifier. *IEEE Access* 9, 34264–34275. doi: 10.1109/ACCESS.2021.3061455

Shan, C. (2012). Learning local binary patterns for gender classification on real-world face images. *Patt. Recogn. Lett.* 33, 431–437. doi: 10.1016/j.patrec.2011.05.016

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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

ORIGINAL RESEARCH
published: 25 October 2022
doi: 10.3389/fnins.2022.830143

Development of a Super-Resolution Scheme for Pediatric Magnetic Resonance Brain Imaging Through Convolutional Neural Networks

Juan Manuel Molina-Maza[1], *Adrian Galiana-Bordera*[1], *Mar Jimenez*[2], *Norberto Malpica*[1] *and Angel Torrado-Carvajal*[1]*

[1] *Medical Image Analysis and Biometry Lab, Universidad Rey Juan Carlos, Madrid, Spain,* [2] *Department of Radiology, Hospital Universitario Quirónsalud, Madrid, Spain*

OPEN ACCESS

Edited by:
Arman Sargolzaei,
Tennessee Technological University,
United States

Reviewed by:
Francesco Grussu,
Vall d'Hebron Institute of Oncology
(VHIO), Spain
Gowtham Murugesan,
BAMF Health LLC, United States

***Correspondence:**
Angel Torrado-Carvajal
angel.torrado@urjc.es

Specialty section:
This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: *06 December 2021*
Accepted: *01 June 2022*
Published: *25 October 2022*

Citation:
Molina-Maza JM, Galiana-Bordera A, Jimenez M, Malpica N and Torrado-Carvajal A (2022) Development of a Super-Resolution Scheme for Pediatric Magnetic Resonance Brain Imaging Through Convolutional Neural Networks. Front. Neurosci. 16:830143. doi: 10.3389/fnins.2022.830143

Pediatric medical imaging represents a real challenge for physicians, as children who are patients often move during the examination, and it causes the appearance of different artifacts in the images. Thus, it is not possible to obtain good quality images for this target population limiting the possibility of evaluation and diagnosis in certain pathological conditions. Specifically, magnetic resonance imaging (MRI) is a technique that requires long acquisition times and, therefore, demands the use of sedation or general anesthesia to avoid the movement of the patient, which is really damaging in this specific population. Because ALARA (as low as reasonably achievable) principles should be considered for all imaging studies, one of the most important reasons for establishing novel MRI imaging protocols is to avoid the harmful effects of anesthesia/sedation. In this context, groundbreaking concepts and novel technologies, such as artificial intelligence, can help to find a solution to these challenges while helping in the search for underlying disease mechanisms. The use of new MRI protocols and new image acquisition and/or preprocessing techniques can aid in the development of neuroimaging studies for children evaluation, and their translation to pediatric populations. In this paper, a novel super-resolution method based on a convolutional neural network (CNN) in two and three dimensions to automatically increase the resolution of pediatric brain MRI acquired in a reduced time scheme is proposed. Low resolution images have been generated from an original high resolution dataset and used as the input of the CNN, while several scaling factors have been assessed separately. Apart from a healthy dataset, we also tested our model with pathological pediatric MRI, and it successfully recovers the original image quality in both visual and quantitative ways, even for available examples of dysplasia lesions. We hope then to establish the basis for developing an innovative free-sedation protocol in pediatric anatomical MRI acquisition.

Keywords: deep learning (DL), magnetic resonance imaging (MRI), pediatric imaging, sedation, super-resolution (SR), convolutional neural networks (CNN)

1. INTRODUCTION

Magnetic Resonance Imaging (MRI) is the key modality for obtaining pathophysiological information in a non-invasive manner, while avoiding the use of ionizing radiation that may be harmful to patients, causing them short- or long-term damaging effects. By using MRI, it is possible to obtain high-resolution images describing the anatomy, function, diffusion, etc. of the patient and capture decisive details. Although there are no major issues among most of the adult populations, pediatric MRI acquisition represents, however, an important challenge, as their cooperation becomes more difficult to attain for the clinical trial success. Children need to stand still and immobile during the examination and get to tolerate the claustrophobic environment inside the scanner. However, this is challenging and motion artifacts appear as a direct consequence, frequently worsening the image quality and sharpness, thus, restraining its usability as a diagnosis and treatment tool. In this scenario, and also due to intense noise and the aforementioned restriction of space inside the MRI scanner, sedation is very advisable for examinations of children in most situations, as stopping an MRI scan is expensive and ineffective, hence the failure rate needs to be mitigated (Schulte-Uentrop and Goepfert, 2010). Nevertheless, despite these artifacts can be refrained with the aid of sedation or anesthesia and reach a high image quality, children could suffer potential long-term neurological and cognitive side effects, such as hallucinations or nightmares, and also cardiovascular affectations, like hypertension, bradycardia, alterations in mean arterial pressure, and myocardial depression (Slovis, 2011; Ahmad et al., 2018). Additionally, because as low as reasonably achievable (ALARA) principles are considered for all imaging studies -especially pediatric ones- the main goal is to restrict its dispense as much as possible and limit it only to imperative cases.

In this sense, there exist distraction methods that have revealed a reduction in the percentage of sedated infants compared to control groups (Harned Ii and Strain, 2001; McGee, 2003; Khan et al., 2007), so parents' stress decrease substantially because sedation and its associated recovery time are luckily skipped. One of the most common distraction methods consists of wearing video goggles and earphones in order to watch and listen to films as a distraction during the examination. Another way to calm children and infants during the acquisition is displaying a video on a mobile screen monitor that can be adjusted according to any position of the patient inside the scanner. Moreover, a certified child-life specialist gets children ready to behave correctly in the scanner and attract their attention during the examination with the direct effect of improving image quality. Sleep manipulation might represent an alternative way to handle children's movements (Edwards and Arthurs, 2011). Deprivation of sleep before the examination has been demonstrated to prevent infants from sedation and acts as a substitute for pharmacological sleep induction.

While all these techniques are extremely useful in shortening scan time, they still result insufficient to avoid sedation in some pediatric patients. In this context, ground-breaking concepts and novel technologies can help to find a solution to these challenges while helping in the search for underlying disease mechanisms. Very recently, the feasibility and acceptability of MRI imaging studies without anesthesia have been demonstrated in adolescent populations with moderate or severe neuropathic pain (Verriotis et al., 2020), or autism spectrum disorders (Smith et al., 2019). In this sense, post-MRI-acquisition super-resolution (SR) methods play an essential role. SR methods can provide high resolution (HR) images from low resolution (LR) images and consequently achieve a resultant image quality similar to ideal HR images with a shorter acquisition time.

Traditional mathematical methods can address the SR problem in MRI. For instance, Peled and Yeshurun (2001) and Yan and Lu (2009) use the iterative back projection method in Diffusion Tensor images and anatomical images, respectively. In the case of Gholipour et al. (2010), a stochastic regularization process is applied to brain fetal MRI. Another popular implementations are sparse methods, which subdivide the image and store the result in a dictionary. Lustig et al. (2007) made subimage sampling for sparse coding on the frequency domain, and Zhang et al. (2019) succeeded to recover details from neonatal T1 and T2-weighted MRI, with the help of older children's images. However, all these classical methods for SR come with performance limitations when applied to pediatric MRI. With that in mind, it is preferred to develop a scheme that profits from deep learning (DL) tools to generate the SR image, as they imply an improvement compared with results using classical methods.

We can find convolutional neural network (CNN) implementations in the literature that successfully solve the SR problem using MRI. Pham et al. (2019) and Chaudhari et al. (2018) implemented a 3D CNN with no feature dimension reduction and evaluated the network for many scaling factors. They also demonstrate that one single network gets to deal with more than one scaling factor. In the case of Pham et al. (2019), the same number of filters is applied for each layer and they use residual operations and a single skip connection. Qualitative evaluation is carried out from HR neonatal reconstructions as the ground truth of real LR data is not available. In Zeng et al. (2020), a fully dense connected cascade network is developed and they compared the results with a Unet, which is the base architecture in our work. Chen et al. (2018) also proposed a 3D densely connected network for brain MRI SR with the same number of filters inside the dense block. Pham et al. (2017) designed a 3D CNN with variation in the number of filters depending on the layer, while Du et al. (2020) opt for a residual CNN whose input is individual 2D MRI slices. As well, Qiu et al. (2020) combine a three hidden layer 2D CNN with a sub-pixel convolutional layer using knee MRI and Zeng et al. (2018) carry out a sophisticated approach where they benefit from distinct modality information and the architecture implies two consecutive CNNs.

Apart from CNN development in brain MRI SR, some works based on Generative Adversarial Networks (GANs) have been published in the last few years. Sánchez and Vilaplana (2018) used T1 MRI for two different scaling factors, and Lyu et al. (2018) takes as input T2 individual MRI slices for training two kinds of GANs. Wang et al. (2020) implement a 3D GAN framework with T1 images. Finally, You et al. (2020) and

Tan et al. (2020) make a more elaborated implementation of the GAN framework.

In this work, we exploit recent development in Graphics Processing Units (GPU) and we present a DL approach to address the SR problem for pediatric MRI. Our CNN is based on the form of a Unet autoencoder and we get to successfully recover the resolution from pediatric LR MRI. The CNN architecture includes skip connections, residual operations, and feature maps with distinct dimensions depending on the layer, allowing the network to learn features with many levels of complexity. Using consecutive convolutions, we intend to extract critical features that determine the SR process during the encoding stage. Then, the objective is to reconstruct the output from deep features maps as dimension and domain from input and output MRI of CNN need to be the same in order to compare both at the evaluation stage.

2. MATERIALS AND METHODS

2.1. Dataset

An open dataset containing brain MRI volumes of 155 healthy controls has been used for the main implementation and technical assessment of our DL architecture (Richardson et al., 2018). This dataset is composed of 33 adults (20 female; 24.8 ± 5.3 years, range of 18–39 years) and 122 children (64 female; 6.7 ± 2.3 years, range of 4–12 years). Anatomical T1-weighted and functional MRI volumes have been acquired on a Siemens TIM Trio 3T MRI scanner (Siemens Healthineers, Erlangen, Germany) located at the *Athinoula A. Martinos Center for Biomedical Imaging (Boston, MA, USA)*. Anatomical volumes were obtained with a 32 channel head coil using an MPRAGE sequence (TR = 2.53 s, TE = 1.64 ms, $FlipAngle$ = 7°, 1 mm isotropic voxel) including two different matrix sizes: 176 × 256 × 256 and 176 × 192 × 192. Specific head coils have been used for subjects younger than 5 years.

