frontiers | Research Topics

Extracting insights from digital public health data using artificial intelligence, volume II

Edited by
Steven Fernandes, Hong Lin, João Manuel R. S. Tavares, Shyamala Guruvare and Yu-Dong Zhang

Coordinated by
Prianna Menezes

Published in
Frontiers in Public Health
Frontiers in Oncology

ISSN 1664-8714
ISBN 978-2-8325-3803-6
DOI 10.3389/978-2-8325-3803-6

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

Extracting insights from digital public health data using artificial intelligence, volume II

Topic editors

Steven Fernandes — Creighton University, United States
Hong Lin — University of Houston–Downtown, United States
João Manuel R. S. Tavares — University of Porto, Portugal
Shyamala Guruvare — Kasturba Medical College, Manipal Academy of Higher Education, India
Yu-Dong Zhang — University of Leicester, United Kingdom

Topic coordinator

Prianna Menezes — Royal College of Surgeons in Ireland, Bahrain

Citation

Fernandes, S., Lin, H., Tavares, J. M. R. S., Guruvare, S., Zhang, Y.-D., Menezes, P., eds. (2024). *Extracting insights from digital public health data using artificial intelligence, volume II*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-3803-6

Table of contents

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 17 October 2022
DOI 10.3389/fpubh.2022.1038502

OPEN ACCESS

EDITED BY
Shyamala Guruvare,
Manipal Academy of Higher Education, India

REVIEWED BY
Naveena A. N. Kumar,
Manipal Academy of Higher Education, India
Vidya Kudva,
Nitte Mahalinga Adyanthaya Memorial Institute of Technology, India

*CORRESPONDENCE
Maoxian Li
596636357@qq.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 07 September 2022
ACCEPTED 29 September 2022
PUBLISHED 17 October 2022

CITATION
Li M, Wang J, Li J, Zhang Y, Zhao X, Lin Y, Deng C, Li F and Peng Q (2022) Develop and validate nomogram to predict cancer-specific survival for patients with testicular yolk sac tumors.
Front. Public Health 10:1038502.
doi: 10.3389/fpubh.2022.1038502

Develop and validate nomogram to predict cancer-specific survival for patients with testicular yolk sac tumors

Maoxian Li[1,2*†], Jinkui Wang[2†], Jinfeng Li[1], Yongbo Zhang[1], Xing Zhao[1], Yang Lin[1], Changkai Deng[1], Fulin Li[1] and Qiang Peng[1]

[1]Department of Pediatric Surgery, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China, [2]Department of Urology, Ministry of Education Key Laboratory of Child Development and Disorders, National Clinical Research Center for Child Health and Disorders, China International Science and Technology Cooperation Base of Child Development and Critical Disorders, Children's Hospital of Chongqing Medical University, Chongqing, China

Purpose: Testicular yolk sac tumor (TYST) is a rare malignant germ cell tumor that mainly occurs in young men. Due to the low incidence of yolk sac tumors, there is a lack of prospective cohort studies with large samples. We aimed to develop a nomogram to predict cancer-specific survival (CSS) in patients with TYST.

Materials and methods: Patient information was downloaded from the Surveillance, Epidemiology and End Results (SEER) database. We enrolled all patients with TYST from 2000 to 2018, and all patients were randomly divided into a training set and a validation set. Univariate and multivariate Cox proportional hazards regression models were used to identify independent risk factors for patients. We constructed a nomogram based on the multivariate Cox regression model to predict 1-, 3-, and 5-year CSS in patients with TYST. We used a series of validation methods to test the accuracy and reliability of the model, including the concordance index (C-index), calibration curve and the area under the receiver operating characteristic curve (AUC).

Results: 619 patients with TYST were enrolled in the study. Univariate and multivariate Cox regression analysis showed that age, T stage, M stage and chemotherapy were independent risk factors for CSS. A nomogram was constructed to predict the patient's CSS. The C-index of the training set and the validation set were 0.901 (95%CI: 0.859–0.847) and 0.855 (95%CI: 0.865–0.845), respectively, indicating that the model had excellent discrimination. The AUC showed the same results. The calibration curve also indicated that the model had good accuracy.

Conclusions: In this study, we constructed the nomogram for the first time to predict the CSS of patients with TYST, which has good accuracy and reliability and can help doctors and patients make clinical decisions.

KEYWORDS

testicular, yolk sac tumors, cancer-specific survival, SEER, nomogram

Introduction

Testicular yolk sac tumor (TYST) is a rare malignant germ cell tumor with an incidence of <1% (1). The disease is often accompanied by elevated serum alpha-fetoprotein (AFP) (1–4). TYST accounts for 11–18% of germ cell tumors and 10–44% of non-seminoma germ cell tumors (4). Due to its high degree of malignancy, insidious onset and rapid progression, TYST seriously threatens the survival of patients (2, 3). Surgical resection of the primary tumor is a key factor in treatment. Targeted therapy with chemotherapy, retroperitoneal lymph node dissection and distant metastasis have made widely disseminated testicular germ cell tumors treatable (5). Due to the low incidence of TYST, there is a lack of studies on TYST with large samples. The related factors affecting the prognosis of TYST are still unclear.

Previous studies have found that the maximum diameter of the tumor, low echo and low blood flow signal under ultrasound can predict the degree of malignancy of testicular tumors and thus predict the prognosis of patients (6). In addition, a study on germ cell tumors (including testicular and ovarian yolk sac tumors) found that patients with slow AFP decline and long half-life after surgery tended to have poor long-term prognoses (7). Unfortunately, the predictive value of these factors remained unclear. Physicians often make empirical judgments on the prognosis of TYST patients based on the patient's age, clinical stage, surgical condition and chemotherapy, but this method cannot intuitively and quantitatively predict patients' survival.

The National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) database provides high confidence data on many cancer patients since 1973. The nomogram prediction model has been widely used to predict the prognosis of a variety of solid tumors and is considered to be one of the most accurate ways to predict tumors (8), including UISS (9) and SSIGN (10). However, for the above model, there are no relevant reports of these clinical variables for TYST cases. This study retrospectively analyzed the data of TYST patients from the SEER database to explore the clinicopathological features affecting their prognosis. According to these features, a nomogram was constructed to predict the cancer-specific survival (CSS) of TYST patients.

Patients and methods

Data source and data extraction

Patient data were downloaded from the SEER database to identify all patients with TYST. The SEER database is the US national cancer database that includes patients from 18 cancer registries, covering about 30% of the US population. The clinicopathological information of the patients was publicly available. Since the patients' personal information could not be identified, ethical approval and informed consent were not required for our study.

We collected demographic information (age, marital status, race), tumor information (laterality, tumor stage), treatment information (surgery, radiotherapy, chemotherapy), and follow-up information (survival time, survival status, cause of death). Inclusion criteria: (1) pathological diagnosis of TYST; (2) Complete follow-up information. Exclusion criteria: (1) Unknown surgical method; (2) The cause of death is unknown; (3) Survival time <1 month. The inclusion and exclusion process of patients is shown in Figure 1.

Patients were classified as white, black, and other (American Indian/AK Native, Asian/Pacific Islander). The patient's marital status was classified as married, single, and divorced. The laterality of the tumor is classified as left, right, and bilateral. Surgery was classified as non-surgery and excision of testicle. According to whether the patient received radiotherapy, it was classified as yes or no. According to whether the patient received chemotherapy, it was classified as yes or no. Causes of death include death from cancer or death from other reasons.

Nomogram development and validation

All patients were randomized into a training set (70%) and a validation set (30%). We first used univariate Cox regression models in the training set to analyze the factors affecting the patients' CSS. Subsequently, the factors identified by univariate Cox regression analysis were included in the multivariate Cox regression model to screen for independent risk factors affecting patients' CSS. We developed a nomogram based on multivariate Cox regression model to predict CSS in patients with TYST. Then, the concordance index (C-index), the calibration curve, and the area under the receiver operating characteristic curve (AUC) were used to validate the accuracy and discrimination of the model in the training set and the validation set.