An additional dataset composed of 12 children (3 female, 9.29 ± 3.76; range of 3–16 years) brain MRI volumes diagnosed with different types of dysplasia and presenting different lesions has been used for a more detailed clinical assessment (**Table 1**). Anatomical T1-weighted volumes have been acquired on two different General Electric (GE) 3T MRI scanners, a Signa HDxt 3.0T and a Signa Premier(GE Healthcare, Waukesha, WI, USA) located at *Quirónsalud Madrid University Hospital (Pozuelo de Alarcón, Madrid)*. Among them, 8 volumes were obtained with an 8 channel head coil using an EFGRE3D sequence (TR = 9.244 ms, TE = 3.428 ms, TI = 750 ms, $FlipAngle$ = 10°, 1.13mm isotropic voxel) and 4 of them were acquired with a 48 channel head coil using an MPRAGE sequence (TR = 2.2 s, TE = 2.8 ms, $FlipAngle$ = 8°, 1 mm isotropic voxel). The use of these images for the present study was approved by the local Institutional Review Board.

2.2. Data Preprocessing

2.2.1. Quality Assurance and Size Standardization

Quality assurance (QA) in the open database revealed that from the initial 155 healthy controls, 32 children volumes (matrix size 176 x 192 x 192) demonstrated evident imaging artifacts that could negatively impact our procedure and, as a consequence, they were excluded from the final dataset. Thus, 123 healthy controls were selected after QA: 33 adults (20 female; 24.8 ± 5.3 years, range of 18–39 years) and 90 children (46 female; 7.1 ± 2.0 years, range of 4–12 years).

The final matrix size of the selected images after QA is 176x256x256. Zero-padding was applied on both sides in the first dimension to ease later patching of the volumes when fed to our CNN architecture while maintaining the brain centered in the final volume. Additionally, the background was homogenized using a binary mask. The images were not denoised, bias corrected, or registered as part of the pre-processing stage.

The same procedure was applied to the clinical pediatric dataset. No images were discarded in this process. However, as our original pipeline was developed for a matrix size of 256 × 256 × 256 in the open access dataset, we resized all MRI volumes from 512 × 512 × 512 to 256 × 256 × 256.

TABLE 1 | Demographic data of the pediatric dataset used for the clinical evaluation of the algorithm.

ID#	Age	Sex	Disease	MRI sequence (Scanner)
001	8	F	Simple cortical dysplasia	EFGRE3D (Signa HDxt)
002	9	F	Taylor dysplasia	EFGRE3D (Signa HDxt)
003	11	M	Bilateral opercular syndrome	EFGRE3D (Signa HDxt)
004	16	M	Aqueductal stenosis; Opercular syndrome; Heterotopia; Cystic brain lesion	EFGRE3D (Signa HDxt)
005	6	F	Opercular syndrome	EFGRE3D (Signa HDxt)
006	8	M	Taylor dysplasia	EFGRE3D (Signa HDxt)
007	5	M	Cingulum dysplasia	EFGRE3D (Signa HDxt)
008	13	M	Opercular syndrome	EFGRE3D (Signa HDxt)
009	3	M	Opercular syndrome; Cerebral fissure malformation	MPRAGE (SIGNA Premier)
010	8	M	Cortical dysplasia	MPRAGE (SIGNA Premier)
011	13	M	Taylor dysplasia	MPRAGE (SIGNA Premier)
012	6	M	Cortical dysplasia	MPRAGE (SIGNA Premier)

2.2.2. Creation of Low Resolution Database

The selected databases are composed of HR images. Therefore, the first step is the artificial production of LR images from HR original ones. First, we simultaneously apply both downsampling and smoothing operations convolving a 3D cubic gaussian kernel of 5 voxels size ($\sigma = 1$) with the MRI volume, in one of each N voxels per dimension. If the original volume dimensions are (d_1, d_2, d_3), dimensions of HR image downsampled will be $(d_1/N, d_2/N, d_3/N)$ with N as the selected scaling factor. This way of deriving the LR counterparts is in consonance with previous works in the state of the art (Rueda et al., 2013; Shi et al., 2015; Pham et al., 2017, 2019; Zeng et al., 2018).

Low resolution downsampled images are then interpolated to HR image size. Although interpolation is an image transformation that could add noise to the downsampled image, it avoids fitting the CNN to an specific scaling factor. In this manner, the designed CNN does not depend on LR image size, and it becomes a robust method to implement on any image size. Hence, LR images are rescaled to original HR image size (d_1, d_2, d_3). New intensity values are created among existing ones to get the necessary matrix size. In the majority of the mentioned studies that have applied CNNs to SR of MRI, bicubic splines is the key method for interpolation, thus used in this work.

Then, we apply min-max normalization to all MRI volumes. The intensity values of every image are transformed to a common $[0, 1]$ range. As every LR volume is created from an HR image, each LR normalization takes the arguments (minimum and maximum intensity values) from the associated HR image.

2.2.3. Volumes Patching

Convolutional neural network architecture determines the way images feed the network. This work explores two network modalities: 2D and 3D. 3D CNNs receive 3D volumes as input, but not full LR volumes in our case. NVIDIA graphic cards cannot store all parameters for the whole volume because of their limited memory, which forces us to extract patches from the images. Patches are cubic volumes of size 32 and they are extracted in train and validation sets using a stride equal to 32, with no overlap, in order to be differentiated from test patching. This patching process is carried out identically for both input and label images of the network. On the other hand, 2D CNNs receive two-dimensional inputs which are complete individual slices of size 256 x 256.

The data augmentation process is carried out to assure a sufficient amount of information and get CNN generalization. It consists of randomly rotating each volume in every dimension by selecting a random angle within the interval $[-30, 30]$ degrees. We select this interval range of angles as 30 degrees is the maximum possible shift of the subject position inside the MRI scanner.

2.3. Hardware and Software

Computational equipment includes a processor *Intel(R) Core(TM) i9-10980XE CPU, 128 GB RAM memory*, *Ubuntu 20.04 operating system*, and *1 TB hard disk storage*. Data preprocessing has been developed using *Python 3*, and CNNs were designed and trained with *Tensorflow 2* library using *GPU NVIDIA GeForce RTX 3080*.

2.4. Super-Resolution Formulation

The DL-based SR process aims to estimate an HR image $\widehat{\mathbf{X}}$ from an LR image $\mathbf{Y} \in \mathbb{R}^n$ that is simulated from an HR ground-truth image $\mathbf{X} \in \mathbb{R}^m$. Thus, the connection between $\mathbf{X}$ and $\mathbf{Y}$ is formulated as follows:

$$\mathrm{Y} = \mathrm{T}(\mathrm{X}) = (\mathrm{D}_{\downarrow}\mathrm{S})\mathrm{X} + \mathrm{N} \tag{1}$$

with $\mathbf{T} \in \mathbb{R}^{m,n}$ $(m > n)$ as the general transformation, in our case, based on applying to the ground truth HR image a combination of downsampling $\mathbf{D}_{\downarrow}$ and smoothing $\mathbf{S}$ processes. $\mathbf{N}$ indicates the noise related to the transformation. For the purpose of finding the inverse mapping estimation, it is convenient to minimize a least-squares cost function:

$$\hat{\mathrm{X}} = \underset{X}{\operatorname{argmin}} \, \|\mathrm{X} - \mathrm{T}^{-1}(\mathrm{Y})\|^2 = \underset{X}{\operatorname{argmin}} \, \|\mathrm{X} - \mathrm{R}(\mathrm{I}^{\uparrow}(\mathrm{Y}))\|^2 \tag{2}$$

being $\mathbf{T}^{-1}$ an ensemble of a recovery matrix $\mathbf{R} \in \mathbb{R}^{m,m}$ and an upsampling operator $\mathbf{I}^{\uparrow}$ regarding the spline interpolation of LR image $\mathbf{Y}$. Since we have multiple subjects available, $\hat{\mathbf{R}}$ can be achieved by minimizing the following additive function:

$$\hat{\mathrm{R}} = \underset{R}{\operatorname{argmin}} \sum_{i=1}^{N} \|\mathrm{X}_i - \mathrm{R}(\mathrm{I}^{\uparrow}(\mathrm{Y}_i))\|^2 \tag{3}$$

where N is the number of subjects available for generating the model. Finally, the optimal $\hat{\mathbf{R}}$ learnt allows to produce an SR estimation $\hat{\mathbf{X}}$ from a new LR image $\mathbf{Y}$:

$$\hat{\mathrm{X}} = \hat{\mathrm{R}}(\mathrm{I}^{\uparrow}(\mathrm{Y})) \tag{4}$$

2.5. CNN Architecture

The neural network chosen to perform SR is the Unet autoencoder, whose basic structure and function are presented in Ronneberger et al. (2015). This structure is taken as the core basis to insert residual operations to improve performance. In the encoding stage, several downsampling transformations are performed by max pooling operations (pool value of 2 voxels per dimension), and the number of filters is double increased. Moreover, the decoding stage contains transposed convolutions (stride value of 2 voxels per dimension) as upsampling operations, where the number of filters is consecutively halved to achieve the initial image size. In the course of the decoding stage, after each transposed convolution, skip connections concatenate feature maps with their homologous encoding phase to preserve information from multiple scales and complexities. Apart from the dimensions of convolution kernels, 2D and 3D CNNs have identical architecture.

Our proposed architecture is divided into 9 stages. The first four stages are focused on encoding, the fifth one is the latent space, whereas the last four ones belong to the decoding stage. The first two stages in encoding and last two stages on decoding have two convolutional blocks each, whereas the rest of the stages contain three convolutional blocks, and we get 23 convolution blocks in all. Each of those blocks consists of a convolutional layer followed by batch normalization and *ReLU* activation function. Also, we have one final convolution at the end of the network and 4 transposed convolutions in the decoding stage, which makes a total of 28 convolutional layers for the whole CNN.

All convolutional layers use (3,3,3) kernel size except the last one. We take into account that the model accuracy may worsen as the network depth increases so the number of convolutional layers is the minimum needed to succeed with generalization. Also, we tested the model with dilated convolutions but they have been discarded as they did not improve the performance. This demonstrates that a more complex and sophisticated implementation of neural networks does not always give better results.

Residual operations intend to avoid the gradient vanishing problem and are located before each max pooling layer and after each transposed convolution. They append the input features maps to the output feature maps of the current stage. These input maps are previously convolved with a kernel size (1,1,1). Finally, the last convolutional layer employs (1,1,1) kernel size and a sigmoid activation function to bound the intensity values, in case some exceeds the interval [0, 1]. **Figure 1** shows a general overview of the CNN architecture.

2.6. Experimentation

Once data preprocessing is implemented on the original volumes and the model CNN architecture is defined, subjects are divided into different subsets. The main purpose of this work is to significantly improve the resolution of MRI among the infant population. As mentioned before, there are images of adults in the healthy database, and we benefit from them to start the training process and provide a first learning step for our model.