Clinical utility

We use decision curve analysis (DCA) to assess the clinical potential practical value of the model. DCA is a new algorithm based on calculating net returns under different thresholds. In addition, we calculated the optimal cut-off value based on the receiver operating characteristic curve (ROC) based on the risk scores of all patients. Then, according to the cut-off value, we developed a risk stratification system that divides patients into high-risk and low-risk groups. We used the log-rank test and

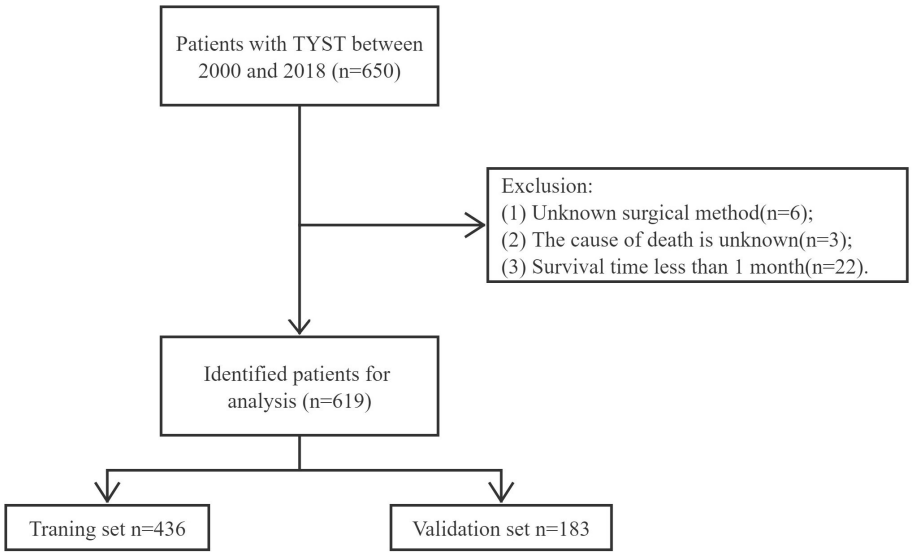


FIGURE 1
Flowchart for inclusion and exclusion of all patients with TYST.

the Kaplan-Meier (K-M) curve to test the survival differences of patients in different risk groups.

Statistical analysis

SPSS 26.0 was used for all statistical analyses. R software 4.1.0 was used to develop and validate the nomogram. Categorical variable data were described by frequency (%), and differences between groups were compared by the chi-square test. Data of continuous variables were described by means and standard deviations, and differences between groups were analyzed by student's *t*-test or non-parametric U test. The Cox regression model analyzed the survival factors of patients. The log-rank test was used to compare differences in survival among different groups of patients. A p value of <0.05 was considered statistically significant.

Result

Clinical features

In total, we enrolled 619 patients with TYST. Patients were randomly assigned to the training set ($n = 436$) and the validation set ($n = 183$). The mean age of the patients was 24.2 $\pm$ 17.0 months. There were 530 (85.6%) white and 131 (21.2%) married patients. There were 300 (48.5%) tumors located in the left testis, 327 (52.8%) tumors confined to the primary site, 275 (44.4%) tumors were stage I, 36 (5.82%) tumors were stage II, and 158 (25.5%) tumors were stage III. The 563 (91.0%) patients underwent orchiectomy, 15 (2.4%) patients received radiotherapy, and 270 (43.6%) patients received chemotherapy. The survival time of all patients was 83.6 $\pm$ 67.1 months. The information of patients in the training and validation sets is shown in Table 1, and there is no significant difference between the two groups.

Univariate and multivariate cox regression analysis

Univariate and multivariate Cox regression analyses were used to explore the prognostic factors for CSS. Specifically, we first used univariate Cox regression analysis to identify factors associated with CSS. The results showed that age, laterality, T stage, M stage, surgery, radiotherapy, and chemotherapy were the factors affecting CSS. These seven factors were then included in a multivariate Cox regression analysis to identify independent risk factors. Finally, we found that age, T stage, M stage, laterality, and chemotherapy were independent predictors of CSS in patients. The results of univariate and multivariate Cox regression analysis are shown in Table 2.

Construction of 1-, 3-, and 5-year CSS nomogram

Based on the multivariate Cox regression model, we constructed a nomogram for predicting CSS in TYST patients. Specifically, the model includes age, T stage, M stage, laterality,

TABLE 1 Clinical characteristics of patients with TYST.

	ALL $N = 619$	Training cohort $N = 436$	Validation cohort $N = 183$	p
Age	24.2 (17.0)	23.9 (17.6)	24.9 (15.8)	0.504
Race				0.875
White	530 (85.6%)	375 (86.0%)	155 (84.7%)	
Black	33 (5.33%)	22 (5.05%)	11 (6.01%)	
Other	56 (9.05%)	39 (8.94%)	17 (9.29%)	
Marital				0.495
Married	131 (21.2%)	93 (21.3%)	38 (20.8%)	
Singled	439 (70.9%)	305 (70.0%)	134 (73.2%)	
Divorced	49 (7.92%)	38 (8.72%)	11 (6.01%)	
Laterality				0.282
Left	300 (48.5%)	219 (50.2%)	81 (44.3%)	
Right	282 (45.6%)	194 (44.5%)	88 (48.1%)	
Unknown	37 (5.98%)	23 (5.28%)	14 (7.65%)	
AJCC				0.148
I	275 (44.4%)	202 (46.3%)	73 (39.9%)	
II	36 (5.82%)	23 (5.28%)	13 (7.10%)	
III	158 (25.5%)	115 (26.4%)	43 (23.5%)	
Unknown	150 (24.2%)	96 (22.0%)	54 (29.5%)	
T				0.522
T1	277 (44.7%)	200 (45.9%)	77 (42.1%)	
T2	102 (16.5%)	76 (17.4%)	26 (14.2%)	
T3	37 (5.98%)	26 (5.96%)	11 (6.01%)	
T4	16 (2.58%)	11 (2.52%)	5 (2.73%)	
TX	187 (30.2%)	123 (28.2%)	64 (35.0%)	
M				0.122
M0	359 (58.0%)	259 (59.4%)	100 (54.6%)	
M1	139 (22.5%)	101 (23.2%)	38 (20.8%)	
MX	121 (19.5%)	89 (20.5%)	32 (17.3%)	
Surgery				1.000
No	56 (9.05%)	39 (8.94%)	17 (9.29%)	
Yes	563 (91.0%)	397 (91.1%)	166 (90.7%)	
Radiotherapy				0.777
No	604 (97.6%)	426 (97.7%)	178 (97.3%)	
Yes	15 (2.42%)	10 (2.29%)	5 (2.73%)	
Chemotherapy				0.766
No	349 (56.4%)	248 (56.9%)	101 (55.2%)	
Yes	270 (43.6%)	188 (43.1%)	82 (44.8%)	
Regional nodes examined				0.128
No	551 (89.0%)	394 (90.4%)	157 (85.8%)	
Yes	68 (11.0%)	42 (9.63%)	26 (14.2%)	
Regional nodes positive				0.273
No	595 (96.1%)	422 (96.8%)	173 (94.5%)	
Yes	24 (3.88%)	14 (3.21%)	10 (5.46%)	
CSS				0.216
Dead	68 (11.0%)	43 (9.86%)	25 (13.7%)	
Alive	551 (89.0%)	393 (90.1%)	158 (86.3%)	

(Continued)

TABLE 1 (Continued)

	ALL $N = 619$	Training cohort $N = 436$	Validation cohort $N = 183$	p
Survival months	83.6 (67.1)	84.7 (66.1)	80.9 (69.4)	0.531
Status				0.621
Dead	93 (15.0%)	63 (14.4%)	30 (16.4%)	
Alive	526 (85.0%)	373 (85.6%)	153 (83.6%)	

AJCC, American Joint Committee on Cancer; T, Tumor; M, Metastasis.

and chemotherapy. As shown in Figure 2, patients' risk of death increased with age. The higher the T stage and M stage, the higher the risk of death. Patients with left-sided tumors have a lower survival rate than those with right-sided tumors.

Validation of nomogram

According to the established nomogram, we first used the C-index to validate the model's accuracy. The range of C-index is between 0.5 and 1, and a high C-index indicates a high discrimination of the model. The C-index of the training and validation sets were 0.886 (95%CI: 0.831–0.941) and 0.876 (95%CI: 0.829–0.923), respectively, indicating the high discrimination of the model. In addition, the calibration curve was used to validate the model's accuracy. As shown in Figure 3, in the training and validation sets, the predicted and observed values were highly consistent, indicating the high accuracy of the nomogram. In addition, AUC was used to validate the discrimination of the model (Figure 4). Compared with the C-index, the time-dependent AUC can more clearly show the trend of model discrimination over time. In the training set, the AUC of nomogram were 93.0, 89.3 and 87.6 at 1-, 3- and 5-year, respectively. In the validation set, the AUC of nomogram were 90.1, 90.3, and 82.3 at 1-, 3- and 5-year, respectively. It shows that the model has high discrimination.