Although adult subjects are not sufficient to reach the expected performance, they generate a first model approach as a starting point for training with children's data. By doing so, the variability and size of the database are increased and we gain a better model generalization. In this manner, we apply transfer learning by retraining healthy children from best adults model, which allows to attain faster and closer the optimal solution of the SR proposal.

There are 33 adult volumes available and they are assigned to the training set (80%, 26 volumes) and test set (20%, 7 volumes). The reason why there is no validation group in adults is that these subjects help to get a first approach to the model but not to evaluate the final model performance, as this work is focused on children SR. For 90 total healthy children, test and validation sets correspond to 18 volumes each (20%) and the training set contains 54 volumes (60%). In the case of dysplasic patients, we obtained worse results doing transfer learning from the healthy

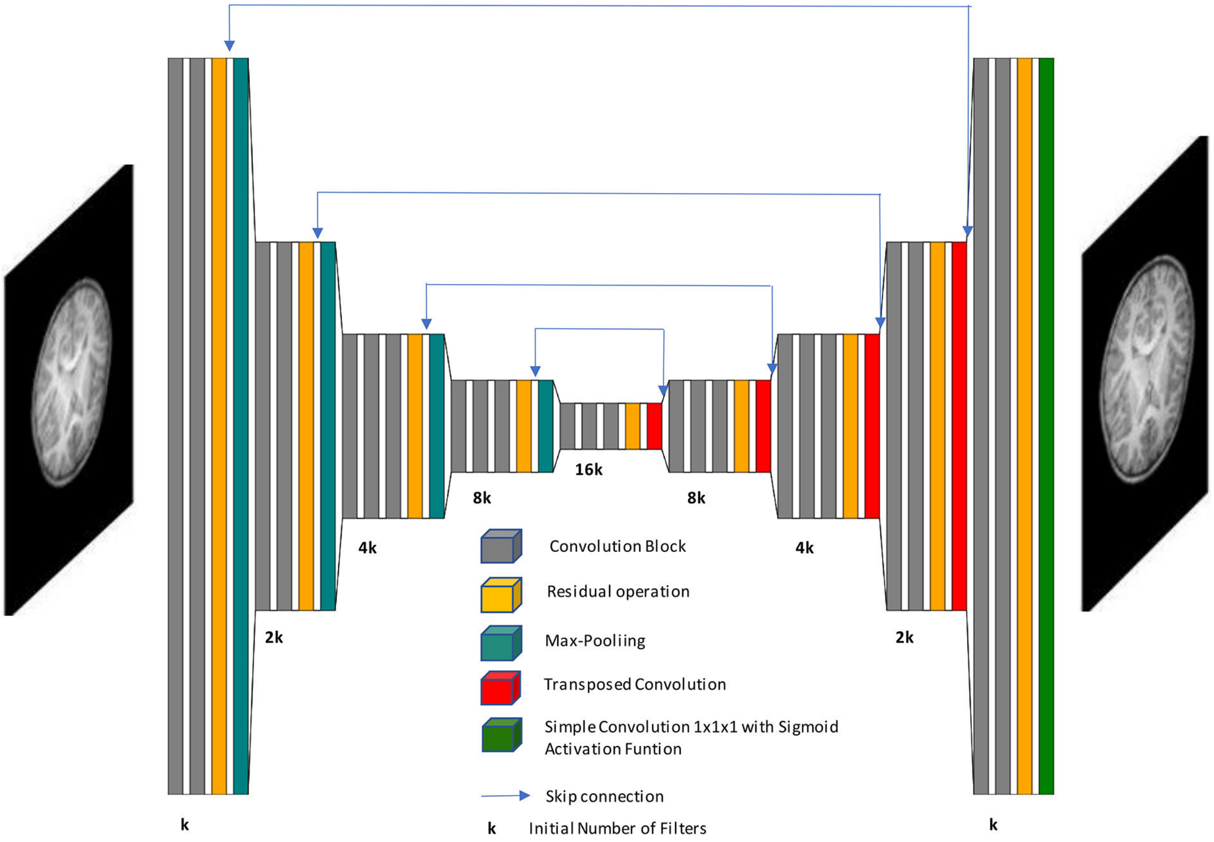

FIGURE 1 | Schematic representation of our network architecture. Our autoencoder implementation processes low-resolution images to generate high-resolution images.

children model than when processing these volumes directly with the aforementioned model. For that reason, this last set of subjects is not divided into subsets and they just form together with the test set (12 volumes).

As optimizer, we chose Adam with an initial learning rate of 0.0001, which is reduced $e^{0.2}$ times if validation loss stops decreasing for more than 4 epochs, being 0.00001 its minimum possible value. Our loss function is the sum of the mean absolute error and a gradient absolute error function. This allows to give value to regions, such as borders or contours, where information could be easily missed otherwise while focusing also on the rest of the image.

Adult and children models have been trained for 50 and 100 epochs, respectively. Those number of epochs are large enough to achieve the convergence of the model. The model selected to perform an evaluation on the test set is the one that gets a smaller value of loss function on the validation group, to avoid both underfitting and overfitting. Furthermore, the loss function is calculated in every iteration only on tissue voxels to prevent background noise from influencing the learning process. Also, batch size has been defined as 12 patches for 3D CNN and 8 patches for 2D CNN, which means that backpropagation updates the parameters after that number of patches. The number of initial filters is 64 for both 3D and 2D CNNs. Even though 2D kernels imply less memory to store, the input patch in 2D CNN contains more voxels than its 3D counterpart and we can not manage with a higher number of filters. Kernel weights are initialized randomly.

Different experiments are performed for every CNN designed. Each CNN (2D and 3D) is trained with three distinct scaling factors (*x2*, *x3*, *x4*) separately. The goal is to analyze at which point there is the best trade-off between the quality of HR estimated image and a significant reduction of MRI acquisition time. In a summary, 6 different training experiments are developed. Spline interpolation is not considered an experiment itself but it is however included in later results for comparing it with CNN-based methods.

2.7. Evaluation and Metrics

Before proceeding with the evaluation stage, the complete estimated volume must be reconstructed by merging the patches or slices generated by the CNN. Image reconstruction for 2D network output is straightforward as it simply consists of attaching one slice above the other. 3D reconstruction requires a more sophisticated procedure. If we merely allocated the produced patches to recover the total volume, borders among them would be evident when visualizing the image. Thus, we extract patches from interpolated LR test images with a stride value lower than the patch side size (in our case this size is 32 voxels and stride value is defined as half size, that is, 16 voxels). Once the CNN outputs the SR estimated patches, they are placed on the final volume and we extract the inner cube of 16 side voxels.

The output of our CNN model is not a single scalar value but a complete image. Thus, we measure the model performance with those mathematical functions able to compare the similarity between the output image and the ground truth image. Three metrics are used in this stage: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Metric (SSIM), and Mean Absolute Error (MAE).

Peak Signal-to-Noise Ratio tries to estimate how noise affects image quality. Its value is commonly expressed in decibel (dB) logarithmic scale:

$$PSNR(\hat{\mathbf{y}}, \mathbf{y}) = 20 \log_{10} \left(\frac{MAX}{\sqrt{MSE(\hat{\mathbf{y}}, \mathbf{y})}} \right) [dB] \tag{5}$$

where *MAX* represents the maximum intensity value of the HR ground truth image, that is 1 in our case, and *MSE* refers to Mean Squared Error.

Mean Absolute Error measures the mean magnitude of errors between images. It calculates the average absolute differences between each pair of equivalent voxels, and it is expressed in intensity values of the image:

$$MAE(\hat{\mathbf{y}}, \mathbf{y}) = \frac{1}{N} \sum_{i=1}^{N} |\hat{\mathbf{y}}(i) - \mathbf{y}(i)| \tag{6}$$

Structural similarity index metric is a metric that quantifies the degradation of the estimated image by extracting 3 key features of the images: luminance, contrast, and structure. It is a non-dimensional metric, much more coherent with human visual perception and is able to imitate it in a quantitative manner. The range of possible values fluctuates in the interval $[-1, +1]$, with 1 meaning exact images. Mathematically, it can be expressed as follows:

$$SSIM(\hat{\mathbf{y}}, \mathbf{y}) = \frac{(2\mu_{\hat{\mathbf{y}}}\mu_{\mathbf{y}} + c_1)(2\sigma_{\hat{\mathbf{y}}\mathbf{y}} + c_2)}{(\mu_{\hat{\mathbf{y}}}^2 + \mu_{\mathbf{y}}^2 + c_1)(\sigma_{\hat{\mathbf{y}}}^2 + \sigma_{\mathbf{y}}^2 + c_1)} \tag{7}$$

where c_1 and c_2 are regularization constants of the metric itself, $\sigma_{\hat{\mathbf{y}}\mathbf{y}}$ is the images joint covariance, $\mu_{\hat{\mathbf{y}}}$ and $\mu_{\mathbf{y}}$ are the averages of images, and $\sigma_{\hat{\mathbf{y}}}^2$ y $\sigma_{\mathbf{y}}^2$ are the images variances.

Although it makes sense to calculate metrics only on the brain tissue using the binary mask and preventing the background of biasing and distorting the evaluation, most of the related literature provides the results on the whole volume and we, therefore, provide metrics for the complete image. While PSNR and SSIM are relative measures and do not depend on the intensity values scale, MAE is however sensitive to it. Thus, we indicate the MAE metric in terms of $[0, 255]$ standard range of possible values.

2.8. Statistical Analysis

Statistical analysis was performed using Python. The quantitative performance of the different methods was assessed by comparing the metrics computed between ground truth images and super-resolved images, using a repeated measures analysis of the variance (ANOVA) to evaluate the effect of the method (interpolation, 2D CNN, 3D CNN), scaling factor (x2, x3, x4), and their interaction, followed by paired-samples Wilcoxon signed rank tests to assess if the performance of each CNN was significantly different from the rest of the methods. Statistical significance was considered when the *p*-value was lower than 0.05.

For the purpose of evaluating the detection of dysplasia lesions in our SR images, an expert radiologist performed two different tests. The first one consists of providing a binary answer whether the lesion can be diagnosed for each of the patients. We only have positive cases in our second database so the sensitivity calculated provides a measure that represents the percentage from all affected SR images where the physician is able to still detect the dysplasias. The second test performed is a 5-point Likert scale. It assigns to each image one of five different levels, where level one means it is impossible to detect the lesions and level five indicates the dysplasia is found perfectly and easily.