Clinical application of the nomogram

DCA is used to validate the practical value of the model. As shown in Figure 5, the DCA of the nomogram was superior to the traditional TNM staging in both the training and validation sets. In addition, according to the nomogram, we construct a new risk stratification system. We calculated the risk value for each patient and then divided the patients into a high-risk group (total score $\geq$94.5) and a low-risk group (total score $<$94.5) using the optimal cut-off value of the ROC. In the high-risk group, patients' 1-, 3- and 5-year survival rates were 87.5, 79.6, and 78.1%, respectively. The patients' 1-, 3-, and 5-year survival rates were 100.0, 98.8, and 97.9% in the low-risk group. In both the training and validation sets, patients in the high-risk group had significantly lower survival than those in the low-risk group (Figure 6).

Discussion

TYST has two peak onset periods, within 2 years after birth and $>$ 15 years of age. With the development of medical technology, the overall survival rate of TYST has gradually increased to 90% (1, 11). To further improve the quality of life of TYST patients and accurately predict the prognosis value of TYST patients concerned. In clinical practice, in addition to TNM staging, there is a lack of a model that can accurately predict the prognosis of patients with TYST.

Nomogram is a data-based graphical computing tool that predicts the risk of developing a disease by incorporating the American Joint Committee on Cancer (AJCC) TMN staging system and other important risk factors associated with prognosis (12, 13). Compared with traditional TMN staging, nomogram has better accuracy in predicting prognosis and can better provide advice and help for clinicians in diagnosis and treatment (14). To our knowledge, no prognostic nomogram of patients with TYST has been reported. Due to the low incidence of testicular yolk sac tumor, it isn't easy to collect a large sample size for a single-center study to obtain reliable conclusions. Therefore, it is very important to establish a reliable and accurate prognosis model for TYST. SEER database covers 18 countries and regions. The biggest advantage of the SEER database is that it can obtain clinical data with large samples, which is especially helpful for studying rare cases (4). This study collected data from the SEER database, effectively avoiding the insufficient sample size of TYST and single clinical data.

In this study, we developed and validated a novel nomogram to accurately predict the specific survival of patients with TYST. It is well known that the chance of genetic mutations causing cancer increases with age. Studies have shown that age plays a key role in the survival of various cancers (15, 16). A prediction study on non-spermatogonial cell tumors found that the survival rate of adolescent patients was higher than that of adults (17). Previous studies have reported that post-pubertal TYST often has a poor prognosis (18). Talerman et al. also found that the

TABLE 2 Univariate and multivariate analyses of CSS in training set.

	Univariate			Multivariate		
	HR	**95%CI**	***P***	**HR**	**95%CI**	***P***
Age	1.04	1.03–1.06	<0.001	1.027	1.01–1.05	0.003
Race						
White	1.22	0.3–5.06	0.782			
Black	0.98	0.35–2.75	0.974			
Other	0	0-Inf	0.997			
Marital						
Married	0.94	0.45–1.98	0.877			
Singled	2.77	0.85–8.99	0.09			
Divorced	0.64	0.08–5.07	0.675			
Laterality						
Left						
Right	8.74	4.26–17.97	<0.001	0.658	0.317–1.365	0.261
Unknown	0.76	0.38–1.5	0.424	3.1	1.405–6.839	0.005
T						
T1						
T2	1.53	0.46–5.07	0.49	2.147	0.622–7.413	0.227
T3	6.73	2.33–19.39	<0.001	3.377	1.113–10.248	0.032
T4	17.06	5.55–52.44	<0.001	5.498	1.702–17.759	0.004
TX	4.34	1.94–9.7	<0.001	3.851	1.554–9.543	0.004
M						
M0						
M1	14.54	6.05–34.95	<0.001	4.209	1.586–11.172	0.004
MX	3.87	1.4–10.68	0.009	1.382	0.42–4.555	0.595
AJCC						
I						
II	2.13	0.24–19.07	0.499			
III	16.24	5.73–46.04	<0.001			
Unknown	4.01	1.26–12.79	0.019			
Surgery						
No						
Yes	0.14	0.07–0.26	<0.001			
Radiotherapy						
No						
Yes	4.06	1.46–11.34	0.007			
Chemotherapy						
No						
Yes	7.32	3.41–15.72	<0.001	2.491	1.041–5.958	0.04

prognosis of adult testicular yolk sac tumor was worse than that of infants (19). Our study identified age as an independent risk factor for testicular yolk sac tumor-specific survival. In other words, the CSS rate of TYST patients decreased gradually with age.

The TNM staging system is often used to evaluate various malignant tumors, helps judge the prognosis of malignant tumors, and guides clinicians to adopt better treatment strategies (20, 21). Eighty to 85% of patients with TYST were stage I, confined to the primary part of the tumor. Radical testicular resection through the inguinal region often has a good prognosis (1, 3, 22). In this study, the operation was also found to be a specific protective factor for long-term survival. This study found that the T and M stages are independent risk factors for predicting the specific survival of patients with TYST. The prognosis is often poor for the high-level T and

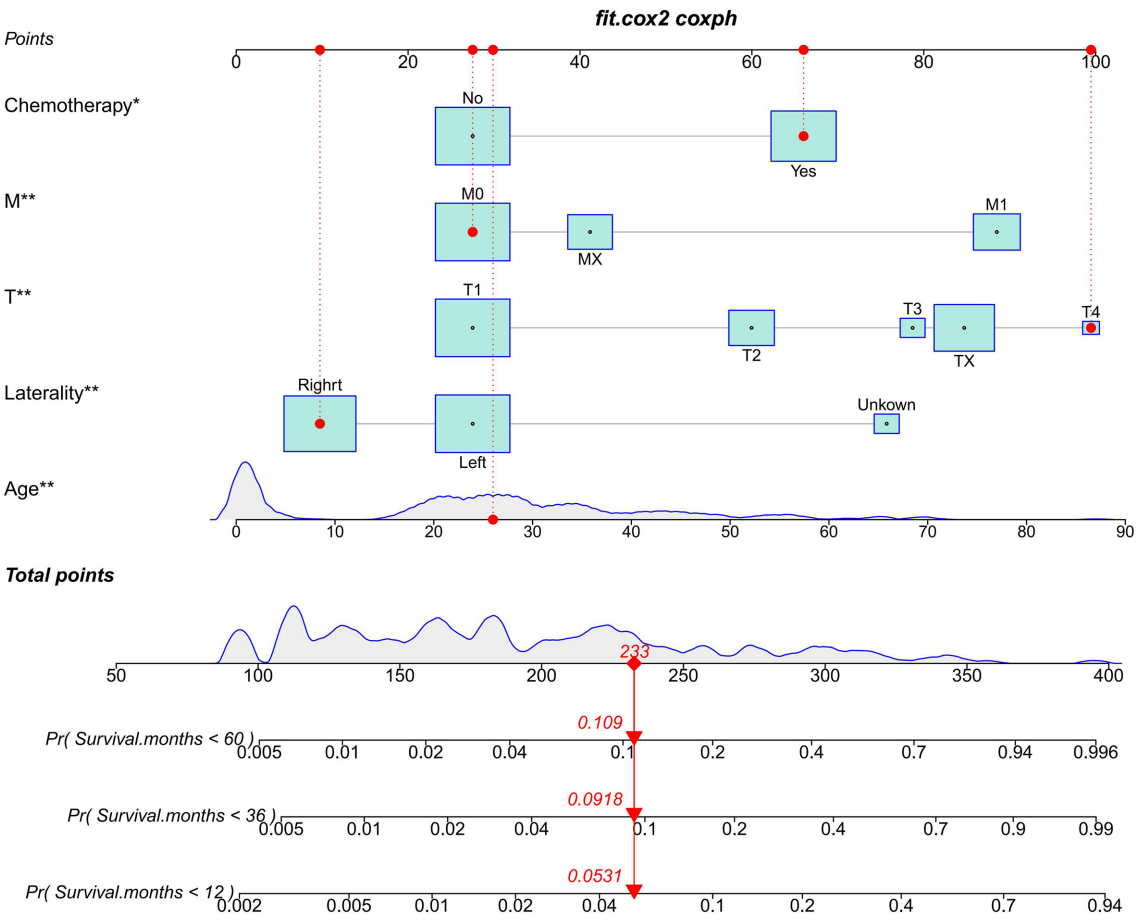


FIGURE 2
The nomogram predicts CSS in patients with TYST at 1-, 3-, and 5-year. * $P < 0.05$, ** $P < 0.01$.