3. RESULTS

In this section, we provide both qualitative and quantitative assessments of our methods. These two assessments were also used to infer which architecture (2D vs. 3D) fits better with the SR task. All models were assessed on pediatric images, and all assessments were performed for the different scaling factors to evaluate the ability of our SR scheme to recover data. Finally, we provide some clinical insights into the best model as assessed in the clinical dataset and a further evaluation of the resulting images in a diagnostic scenario.

3.1. Qualitative Assessment

An expert radiologist performed a qualitative assessment by visually inspecting the resulting images to confirm the accuracy of our method; our pipeline performed equally well in all cases considered in the original dataset. **Figures 2**, **3** show a visual comparison between different super-resolution methods (Interpolation, CNN-2D, CNN-3D) and scaling factors (*x2*, *x3*, *x4*) compared to the LR input and the HR ground truth image in a representative case of a 5-year-old subject in the dataset. We can see how the bicubic interpolation method provides an overall improvement over LR images. However, despite solving the loss of spatial resolution, the resultant images are still excessively smoothed, presenting a lack of detail and contrast. The detailed definition of

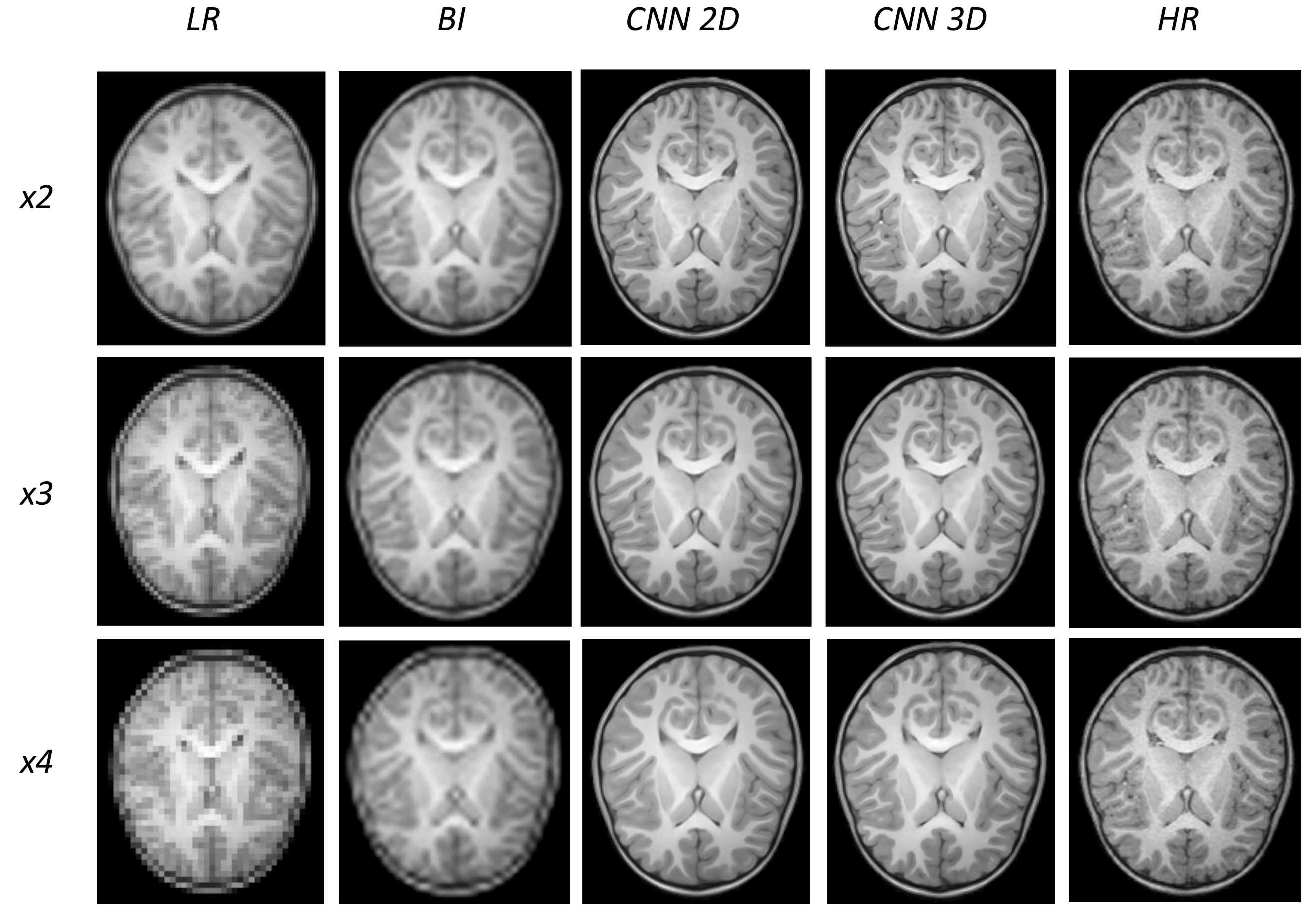

FIGURE 2 | Results of both CNN methods on three scaling factors of super-resolution. A representative slice is shown, together with the original high-resolution image, the low resolution image and the results of bicubic interpolation.

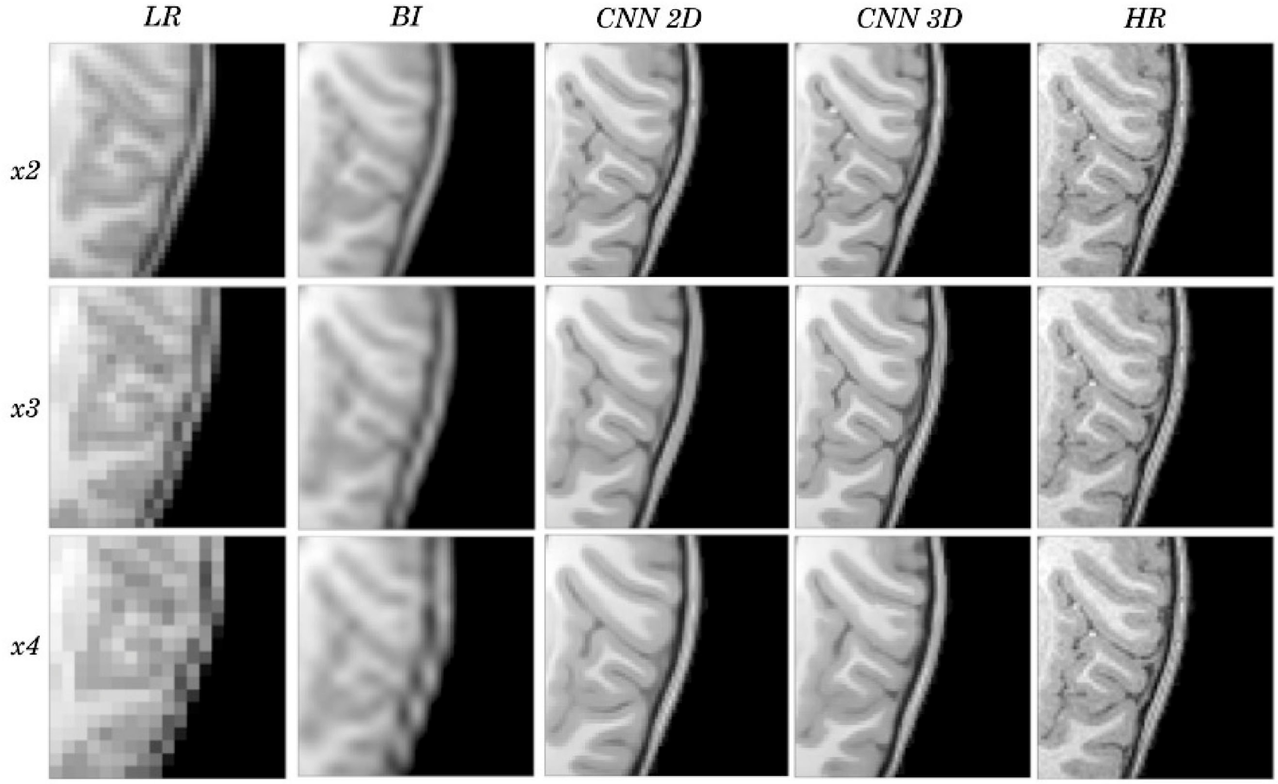

FIGURE 3 | Detail on the images of **Figure 2**.

structures and contours is not well recovered in the interpolated image because it is based on "blind" learning, just on neighboring information from the LR image, thus the newly generated intensity values are basically noise added to the LR image.

Visual inspection of the images generated by both CNNs (2D and 3D) shows that this problem disappears, and the contours are successfully recovered. The granulation that exists in the original HR image is lost, and our resulting images show smoother intensities than the original ones due to some inherent filtering/denoising performed by the CNN, but the shape of the brain was reconstructed generally well despite patient-specific anatomic variations. Actually, the comparison between the patient-specific HR volume and different SR volumes shows that our method accurately estimates the ground truth, delimiting the contours and differentiating tissues for all scaling factors. Visual inspection of the results show the high quality of the SR estimation and the robustness of the method, which is able to capture details of the different tissues in non smooth areas such as the brain circumvolutions. However, as expected, results from higher scaling factors (*x3* and *x4*) present fewer details than those from *x2* scaling factor, losing certain high pixel intensities. Additionally, the 2D CNN slightly loses contour definition compared to the 3D CNN.

TABLE 2 | Results for different metrics and scaling factors among healthy children test set using bicubic interpolation method.

Healthy children	BICUBIC INTERPOLATION		
	x2	x3	x4
PSNR (dB)	33.543 ± 1.404	33.382 ± 1.400	31.841 ± 1.374
MAE	1.569 ± 0.233	1.610 ± 0.238	1.919 ± 0.280
SSIM	0.966 ± 0.004	0.961 ± 0.005	0.945 ± 0.006

TABLE 3 | Results for different metrics and scaling factors among healthy children test set using 2D CNN method.

Healthy children	CNN 2D		
	x2	x3	x4
PSNR (dB)	38.596 ± 1.239	36.852 ± 1.189	35.207 ± 1.182
MAE	0.892 ± 0.123	1.081 ± 0.144	1.293 ± 0.173
SSIM	0.988 ± 0.001	0.983 ± 0.002	0.974 ± 0.003

3.2. Quantitative Assessment

Besides qualitative evaluation, various metrics (PSNR, MAE, and SSIM) have been calculated using the test children's images from the open dataset. **Tables 2–4** show the resultant mean and SD for the different metrics, scaling factors, and SR method, computed as the average among children in the dataset. Similarly, **Figures 4–6** display the statistical distribution of these participants regarding the same aspects. Adult models are included in the violin plots to show that results improve when performing transfer learning from the adult dataset to the pediatric dataset in all cases considered. Statistical tests are performed to compare interpolation and the pediatric CNNs, not including the adult CNNs in these analyses.