M stages, consistent with previous studies' results (23). In our prediction model, the *n* stage was not an independent risk factor for patients with testicular yolk sac prognosis. For patients with retroperitoneal metastasis, retroperitoneal lymph node dissection is used to improve the prognosis of patients. However, this study found that 68 patients underwent retroperitoneal lymph node dissection (RPLND), and only 24 were positive. Standard retroperitoneal lymphadenectomy involves the removal of all lymphoid tissue from the hilum to 2cm distal to the common iliac artery bifurcation and from both sides to the level of the ureter. All large vessels need to be skeletal. RPLND is prone to erectile dysfunction, lymphatic leakage (lymphedema, chylous ascites), surgical injury (most commonly vascular injury and ureteral injury), surgical infection (lung, incision), intestinal obstruction and other complications (24, 25). Besides, infertility was the major long-term toxicity associated with RPLND (24). Therefore, routine retroperitoneal lymphadenectomy is not recommended for children with TYST (1, 5, 26), and even RPLND is mainly resection of retroperitoneal mass rather than radical RPLND. This study showed that retroperitoneal lymph node dissection did not significantly improve long-term survival. Accordingly, we do not recommend RPLND for TYST patients without a high risk of recurrence. This can improve a patient's quality of life without reducing the long-term specific survival rate.

Although postoperative radiotherapy has a good effect in treating various tumors and has been included in various cancer guidelines, TYST is not sensitive to radiotherapy, and it was found in this study that postoperative radiotherapy could not improve the cancer-specific survival rate of TYST patients. In addition, this study found that chemotherapy is the main risk factor affecting the prognosis of TYST, and chemotherapy patients often indicate a poor prognosis. Geng et al. reported similar results when they established a nomogram to predict the survival rate of yolk sac tumors (4). According to the 2015 European guidelines for Urology treatment, chemotherapy is not recommended for patients in stage I without high-risk factors. Chemotherapy is recommended for patients unwilling

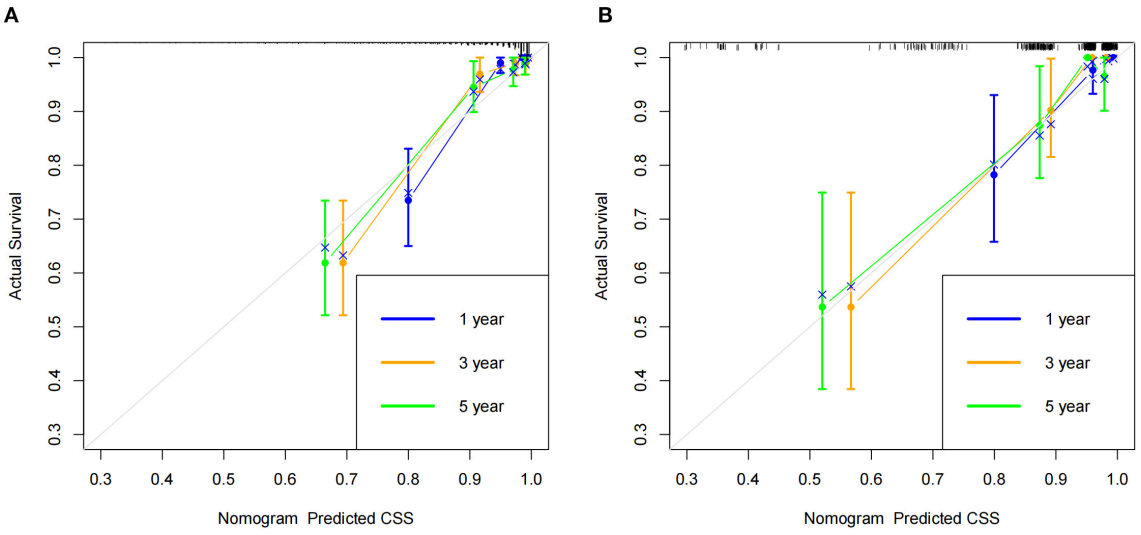


FIGURE 3
Calibration curve of the nomogram. **(A)** Calibration curves of 1-, 3-, and 5-year CSS in the training set; **(B)** Calibration curves of 1-, 3-, and 5-year CSS in the validation set.

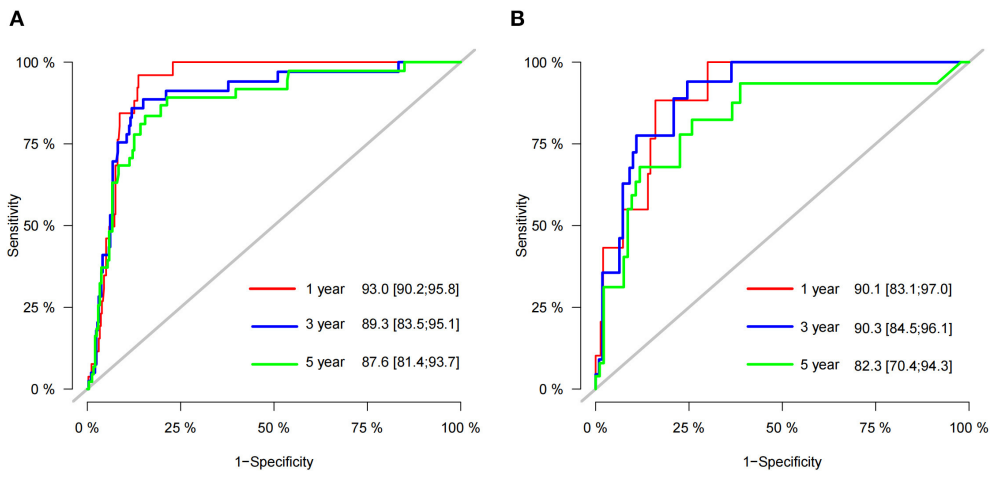


FIGURE 4
AUC for predicting 1-, 3-, and 5-year CSS in the training set **(A)** and validation set **(B)**.

to undergo surveillance or those with high-risk stage I (27). Our study showed doctors' importance in identifying patients with high-risk stage I TYST to avoid unnecessary chemotherapy. So 80–95% of TYST patients could be protected from the toxic effects of chemotherapy (24). Besides, chemotherapy treatment for high-stage patients might not predict a good prognosis. At the same time, it was impossible to further investigate the confounders due to the lack of details on chemotherapy and surgery.

Finally, the newly constructed Nomogram model for predicting the CSS rate of TYST patients includes multiple factors such as age at diagnosis, TM stage, and chemotherapy, which is convenient for clinical information collection. Although the International Germ Cell Cancer Collaborative Group (IGCCCG) is commonly used to predict the prognosis of metastatic non-spermatogenial germ cell tumor, it targets all testicular cancers and predict the individual 3-year progression-free survival rate of patients (28). Our present developed nomogram depending on SEER can accurately predict CSS at 1, 3, and 5 years in patients with TYST.

However, this study still has some limitations. First, although the SEER database collected clinical data of patients from

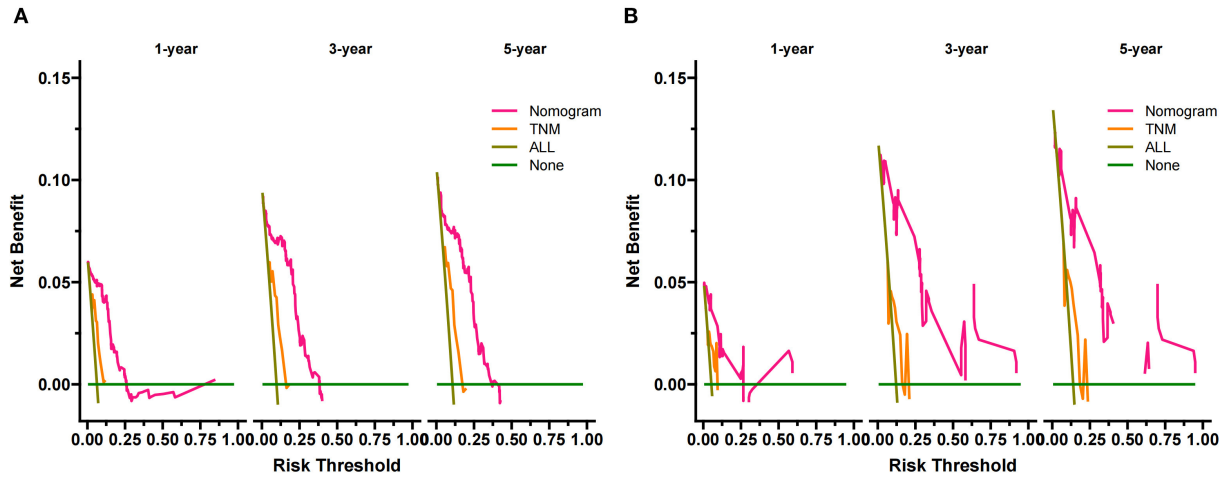


FIGURE 5
DCA of the nomogram in the training set **(A)** and the validation set **(B)**. The Y-axis represents a net benefit, and the X-axis represents threshold probability. The green line means no patients died, and the dark green line means all patients died. When the threshold probability is between 0 and 60%, the net benefit of the model exceeds all deaths or none.

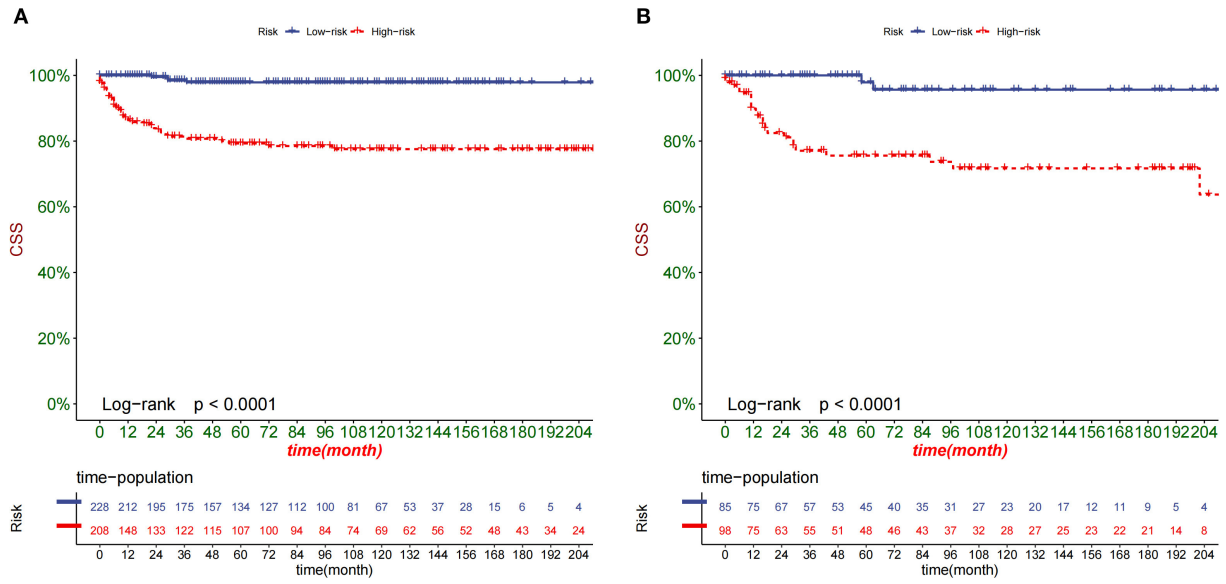


FIGURE 6
Kaplan-Meier curves of children in the low-risk and high-risk groups in the training set **(A)** and validation set **(B)**.

multiple medical centers in the United States, much detailed clinical information (the patient's family history, laboratory test results (AFP and LDH), maximum tumor diameter, surgical scope and intraoperative conditions, pathological history, chemotherapy regimen, recurrence and metastasis) were not recorded. This clinical information has been proven to be related to the prognosis of TYST (1, 6, 11, 28, 29), and adding such information can further improve the model's accuracy. Second, this study used an internal validation method to validate the nomogram model, and an external cohort was needed to further validate the accuracy of the model. Finally, since this study is retrospective, there must be selection bias. Therefore, prospective studies are needed to further validate the model.

Conclusion

In the study, we explored prognostic factors in TYST patients. The results showed that the patient's age, T stage, M stage and chemotherapy were independent risk factors affecting

the CSS of the patients. We constructed a nomogram to predict the CSS of TYST patients. After internal validation, the nomogram has been proven to have good accuracy and reliability. This predictive tool can help physicians predict the prognosis of TYST patients and develop appropriate monitoring and follow-up plans.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: The SEER data analyzed in this study is available at https://seer.Cancer.gov/.

Author contributions

ML and JW designed the study. JW, JL, YZ, and XZ collected and analyzed the data. ML drafted the initial manuscript and reviewed and edited the article. YL, CD, FL, and QP revised the article critically. All authors approved the final manuscript, contributed to the article, and approved the submitted version.

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.

References

1. Gilligan T, Lin DW, Aggarwal R, Chism D, Cost N, Derweesh IH, et al. Testicular Cancer, Version 2.2020, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. (2019) 17:1529–54. doi: 10.6004/jnccn.2019.0100

2. Jarvis H, Cost NG, Saltzman AF. Testicular tumors in the pediatric patient. *Semin Pediatr Surg*. (2021) 30:151079. doi: 10.1016/j.sempedsurg.2021.151079

3. Stein R, Quaedackers J, Bhat NR, Dogan HS, Nijman RJM, Rawashdeh YF, et al. EAU-ESPU pediatric urology guidelines on testicular tumors in prepubertal boys. *J Pediatr Urol*. (2021) 17:529–33. doi: 10.1016/j.jpurol.2021.06.006

4. Geng R, Zheng Z, Lin Y, Li Y, Ge G, Zhang J, et al. Clinical characteristics and prognostic factors of male yolk sac tumor: a surveillance, epidemiology, and end results program study. *World J Urol*. (2021) 39:1211–7. doi: 10.1007/s00345-020-03311-y

5. Grantham EC, Caldwell BT, Cost NG. Current urologic care for testicular germ cell tumors in pediatric and adolescent patients. *Urol Oncol*. (2016) 34:65–75. doi: 10.1016/j.urolonc.2015.06.008

6. Song G, Xiong GY, Fan Y, Huang C, Kang YM Ji GJ, et al. The role of tumor size, ultrasonographic findings, and serum tumor markers in predicting the likelihood of malignant testicular histology. *Asian J Androl*. (2019) 21:196–200. doi: 10.4103/aja.aja_119_18

7. O'Neill AF, Xia C, Krailo MD, Shaikh F, Pashankar FD, Billmire DF, et al. α-Fetoprotein as a predictor of outcome for children with germ cell tumors: a report from the malignant germ cell international consortium. *Cancer*. (2019) 125:3649–56. doi: 10.1002/cncr.32363

8. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. (2015) 16:e173-80. doi: 10.1016/S1470-2045(14)71116-7

9. Capogrosso P, Larcher A, Sjoberg DD, Vertosick EA, Cianflone F, Dell'Oglio P, et al. Risk based surveillance after surgical treatment of renal cell. *Carcinoma J Urol*. (2018) 200:61–7.