TABLE 4 | Results for different metrics and scaling factors among healthy children test set using 3D CNN method.

Healthy children	CNN 3D		
	x2	x3	x4
PSNR (dB)	40.167 ± 1.298	37.290 ± 1.239	35.637 ± 1.273
MAE	0.750 ± 0.107	1.022 ± 0.142	1.229 ± 0.176
SSIM	0.992 ± 0.001	0.985 ± 0.002	0.977 ± 0.003

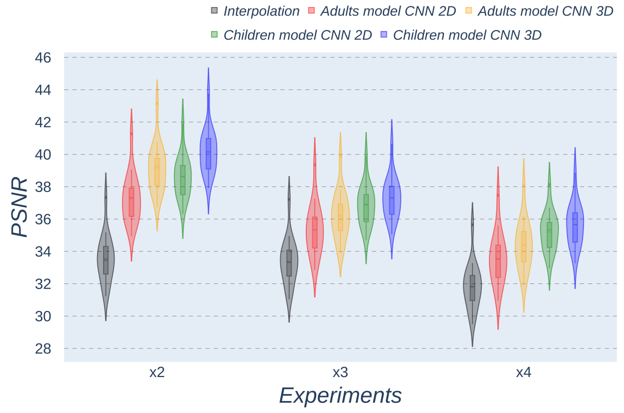

FIGURE 4 | Violin plots for PSNR metric comparing the different methods at the distinct scaling factors. Each violin plot corresponds to one specific experiment, and it is calculated from healthy subjects.

All metrics reveal a considerable improvement of CNNs compared to the interpolation approach. In particular, the 3D CNN reaches better results compared to its 2D counterpart. This could be mainly due to a leakage of information along the third dimension when using 2D convolutions, affecting recovering continuity of anatomical structures within that dimension. Nevertheless, the 2D version of the CNN represents a valuable alternative in cases where complete volumes cannot be acquired and isolated two-dimensional slices are the only data available.

Analysis of variance test for PSNR reveals a statistically significant effect for the *"method"* in all the scaling factors: *x2*, *x3*, and *x4* [$F_{(2,\,51)} = 44.91$; $p = 5.65 \text{ x } 10^{-12}$], and the SR procedures: interpolation, 2D, and 3D [$F_{(2,\,51)} = 42.71$; $p = 7.73 \text{ x } 10^{-3}$]. ANOVA test for MAE also reveals a statistically significant effect for the *"method"* in all the scaling factors: *x2*, *x3*, and *x4* [$F_{(2,\,51)} = 53.01$; $p = 3.52 \text{ x } 10^{-13}$] and the SR procedures: interpolation, 2D, and 3D [$F_{(2,\,51)} = 9.82$; $p = 2.47 \text{ x } 10^{-4}$]. In the SSIM case, ANOVA test reveals a statistically significant effect for the *"method"* in all the scaling factors: *x2*, *x3*, and *x4* [$F_{(2,\,51)} = 307.37$; $p = 3.54 \text{ x } 10^{-29}$] and the SR procedures: interpolation, 2D, and 3D [$F_{(2,\,51)} = 87.64$; $p = 3.16 \text{ x } 10^{-17}$]. For all three metrics, the decomposition of such statistical results using paired comparisons by means of the Wilcoxon test reveals that all pediatric methods are statistically different from each other at all scaling factors and procedures ($p's < 0.01$).

The same trends can be observed in the quantitative results over the clinical dataset, where CNNs reveal an improvement compared to interpolation methods, with the 3D CNN demonstrating the best performance in all metrics (**Tables 5–7**, **Figures 7–9**).

Violin plots reveal that only the interpolation method and the adult models present outliers, confirming that CNNs are

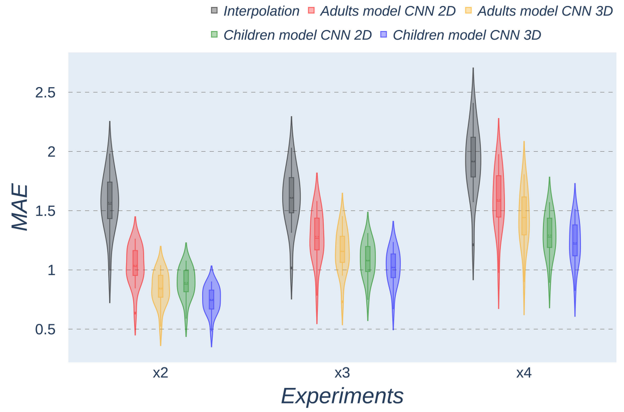

FIGURE 5 | Violin plots for MAE metric comparing the different methods at the distinct scaling factors. Each violin plot corresponds to one specific experiment, and it is calculated from healthy subjects.

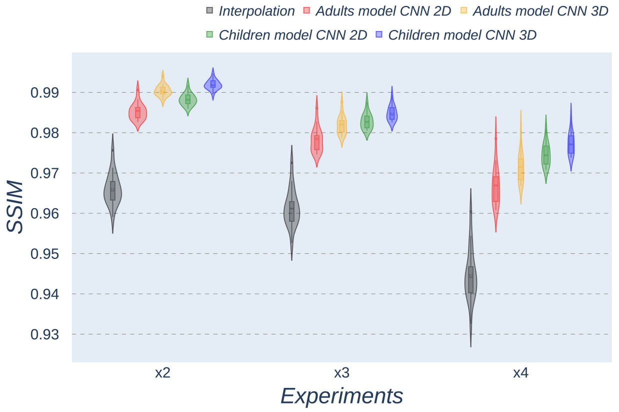

FIGURE 6 | Violin plots for SSIM metric comparing the different methods at the distinct scaling factors. Each violin plot corresponds to one specific experiment, and it is calculated from healthy subjects.

TABLE 5 | Results for different metrics and scaling factors among dysplasic children test set using bicubic interpolation method.

Dysplasic children	Bicubic interpolation		
	x2	x3	x4
PSNR (dB)	32.944 ± 2.097	32.553 ± 2.188	31.211 ± 2.232
MAE	1.484 ± 0.750	1.570 ± 0.798	1.870 ± 0.966
SSIM	0.967 ± 0.002	0.962 ± 0.002	0.946 ± 0.003

TABLE 6 | Results for different metrics and scaling factors among dysplasic children test set using 2D CNN method.

Dysplasic children	CNN 2D		
	x2	x3	x4
PSNR (dB)	35.406 ± 1.851	33.655 ± 2.088	32.028 ± 2.271
MAE	1.089 ± 0.510	1.366 ± 0.686	1.690 ± 0.903
SSIM	0.983 ± 0.006	0.9730 ± 0.011	0.9586 ± 0.020

TABLE 7 | Results for different metrics and scaling factors among dysplasic children test set using 3D CNN method.

Dysplasic children	CNN 3D		
	x2	x3	x4
PSNR (dB)	36.164 ± 2.688	33.642 ± 2.381	32.103 ± 2.409
MAE	0.928 ± 0.457	1.332 ± 0.700	1.663 ± 0.912
SSIM	0.987 ± 0.005	0.974 ± 0.010	0.959 ± 0.020

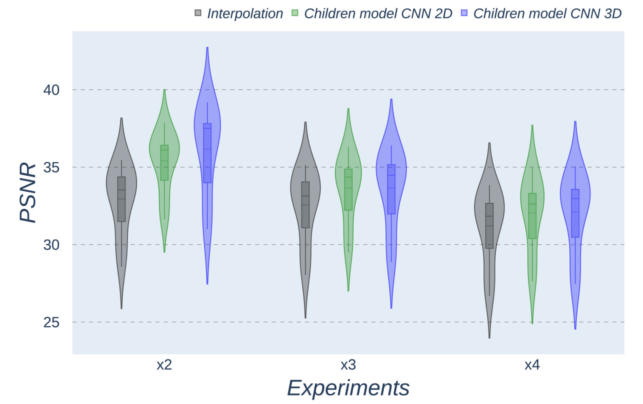

FIGURE 7 | Violin plots for PSNR metric comparing the different methods at the distinct scaling factors. Each violin plot corresponds to one specific experiment, and it is calculated from dysplasic patients.

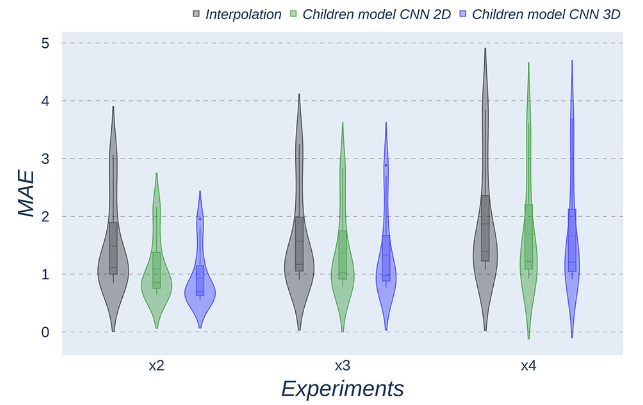

FIGURE 8 | Violin plots for MAE metric comparing the different methods at the distinct scaling factors. Each violin plot corresponds to one specific experiment, and it is calculated from dysplasic patients.

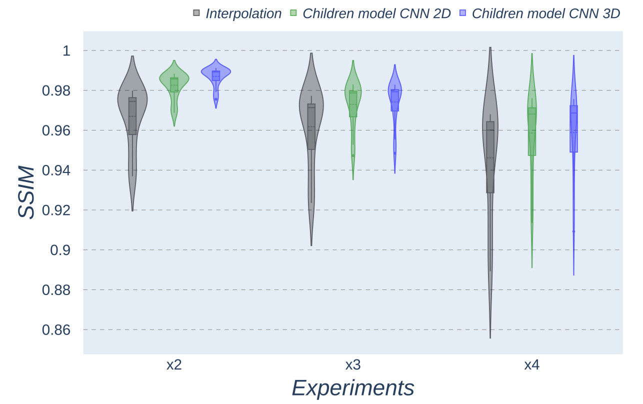

FIGURE 9 | Violin plots for SSIM metric comparing the different methods at the distinct scaling factors. Each violin plot corresponds to one specific experiment, and it is calculated from dysplasic patients.

more accurate and robust methods, introducing less variability among the results for several subjects, especially when trained with specific data. PSNR shows similar trends and dispersion for different scaling factors; MAE values present a more biased widespread distribution in the values above the median than in the values below it; SSIM values are much more concentrated around the median than the rest of the metrics. Moreover, as expected, the *x2* scaling factor is the one that provides the best performance among all the scaling factors, with the 3D CNN demonstrating being the one that shows the best results.