10. Parker WP, Cheville JC, Frank I, Zaid HB, Lohse CM, Boorjian SA, et al. Application of the stage, size, grade, and necrosis (SSIGN) score for clear cell renal cell carcinoma in contemporary patients. *Eur Urol*. (2017) 71:665–73. doi: 10.1016/j.eururo.2016.05.034

11. Ye YL, Zheng FF, Chen D, Zhang J, Liu ZW, Qin ZK, et al. Relapse in children with clinical stage I testicular yolk sac tumors after initial orchiectomy. *Pediatr Surg Int*. (2019) 35:383–9. doi: 10.1007/s00383-018-04426-5

12. Pan H, Shi X, Xiao D, He J, Zhang Y, Liang W, et al. Nomogram prediction for the survival of the patients with small cell lung cancer. *J Thorac Dis*. (2017) 9:507–18. doi: 10.21037/jtd.2017.03.121

13. Liao Y, Wang X, Zhong P, Yin G, Fan X, Huang C, et al. nomogram for the prediction of overall survival in patients with stage II and III non-small cell lung cancer using a population-based study. *Oncol Lett*. (2019) 18:5905–16. doi: 10.3892/ol.2019.10977

14. Duan J, Xie Y, Qu L, Wang L, Zhou S, Wang Y, et al. A nomogram-based immunoprofile predicts overall survival for previously untreated patients with esophageal squamous cell carcinoma after esophagectomy. *J Immunother Cancer*. (2018) 6:100. doi: 10.1186/s40425-018-0418-7

15. Zhanghuang C, Wang J, Zhang Z, Jin L, Tan X, Mi T, et al. A web-based prediction model for cancer-specific survival of elderly patients with clear cell renal cell carcinoma: a population-based study. *Front Public Health*. (2022) 9:833970. doi: 10.3389/fpubh.2021.833970

16. Dias-Santos D, Ferrone CR, Zheng H, Lillemoe KD, Fernández-Del Castillo C. The Charlson age comorbidity index predicts early mortality after surgery for pancreatic cancer. *Surgery*. (2015) 157:881–7. doi: 10.1016/j.surg.2014.12.006

17. Amini A, Waxweiler TV, Maroni PD, Kessler ER, Cost CR, Greffe BS, et al. Survival outcomes of adolescent and adult patients with non-seminomatous testicular germ-cell tumors: a population-based study. J Pediatr Urol. (2016) 12:405. doi: 10.1016/j.jpurol.2016.06.014

18. Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM. The (2016). WHO Classification of tumors of the urinary system and male genital organs-Part A: renal, penile, and testicular tumors. *Eur Urol*. (2016) 70:93–105. doi: 10.1016/j.eururo.2016.02.029

19. Talerman A. The incidence of yolk sac tumor (endodermal sinus tumor) elements in germ cell tumors of the testis in adults. *Cancer*. (1975) 36:211–5.

20. Zhou H, Zhang Y, Song Y, Tan W, Qiu Z, Li S, et al. Marital status is an independent prognostic factor for pancreatic neuroendocrine tumors patients: an analysis of the Surveillance, Epidemiology, and End Results (SEER) database. *Clin Res Hepatol Gastroenterol*. (2017) 41:476–86. doi: 10.1016/j.clinre.2017.02.008

21. Zahoor H, Elson P, Stephenson A, Haber GP, Kaouk J, Fergany A, et al. Patient characteristics, treatment patterns and prognostic factors in squamous cell bladder cancer. *Clin Genitourin Cancer*. (2018) 16:e437–42. doi: 10.1016/j.clgc.2017.10.005

22. Grady RW. Current management of prepubertal yolk sac tumors of the testis. *Urol Clin North Am*. (2000) 27:503–8. doi: 10.1016/S0094-0143(05)70097-5

23. Frazier AL, Hale JP, Rodriguez-Galindo C, Dang H, Olson T, Murray MJ, et al. Revised risk classification for pediatric extracranial germ cell tumors based on 25 years of clinical trial data from the United Kingdom and United States. *J Clin Oncol.* (2015) 33:195–201. doi: 10.1200/JCO.2014.58.3369

24. Pierorazio PM, Cheaib JG, Patel HD, Gupta M, Sharma R, Zhang A, et al. Comparative effectiveness of surveillance, primary chemotherapy, radiotherapy and retroperitoneal lymph node dissection for the management of early stage testicular germ cell tumors: a systematic review. *J Urol.* (2021) 205:370–82. doi: 10.1097/JU.0000000000001364

25. O'Shea K, Tong A, Farrelly P, Craigie R, Cheesman E, Shukla R, et al. Management and outcome of pediatric testicular tumors—a 20 year experience. *J Pediatr Surg.* (2021) 56:2032–6. doi: 10.1016/j.jpedsurg.2021.02.063

26. Cornejo KM, Frazier L, Lee RS, Kozakewich HP, Young RH. Yolk sac tumor of the testis in infants and children: A clinicopathologic analysis of 33 cases. *Am J Surg Pathol.* (2015) 39:1121-31. doi: 10.1097/PAS.0000000000000432

27. Albers P, Albrecht W, Algaba F, Bokemeyer C, Cohn-Cedermark G, Fizazi K, et al. European association of urology: guidelines on testicular cancer: update. *Eur Urol.* (2015) 68:1054–68. doi: 10.1016/j.eururo.2015.07.044

28. Gillessen S, Sauvé N, Collette L, Daugaard G, de Wit R, Albany C, et al. International germ cell cancer classification update consortium. predicting outcomes in men with metastatic nonseminomatous germ cell tumors (NSGCT): results From the IGCCCG Update consortium. *J Clin Oncol.* (2021) 39:1563–74. Erratum in: *J Clin Oncol.* (2022) 40:2283. doi: 10.1200/JCO.20.03296

29. Seidel C, Daugaard G, Tryakin A, Necchi A, Cohn-Cedermark G, Ståhl O, et al. The prognostic impact of different tumor marker levels in nonseminomatous germ cell tumor patients with intermediate prognosis: a registry of the International Global Germ Cell Tumor Collaborative Group (G3). *Urol Oncol.* (2019) 37:809. doi: 10.1016/j.urolonc.2019.07.020

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 20 October 2022
DOI 10.3389/fpubh.2022.1034772

OPEN ACCESS

EDITED BY
Yu-Dong Zhang,
University of Leicester,
United Kingdom

REVIEWED BY
Jatinderkumar R. Saini,
Symbiosis Institute of Computer Studies and Research (SICSR), India
Francisco Solano,
University of Murcia, Spain

*CORRESPONDENCE
Yaping Lu
luyaping@sinopharm.com
Liuqing Chen
chlq35@126.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 02 September 2022
ACCEPTED 30 September 2022
PUBLISHED 20 October 2022

CITATION
Wan L, Ai Z, Chen J, Jiang Q, Chen H, Li Q, Lu Y and Chen L (2022) Detection algorithm for pigmented skin disease based on classifier-level and feature-level fusion.
Front. Public Health 10:1034772.
doi: 10.3389/fpubh.2022.1034772

Detection algorithm for pigmented skin disease based on classifier-level and feature-level fusion

Li Wan[1,2†], Zhuang Ai[3†], Jinbo Chen[1†], Qian Jiang[1], Hongying Chen[1], Qi Li[3], Yaping Lu[3*] and Liuqing Chen[1*]

[1]Dermatology Department, Wuhan No.1 Hospital, Hubei, China, [2]Dermatology Hospital of Southern Medical University, Guangzhou, China, [3]Department of Research and Development, Sinopharm Genomics Technology Co., Ltd., Jiangsu, China

Pigmented skin disease is caused by abnormal melanocyte and melanin production, which can be induced by genetic and environmental factors. It is also common among the various types of skin diseases. The timely and accurate diagnosis of pigmented skin disease is important for reducing mortality. Patients with pigmented dermatosis are generally diagnosed by a dermatologist through dermatoscopy. However, due to the current shortage of experts, this approach cannot meet the needs of the population, so a computer-aided system would help to diagnose skin lesions in remote areas containing insufficient experts. This paper proposes an algorithm based on a fusion network for the detection of pigmented skin disease. First, we preprocess the images in the acquired dataset, and then we perform image flipping and image style transfer to augment the images to alleviate the imbalance between the various categories in the dataset. Finally, two feature-level fusion optimization schemes based on deep features are compared with a classifier-level fusion scheme based on a classification layer to effectively determine the best fusion strategy for satisfying the pigmented skin disease detection requirements. Gradient-weighted Class Activation Mapping (Grad_CAM) and Grad_CAM++ are used for visualization purposes to verify the effectiveness of the proposed fusion network. The results show that compared with those of the traditional detection algorithm for pigmented skin disease, the accuracy and Area Under Curve (AUC) of the method in this paper reach 92.1 and 95.3%, respectively. The evaluation indices are greatly improved, proving the adaptability and accuracy of the proposed method. The proposed method can assist clinicians in screening and diagnosing pigmented skin disease and is suitable for real-world applications.