Additionally, it is obvious that analyzing each method's results worsen as LR image size decreases. In other words, a greater scaling factor involves the loss of tissue/intensity information but implies a significant reduction on MRI acquisition time. Thus, the final acceptable scaling factor would depend on how acceptable these results with distinct scaling factors are in the final clinical practice, as shown and discussed in the following subsection.

3.3. Clinical Assessment

Our expert radiologist performed a clinical assessment by visually inspecting the resulting images to provide with her overall qualitative opinion of the reconstructed images regarding diagnosis (5-point Likert scale) and their ability to use them for actual diagnosis in the clinical dataset. A first inspection of the reconstructed images demonstrated that the 3D CNN method visually outperformed the 2D method; thus, subsequent analyses were performed on the resulting 3D images. It is also worth mentioning that images in the clinical dataset were acquired in two different MRI scanners implementing two different anatomical T1-weighted sequences. Thus, we have separated

TABLE 8 | Results for two different qualitative metrics provided by the expert radiologist.

	Sensitivity			5-Point likert scale		
	x2	**x3**	**x4**	**x2**	**x3**	**x4**
EFGRE3D	0.80 ± 0.42	0.70 ± 0.38	0.10 ± 0.32	3.60 ± 0.97	2.50 ± 1.08	1.60 ± 1.07
MPRAGE	1.00 ± 0.00	1.00 ± 0.00	0.75 ± 0.50	5.00 ± 0.00	4.30 ± 0.50	3.30 ± 0.50
BOTH	0.83 ± 0.39	0.75 ± 0.45	0.25 ± 0.45	3.80 ± 1.00	2.80 ± 1.30	1.80 ± 1.10

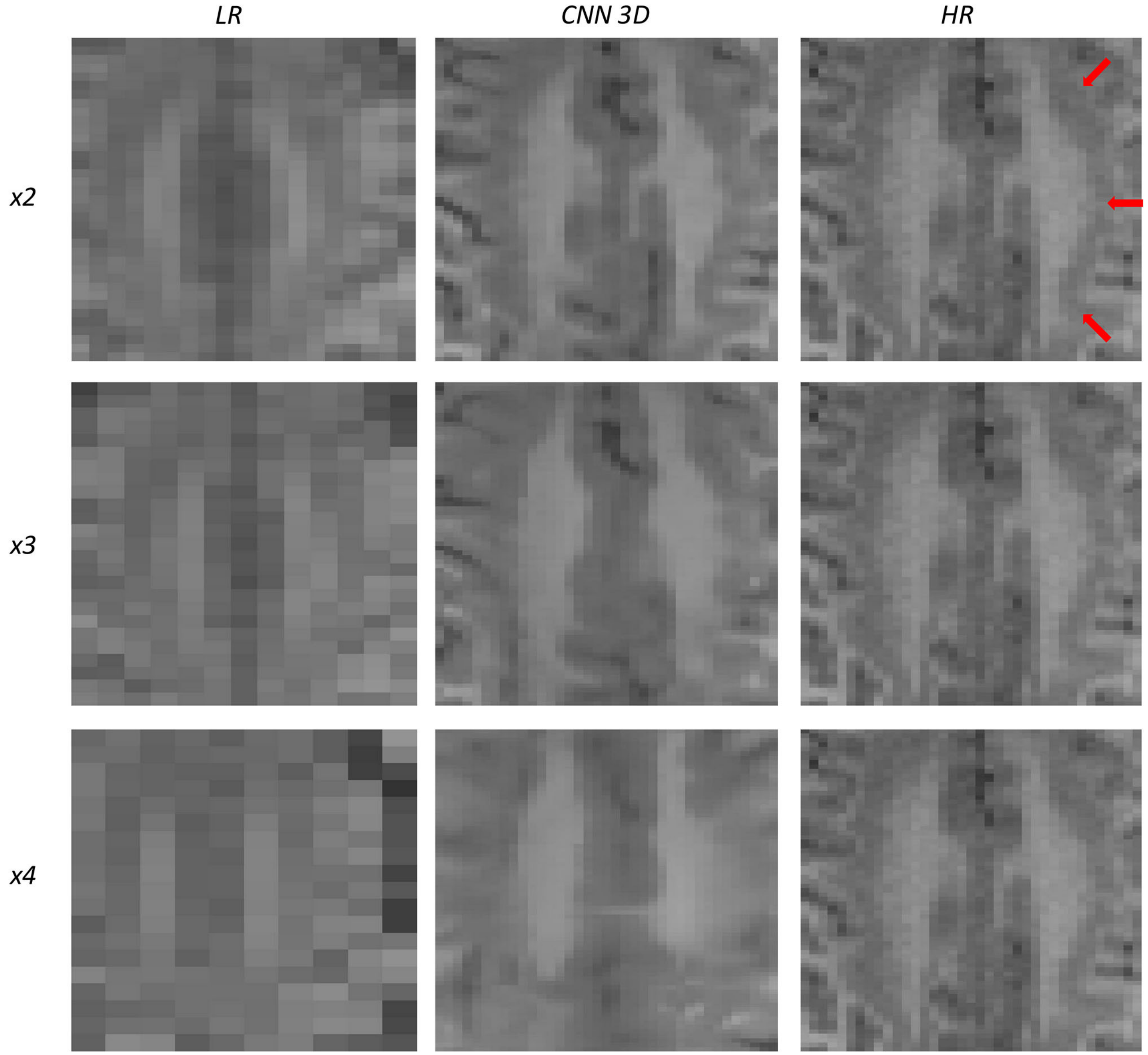

FIGURE 10 | Results of both CNN methods on three scaling factors of super-resolution on images of a patient with dysplasia. A specific region of a slice is shown, with the three red arrows pointing to a lesion.

the corresponding analyses to assess the actual potential of the method in a clinical scenario.

Visual inspection of the images generated by our 3D CNN shows that, depending on the nature of the disease, identification of dysplasias could be straightforward or challenging depending on the baseline scaling factor. **Table 8** shows the decomposition of the results for the 5-point Likert scale and the sensitivity by scaling factor and sequence. As expected, both metrics decrease their values as the scaling factor increases. The 5-point Likert scale reaches a 3.8 ± 1.0 score and a sensitivity of 0.83 ± 0.39 for the *x2* scaling factor when assessing the whole clinical dataset. Additionally, in all these cases, as expected, the CNN provides better results in those images acquired with the same type of sequence that was used for training (MPRAGE), with the 5-point Likert scale and the sensitivity increasing up to 5.0 ± 0.0 and 1.0 ± 0.0, respectively, for the same scaling factor, when analyzing the MPRAGE images only.

Figures 10, **11** show representative examples of patients acquired using the EFGRE sequence on a GE Signa HDxt 3.0T MRI scanner, while **Figures 12**, **13** show representative examples of patients acquired using the MPRAGE sequence on a GE Signa Premier MRI scanner. While these images show a representative

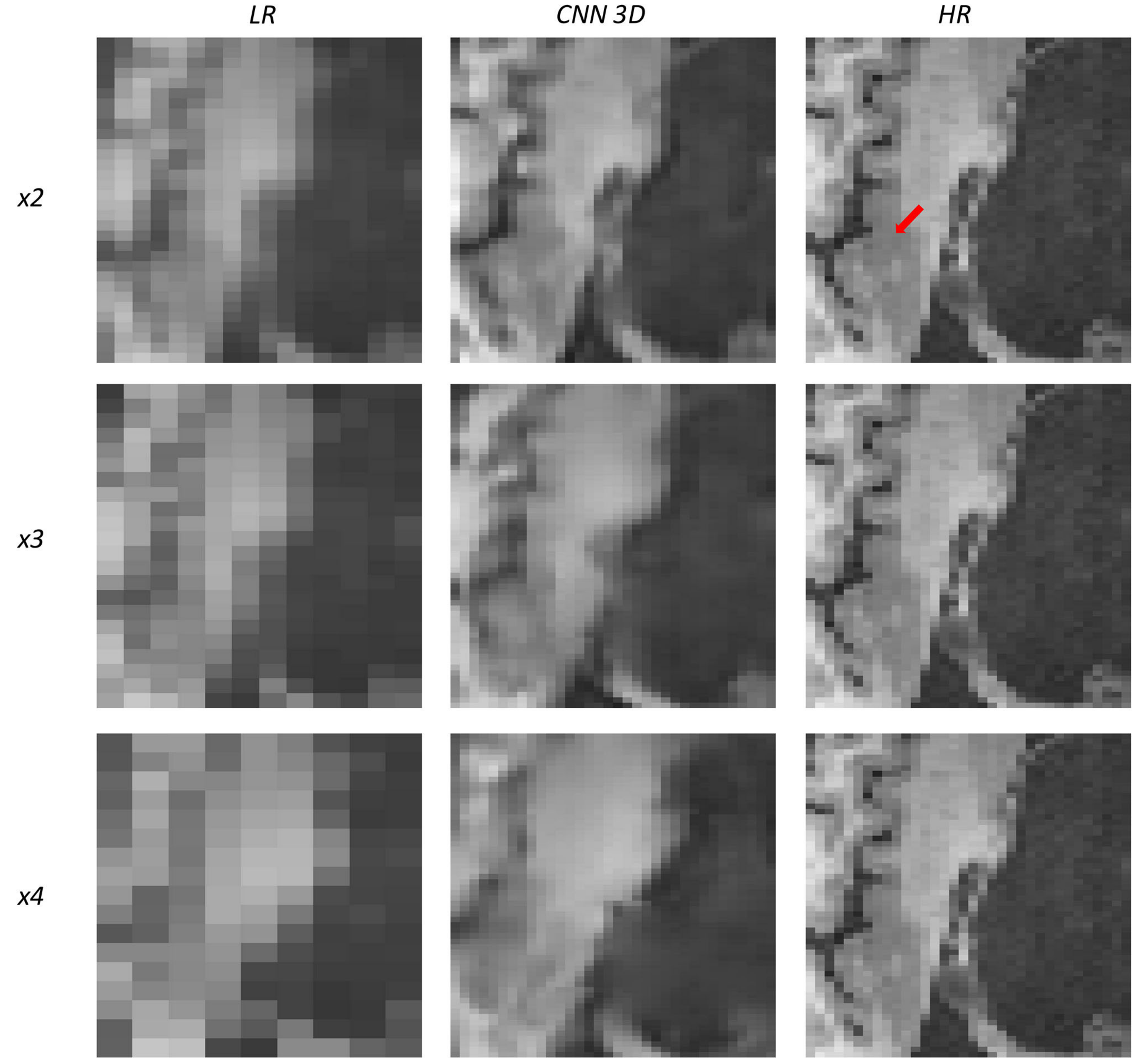

FIGURE 11 | Results of both CNN methods on three scaling factors of super-resolution on images of a patient with dysplasia. A specific region of a slice is shown, with the red arrow pointing to a lesion.

slice, our expert radiologist used the whole volume for the clinical assessment.