KEYWORDS

fusion network, pigmented skin disease, attention mechanism, image style transfer, model interpretability

1. Introduction

Skin, as the first layer of protection for the human body, has important physiological protection functions, such as excretion, regulating body temperature and feeling external stimuli. It is also the largest organ in the human body. However, the incidence of skin diseases is extremely high, and there are many types of skin diseases, among which pigmented skin lesions are common; most pathological areas are black, brown or other dark colors, which is mainly due to the increase or decrease in regional melanin caused by ultraviolet radiation or other external factors. In 2021, skin melanoma in pigmented skin disease accounts for 5.6% of all new cancers in the United States, and the number of skin melanoma patients has increased at an annual rate of ∼1.4% over the past 10 years (1). However, melanoma that is detected early has a very high cure rate. Studies have shown that if abnormal skin melanocyte proliferation is found early, the survival rate is 96%. If late-stage melanoma is detected, the survival rate is reduced to only 5% (2), and its color is easily confused with that of other common skin pigmented diseases, leading to misdiagnosis. The diagnosis of pigmented skin lesions requires trained specialists, but the number of specialist doctors is grossly inadequate compared to the number of cases. Therefore, it is necessary to develop an algorithm for the automatic diagnosis of pigmented skin lesions.

In recent years, deep learning has been widely used in feature extraction, object classification and detection. Compared with machine learning, deep learning can automatically and efficiently extract features from medical images. Since 2012, various deep Convolutional Neural Network (CNN) models based on the "ImageNet" dataset have been proposed. AlexNet (ImageNet classification with deep convolutional neural networks), a network architecture proposed by Krizhevsky et al. (3), was the winner of the first ImageNet Challenge classification task in 2012; ZFNet (4) (Visualizing and understanding convolutional networks) is a large convolutional network based on AlexNet; VGGNET (5) (Very deep convolutional networks for large-scale image recognition) was proposed by Visual Geometry Group (VGG), a famous research group at Oxford University, and won the first place in localization and the second place in classification in that year's ImageNet competition. GoogleNet (6) (Going deeper with convolutions) was proposed by the Google team and won the first place in the ImageNet competition for the classification task; ResNet (7) (Deep residual learning for image recognition), proposed by Microsoft Research, won the first place in classification task and the first place in target detection in that year's ImageNet competition, and the first place in target detection and image segmentation in COCO dataset. ResNeXt (8) (Aggregated residual transformations for deep neural networks) is a new image classification network proposed by Kaiming He's team at CVPR 2017. ResNeXt is an upgraded version of ResNet; SENET (9) (Squeeze-and-Excitation Networks) is a new image recognition architecture announced by the self-driving company Momenta in 2017. This structure is the first place in the ImageNet competition in that year in the classification task; NASNet (Learning Transferable Architectures for Scalable) is a deep network model proposed by Zoph et al. (10) that can automatically generate network structures without manually designing network models; EfficientNet (11) (EfficientNet: Rethinking model scaling for convolutional neural networks) is proposed by Google team to obtain better performance by deepening the model, widening the model or increasing the resolution of the model input. These network models have ranked highly in competitions. The prediction effects of different network structures in various fields are inconsistent, so researchers cannot quickly find appropriate network models. Many scholars have thus conducted research to solve this problem. Researchers must test the outstanding network models one by one to find the most appropriate network model for their scenario (12–15). This strategy wastes time and resources. Therefore, an ensemble network can obtain an algorithmic model that is better than the model produced by the best individual network by setting the weights of different networks (16–18). However, at present, most network fusion approaches use majority voting, mean voting or the weights of the base classifiers to obtain the output of various networks through one-to-one testing, which cannot give full play to the various effects of different classifiers on different tasks. Therefore, this paper proposes a variety of fusion strategies and optimizes the weight of each classifier through the loss function of the network model to fully utilize the ability of each classifier for the detection of pigmented skin diseases.

Therefore, building a pigmented skin disease detection algorithm based on classifier-level and feature-level fusion encounters the following problems.

(1) How to handle unbalanced pigmented skin disease datasets.
(2) How to build an effective network fusion strategy.

2. Related work

In recent years, the applications of Artificial Intelligence (AI) in various fields have developed rapidly, especially in the fields of medical image analysis and bioinformatics. At present, AI is widely used in skin cancer diagnosis (19–21). From the point of view of whether features can be extracted automatically, the AI approaches in this area can be divided into skin cancer classification methods based on machine learning and skin cancer classification methods based on deep learning.

Skin cancer classification based on machine learning generally involves manually extracting image features and then inputting the extracted features into a machine learning algorithm to obtain classification results (22–25). Varalakshmi (26) first used an upsampling method called the Synthetic Minority Oversampling Technique (SMOTE) to balance his

dataset, greatly improving the accuracy of various machine learning models. The accuracies of different machine learning algorithms were then analyzed. Support Vector Machine (SVM) algorithms with polynomial kernels provide better accuracy than other machine learning algorithms, such as decision trees using Gini indices and entropy, naive Bayes classifiers, extreme gradient boosting (XGBoost) classifiers, random forests, and logistic regression algorithms. Sabri (19) first extracted the shapes, colors, textures and skeletons of skin image lesions, then used the information gain method to determine the best combination of features, and finally input this feature combination into a commonly used machine learning algorithm to predict the categories of legions. Vidya (27) first extracted skin image asymmetry, border, color, and diameter information. A Histogram of Oriented Gradients (HOG) and a Gray Level Co-occurrence Matrix (GLCM) were used to extract texture features. The extracted features were passed directly to classifiers utilizing different machine learning techniques [such as an SVM, K-Nearest Neighbors (KNN) and a naive Bayes classifier] to classify skin lesions as benign or melanoma. Kalwa (28) presents a smartphone application that combines image capture capabilities with preprocessing and segmentation to extract the Asymmetry, Border irregularity, Color variegation, and Diameter (ABCD) features of a skin lesion. Using the feature sets, classification of malignancy is achieved through support vector machine classifiers.

Skin cancer classification approaches based on deep learning usually adopt a network model for automatic feature extraction, and thus feature extraction and classification can be completed in the same algorithm (20, 21, 29–31). Skin cancer detection algorithms based on deep learning can be divided into single-classifier detection methods and fusion detection methods based on multiple classifiers according to the number of utilized classifiers.

Based on single-classification detection, Sevli (32) proposed using a CNN model to classify seven different skin lesions in the HAM10000 dataset, and the model achieved 91.51% classification accuracy. The model linked its results to a web application and was assessed in two stages by seven dermatologists. Milton (12) first appropriately processed and enhanced skin images and then carried out experiments on various neural networks, including the progressive NASNet (PNASNet)-5-Large, InceptionResNet V2, SENet154, InceptionV4, etc. Finally, the PNASNet-5-Large model achieved the best validation result of 0.76.

Regarding detection based on multiple classifiers, Pal (33) solved the data imbalance problem in the training dataset by setting a propagation-weighted loss from the loss correspondence. For classifier model construction, the pretraining weights of these models were fine-tuned (by ResNet50, DenseNet-121, and MobileNet). Finally, the average category prediction probabilities obtained from these trained networks were used to determine the category labels of the test images. Xie (34) used four pretrained ResNet50 networks to characterize the multiscale information of skin lesions and combined them by using adaptive weighting schemes that could be learned during error propagation. The proposed model achieved an average Area Under Curve (AUC) value of 86.5% on the official ISIC-Skin 2018 validation database. Aldwgeri (35) aimed to solve the data imbalance problem in the training dataset and realized the equalization of each category through flipping, rotation, shifting, and scaling techniques. The equalized image data were then input into different pretraining models, including VGG-Net, ResNet50, Inception V3, Xception, and DenseNet-121. The outputs of the five pretraining models were averaged to produce the final prediction results.

Therefore, the innovations of this paper include the following aspects.

(1) An image style transfer algorithm is applied to the detection of pigmented skin diseases for the purpose of image augmentation.
(2) To prevent image augmentation noise, the required upsampling image is applied to each class image.
(3) Attention mechanisms and common network architectures should be combined to achieve improved detection efficiency.
(4) Two feature-level fusion optimization schemes based on deep features and a classifier-level fusion method based on a classification layer are proposed.
(5) Two visualization algorithms, Grad_CAM and Grad_CAM++, are used to verify the validity of the fusion network.

3. Detection algorithm for pigmented skin diseases based on classifier-level and feature-level fusion

3.1. System architecture

This paper proposes a detection algorithm for pigmented skin diseases based on a fusion network (Figure 1). This approach can be divided into three modules: image preprocessing, image augmentation, and model building and prediction.

Image preprocessing: First, the obtained pigmented skin disease images are normalized, and the pixel values of the images are limited to 0–1, which can effectively reduce the number of calculations required for the images in the neural network. Then, the height and width of each normalized image are unified to 450*600 (*via* nearest-neighbor interpolation). Finally, the preprocessed image dataset (three-channel color images with heights of 450 and widths of 600) for pigmented skin diseases can be obtained. As seen from Figure 1, the proportions of the different categories after image pretreatment are seriously

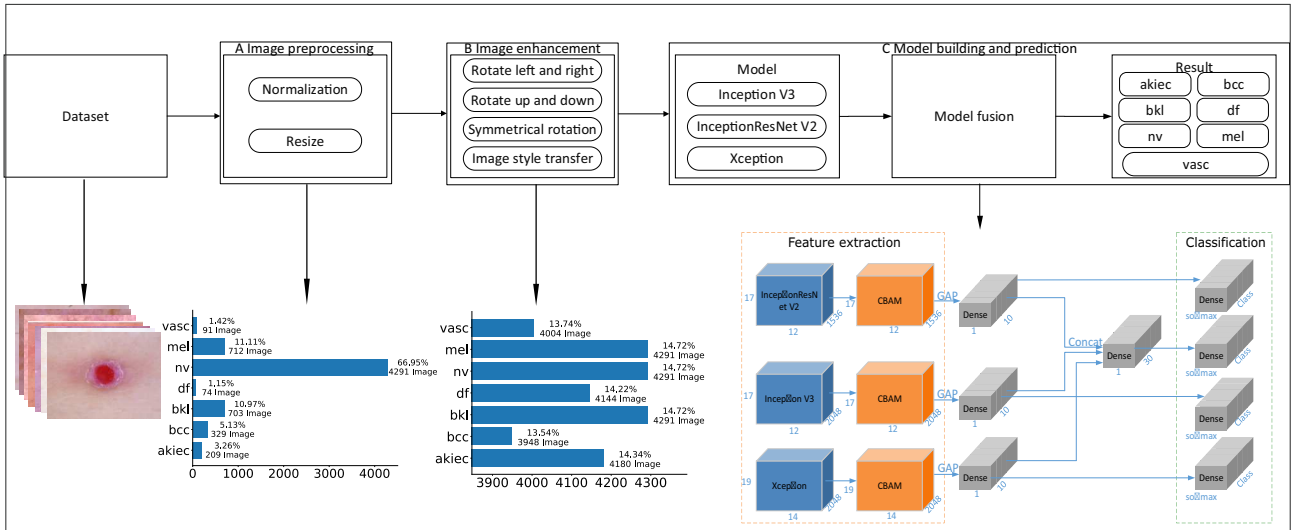


FIGURE 1
An overview of the proposed method. **(A)** "Image preprocessing," including image normalization and image resizing, is performed on an input image before feature extraction. **(B)** "Image augmentation," including operations such as image rotation and image style transfer, performs upsampling on the input image to keep the various categories in a balanced state. **(C)** "Model building and prediction" carries out model training and prediction on the input image, wherein the feature extraction part is the fusion of three base classifiers and an attention mechanism [the convolutional block attention module ("CBAM")], "GAP" denotes global average pooling, "Dense" is a fully connected layer, "Concat" is the fusion of the output results of the three branches, and "Class" is the number of categories. In this article, Class is 7. "Softmax" is the activation function of the classification output layer, "Classification" is the prediction result output layer, and the number represents the change in the characteristic dimensionality at each stage.

unbalanced; among them, the "nv" category occupies 66.95% of the dataset. If no processing is performed, the neural network will seriously prefer this category in model training.

Image augmentation: As the nv category accounts for 66.95% of the dataset, if dataset balance needs to be achieved, other categories need to be upsampled. First, skin images (except those in the nv category) are preprocessed by turning them left and right, reversing up and down, symmetric rotation (the calculation process is shown in Algorithm 1) and performing image style transfer (the calculation process is shown in Algorithm 2) to achieve a balance between the various categories of images. As seen from Figure 1, the proportion of each category after image augmentation is relatively balanced, accounting for ∼14% of the whole dataset of pigmentosa skin disease images.

Model building and prediction: The enhanced images of pigmented skin diseases are first input into three different base classifiers (i.e., Inception V3, InceptionResNet V2, and Xception), and the outputs of the three base classifiers are then fused. Finally, the fusion result is used as the pigmented skin disease prediction result.

Input: Dataset after image preprocessing : *Data*.
Output: training set, validation set, test set.

Define the list of stored images after augmentation: *Data_train_process* = [].
The *Data* are divided into a training set *Data_train*, a validation set *Data_valid* and a test set *Data_test* at a 3:1:1 ratio.
for *image* → *Data_train* **do**
  if 'image' belongs to category 'nv' **then**
    Continue.
  end if
  Add *image* to *Data_process*.
  Rotates_l_r = Rotate *image* left and right.
  Add *Rotates_l_r* to *Data_train_process*.
  Rotates_u_d = Rotate *image* up and down.
  Add *Rotates_u_d* to *Data_train_process*.
  Rotates_s = Rotate *image* Symmetrical.
  Add *Rotates_s* to *Data_train_process*.
end for
return *Data_train_process*, *Data_valid*, *Data_test*.

Algorithm 1. Image augmentation—rotation.

3.2. Image preprocessing module

3.2.1. Dataset

The dataset used in this paper is provided by Tschandl et al. (36), and it contains 10,015 pictures of seven types of skin diseases. Cases include a representative collection of all import diagnostic categories in the realm of pigmented lesions. The seven types are melanocytic Nevi (nv), Melanoma (mel), Benign Keratosis-like Lesions (solar lentigines/seborrheic keratoses and lichen-planus-like keratoses) (bkl), Basal Cell Carcinoma (bcc), Actinic Keratoses and Intraepithelial

Input: Training set after image rotation: *Data_train_process*.
Output: training set *Data_train_augmentation*.
Define the list of stored images after augmentation: *Data_train_augmentation* = [].
Obtain a set of images for each category in *data_train*: data_train0, data_train1, data_train2, data_train3, data_train4, data_train5, data_train6.
data_train_list = (data_train0, data_train1, data_train2, data_train3, data_train5, data_train6).
for *data_train_i* → *data_train_list* **do**
  Calculate the difference between the sample sizes of category *data_train_i* and category *data_train4* (nv sample): *numSub*.
  According to *data_train_i* and *numSub*, calculate the number of images to be upsampled for each category: *numAdd*.
  if *numAdd* ≥ 1 **then**
    for *contentImage* → *data_train_i* **do**
      numAdd images are randomly selected from *data_train_i*: *styleImageList*.
      for *styleImage* → *styleImageList* **do**
        Perform image style transfer using the style image *styleImage* and *contentImage*: *newImage*.
        Add the image *newImage* to *Data_train_augmentation*.
      end for
    end for
  else
    numSub images are randomly selected from *data_train_i*: *contentImageList*.
    for *contentImage* → *contentImageList* **do**
      A images are randomly selected from *data_train_i*: *StyleImage*.
      Perform image style transfer using the style image *styleImage* and *contentImage*: *newImage*.
      Add the image *newImage* to *Data_train_augmentation*.
    end for
  end if
end for
return *Data_train_augmentation*.

Algorithm 2. Image augmentation—image style transfer.

Carcinoma/Bowen's disease (akiec), Vascular lesions (angiomas, angiokeratomas, pyogenic granulomas, and hemorrhage) (vasc), and Dermatofibroma (df). The corresponding amounts of image data are 6,705, 1,113, 1,099, 514, 327, 142, and 115, respectively. The proportion of each category is shown in Figure 2A. Typical images for each category are shown in Figure 2B. In Figure 2A, the selected dataset of pigmented skin diseases is severely imbalanced between categories, and the imbalance in the dataset causes the model to completely bias the prediction results to the side with a large sample size (18), and the model does not have any prediction effect on the other categories of sample classification, so a processing step for the imbalance in the dataset is necessary.

3.2.2. Image preprocessing