Figure 10 shows a 13-years-old male diagnosed with band heterotopia (double cortex syndrome), a form of diffuse gray matter heterotopia due to neuronal migration disorders. In this specific case, the CNN is not able to recover all the details for the *x4* scaling factor, producing a very blurred image; however, as we move forward to the *x3* and the *x2* scaling factors, we can observe further improvements, allowing for the assessment of the bands along the brain.

Figure 11 shows a 16-years-old male diagnosed with aqueductal stenosis, opercular syndrome, heterotopia, and a cystic brain lesion. This patient is specifically difficult to assess due to the presence of diverse diseases. Again, our method fails to recover the details for the *x4* scaling factor and improves moving forward to the *x3* and the *x2* scaling factors. The ventricle enlargement is easy to assess, but the boundaries are better displayed in the image reconstructed from the *x2* scaling factor. The opercular dysplasia can be identified in the *x3* and the *x2* scaling factors.

Figure 12 shows a 9-years-old male diagnosed with opercular syndrome and cerebral fissure malformation. As we move forward to MPRAGE images, we can see the abrupt improvement in the reconstructed images for all scaling factors. Despite the lack in detail for circumvolutions in the dysplasia for the *x4* scaling factor, the thickening of the gray matter can still be assessed.

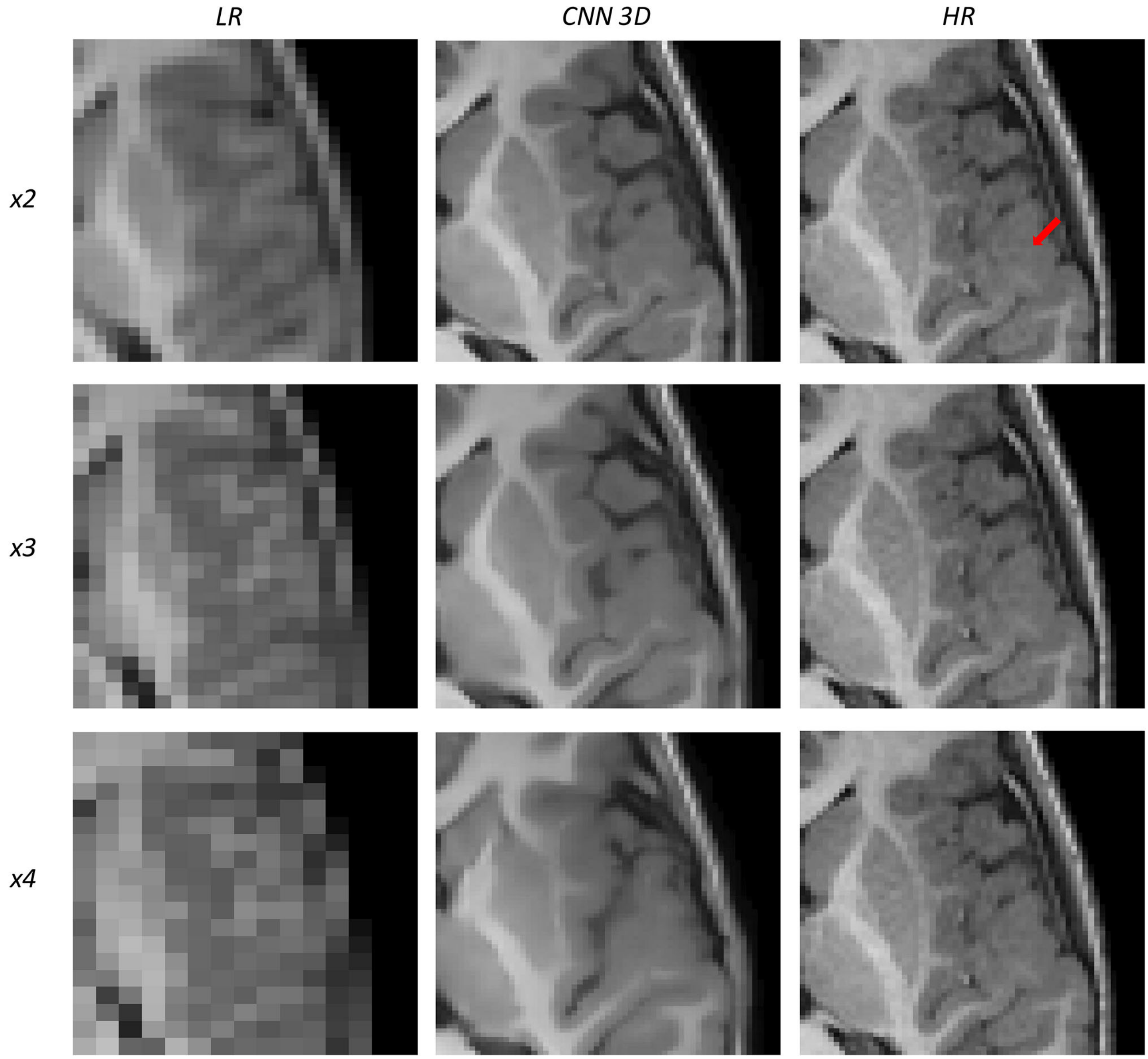

FIGURE 12 | Results of both CNN methods on three scaling factors of super-resolution on images of a patient with dysplasia. A specific region of a slice is shown, with the red arrow pointing to a lesion.

Reconstructions based on the *x3* and *x2* scaling factors allow for delineating even better dysplasia contours and edges.

Figure 13 shows a 12-years-old male diagnosed with a small cortical dysplasia. This patient is especially challenging due to the small size of the dysplasia. However, the lesion can be identified in images reconstructed from all scaling factors. The reconstruction based on the *x2* scaling factor provides an excellent contrast resolution, allowing even to identify the shape of the lesion and presenting less noise than the original HR image.

All these results show that, at least, the *x2* scaling factor is able to recover most of the details in the images, providing with good enough reconstructions to perform disease identification.

4. DISCUSSION

Brain related pathologies often need MRI as a decisive imaging acquisition technique to find the correct diagnosis. Nevertheless, motion artifacts usually appear in pediatric MRI and a method is required to improve the resolution from LR images correctly acquired in a reduced examination time to exclude the use of either sedatives or anesthetics. This project has been based on the design and implementation of a CNN that learns how to automatically increase the resolution of an LR MRI without the need to know its HR counterpart.

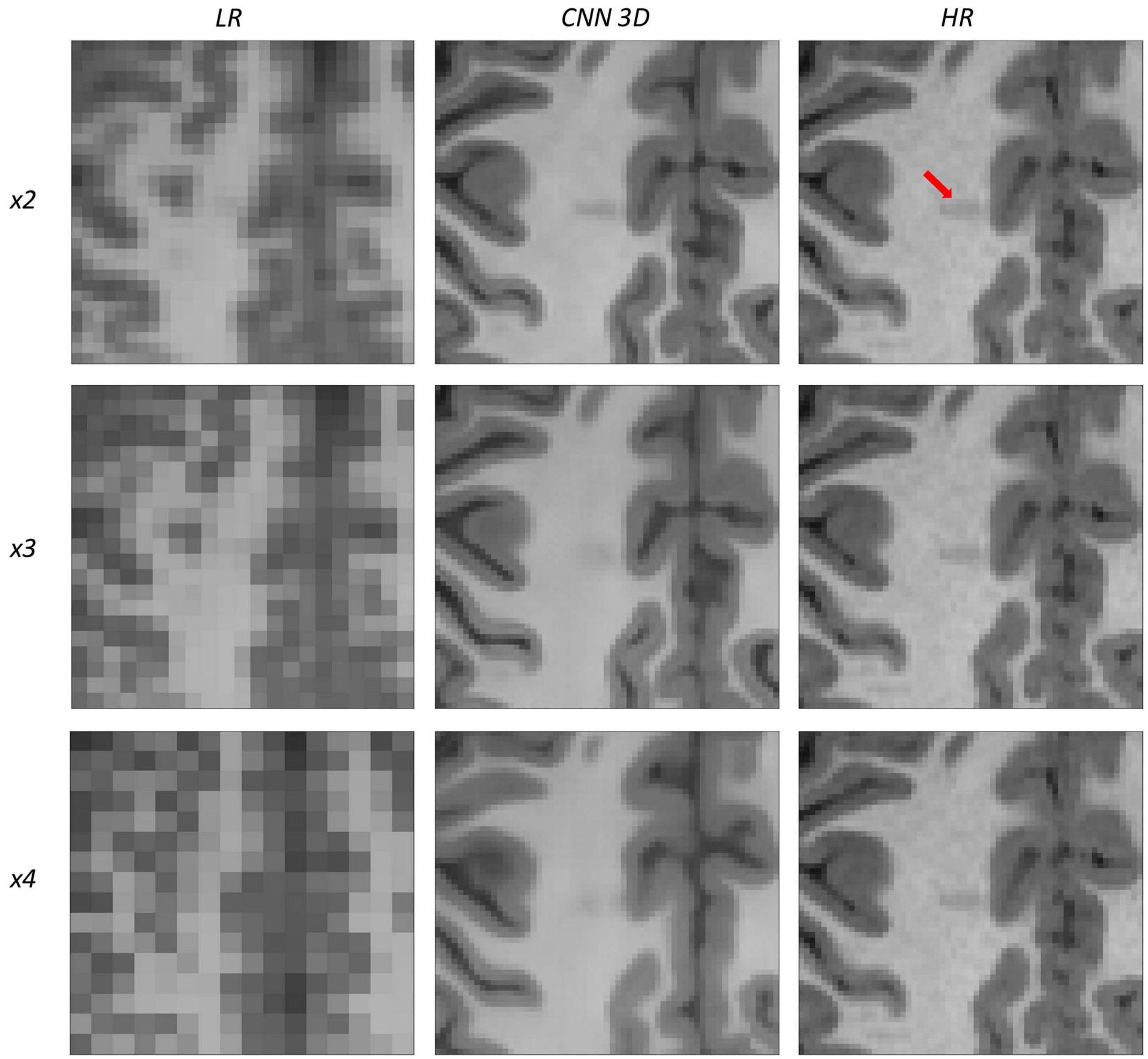

FIGURE 13 | Results of both CNN methods on three scaling factors of super-resolution on images of a patient with dysplasia. A specific region of a slice is shown, with the red arrow pointing to a lesion.

In this work, we have proved how the use of CNNs can satisfactorily solve the initial SR proposal. The database chosen consists of a sufficient number of subjects for the CNN to be able to generalize. In addition, the preprocessing carried out on the original MRI images has managed to adapt them correctly as input to the CNN and the drawback of the artificial generation of the LR database has not adversely affected the performance of the CNN. We obtain visual-pleasing results which agree with the metrics employed to assess the correlation and difference between each pair of images. No matter which scaling factor is utilized, different brain tissues can always be differentiated and most of their details are maintained on the SR image.

We have also carried out an MRI simulation for the original matrix size with 1 mm isotropic voxel and its related scaling factor images (2, 3, and 4 mm). The decrease in acquisition time was around 50% for the first scaling factor and 60% for the second one. This reduction could feed physicians with good quality MRI that aid the clinical routine. Depending on the MRI scanner, the time employed to complete a study ranges from 30 min to 2 h and brain MRI takes an average of 45 min or even longer (Slovis, 2011). It then becomes possible to decrease the examination time by about 20 min.

Our results are compared with the state of art methods that use different datasets than ours. These differences include the MRI sequence, the size and age range of the database, and the scanner specifications. We, therefore, need to be careful when comparisons are made with reviewed literature. Furthermore, the LR images were not directly acquired from the scanner and we made the assumption of resemblance between our artificial LR images and the hypothetical LR acquired ones, in order to be able to discuss the different results.

In future work, it would be desirable to extrapolate the results obtained to different datasets and other modalities of MRI, such as T2-weighted, Diffusion Tensor imaging, or functional MRI. Transfer learning from this work model could be applied to those other modalities and build a stronger method that does not depend on a specific MRI modality. If possible, LR images should be acquired directly from MRI equipment and physicians should check in a clinical study that SR images preserve the same diagnostic value than original acquired images.

5. CONCLUSION

This paper proposes a new CNN pipeline to perform superresolution on pediatric MRI and allows a reduction of MRI examination time. We assessed several CNN architectures on different scaling factors. Our work represents a substantial innovation in the pediatric MRI field, proposing an initial starting point to eliminate the need for a sedation protocol among the infant population.

DATA AVAILABILITY STATEMENT

The healthy participants dataset analyzed in this study is publicly available in the Open Neuro Repository (https://openneuro.org/datasets/ds000228/versions/1.1.0). The clinical image dataset is not publicly available due to IRB restrictions. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Dysplasic children MRI data were retrospectively collected and anonymized from existing clinical datasets after review and approval of the Research Committee from Hospital Universitario Quirónsalud (Madrid, Spain). Patient consent was not required for the study on human participants in accordance with local legislation and institutional requirements.

AUTHOR CONTRIBUTIONS

JM-M and AG-B developed and implemented the methods. JM-M was responsible for data curation, conduction of the experiments, and writing the first draft. MJ was responsible for clinical data curation, conduction of the clinical assessments, and validation of the results concerning the pathological dataset. JM-M, AG-B, MJ, NM, and AT-C analyzed and interpreted the final data and results. MJ, NM, and AT-C critically revised the manuscript. AT-C was responsible for conceptualization, methodology, supervision, resources, and overall project administration. All authors listed have made a substantial, direct and intellectual contribution to the work and design of the study, contributed to manuscript revision, proofreading, and approved the submitted version for publication.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their valuable comments and suggestions that helped to improve the quality of the study.

REFERENCES

Ahmad, R., Hu, H. H., Krishnamurthy, R., and Krishnamurthy, R. (2018). Reducing sedation for pediatric body MRI using accelerated and abbreviated imaging protocols. *Pediatr. Radiol.* 48, 37–49. doi: 10.1007/s00247-017-3987-6

Chaudhari, A. S., Fang, Z., Kogan, F., Wood, J., Stevens, K. J., Gibbons, E. K., et al. (2018). Super-resolution musculoskeletal MRI using deep learning. *Magn. Reson. Med.* 80, 2139–2154. doi: 10.1002/mrm.27178

Chen, Y., Xie, Y., Zhou, Z., Shi, F., Christodoulou, A. G., and Li, D. (2018). "Brain MRI super resolution using 3D deep densely connected neural networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (Washington, DC: IEEE), 739–742.

Du, J., He, Z., Wang, L., Gholipour, A., Zhou, Z., Chen, D., et al. (2020). Super-resolution reconstruction of single anisotropic 3D MR images using residual convolutional neural network. *Neurocomputing* 392, 209–220. doi: 10.1016/j.neucom.2018.10.102

Edwards, A. D., and Arthurs, O. J. (2011). Paediatric MRI under sedation: is it necessary? what is the evidence for the alternatives? *Pediatric radiology* 41, 1353–1364. doi: 10.1007/s00247-011-2147-7

Gholipour, A., Estroff, J. A., and Warfield, S. K. (2010). Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain MRI. *IEEE Trans. Med. Imaging* 29, 1739–1758. doi: 10.1109/TMI.2010.2051680

Harned Ii, R. K., and Strain, J. D. (2001). MRI-compatible audio/visual system: impact on pediatric sedation. *Pediatr. Radiol.* 31, 247–250. doi: 10.1007/s002470100426

Khan, J. J., Donnelly, L. F., Koch, B. L., Curtwright, L. A., et al. (2007). A program to decrease the need for pediatric sedation for CT and MRI. *Appl. Radiol.* 36, 30. doi: 10.37549/AR1505

Lustig, M., Donoho, D., and Pauly, J. M. (2007). Sparse mri: The application of compressed sensing for rapid mr imaging. *Magn. Reson. Med.* 58, 1182–1195. doi: 10.1002/mrm.21391

Lyu, Q., You, C., Shan, H., and Wang, G. (2018). Super-resolution MRI through deep learning. *arXiv [Preprint] arXiv*:1810.06776. doi: 10.1117/12.2530592

McGee, K. (2003). The role of a child life specialist in a pediatric radiology department. *Pediatr. Radiol.* 33, 467–474. doi: 10.1007/s00247-003-0900-2

Peled, S., and Yeshurun, Y. (2001). Superresolution in MRI: application to human white matter fiber tract visualization by diffusion tensor imaging. *Magn. Reson. Med.* 45, 29–35. doi: 10.1002/1522-2594(200101)45:1andlt;29::AID-MRM1005andgt;3.0.CO;2-Z

Pham, C.-H., Ducournau, A., Fablet, R., and Rousseau, F. (2017). "Brain MRI super-resolution using deep 3D convolutional networks," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (Melbourne, VIC: IEEE), 197–200.

Pham, C.-H., Tor-Díez, C., Meunier, H., Bednarek, N., Fablet, R., Passat, N., et al. (2019). Multiscale brain MRI super-resolution using deep 3D convolutional networks. *Comput. Med. Imaging Graph.* 77, 101647. doi: 10.1016/j.compmedimag.2019.101647

Qiu, D., Zhang, S., Liu, Y., Zhu, J., and Zheng, L. (2020). Super-resolution reconstruction of knee magnetic resonance imaging based on deep learning. *Comput. Methods Progr. Biomed.* 187, 105059. doi: 10.1016/j.cmpb.2019.105059

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., and Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nat. Commun.* 9, 1–12. doi: 10.1038/s41467-018-03399-2

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* (Munich: Springer), 234–241.

Rueda, A., Malpica, N., and Romero, E. (2013). Single-image super-resolution of brain MR images using overcomplete dictionaries. *Med. Image Anal.* 17, 113–132. doi: 10.1016/j.media.2012.09.003

Sánchez, I., and Vilaplana, V. (2018). Brain MRI super-resolution using 3D generative adversarial networks. *arXiv [Preprint] arXiv*:1812.11440. doi: 10.48550/arXiv.1812.11440

Schulte-Uentrop, L., and Goepfert, M. S. (2010). Anaesthesia or sedation for MRI in children. *Curr. Opin. Anesthesiol.* 23, 513–517. doi: 10.1097/ACO.0b013e32833bb524

Shi, F., Cheng, J., Wang, L., Yap, P.-T., and Shen, D. (2015). LRTV: MR image super-resolution with low-rank and total variation regularizations. *IEEE Trans. Med. Imaging* 34, 2459–2466. doi: 10.1109/TMI.2015.2437894

Slovis, T. L. (2011). Sedation and anesthesia issues in pediatric imaging. *Pediatr. Radiol.* 41, 514–516. doi: 10.1007/s00247-011-2115-2

Smith, C., Bhanot, A., Norman, E., Mullett, J., Bilbo, S., McDougle, C., et al. (2019). A protocol for sedation free MRI and PET imaging in adults with autism spectrum disorder. *J. Autism. Dev. Disord.* 49, 3036–3044. doi: 10.1007/s10803-019-04010-3

Tan, C., Zhu, J., and Lio, P. (2020). "Arbitrary scale super-resolution for brain MRI images," in *IFIP International Conference on Artificial Intelligence Applications and Innovations* (Neos Marmaras: Springer), 165–176.

Verriotis, M., Moayedi, M., Sorger, C., Peters, J., Seunarine, K., Clark, C. A., et al. (2020). The feasibility and acceptability of research magnetic resonance imaging in adolescents with moderate-severe neuropathic pain. *Pain Rep.* 5, e807. doi: 10.1097/PR9.0000000000000807

Wang, J., Chen, Y., Wu, Y., Shi, J., and Gee, J. (2020). "Enhanced generative adversarial network for 3d brain mri super-resolution," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (Snowmass, CO: IEEE), 3627–3636.

Yan, Z., and Lu, Y. (2009). "Super resolution of MRI using improved IBP," in *2009 International Conference on Computational Intelligence and Security, Vol. 1* (Beijing: IEEE), 643–647.

You, S., Liu, Y., Lei, B., and Wang, S. (2020). Fine perceptive gans for brain MR image super-resolution in wavelet domain. *arXiv [Preprint] arXiv*:2011.04145. doi: 10.48550/arXiv.2011.04145

Zeng, K., Zheng, H., Cai, C., Yang, Y., Zhang, K., and Chen, Z. (2018). Simultaneous single-and multi-contrast super-resolution for brain MRI images based on a convolutional neural network. *Comput. Biol. Med.* 99, 133–141. doi: 10.1016/j.compbiomed.2018.06.010

Zeng, W., Peng, J., Wang, S., and Liu, Q. (2020). A comparative study of cnn-based super-resolution methods in mri reconstruction and its beyond. *Signal Process.* 81, 115701. doi: 10.1016/j.image.2019.115701

Zhang, Y., Yap, P.-T., Chen, G., Lin, W., Wang, L., and Shen, D. (2019). Super-resolution reconstruction of neonatal brain magnetic resonance images via residual structured sparse representation. *Med. Image Anal.* 55, 76–87. doi: 10.1016/j.media.2019.04.010

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: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership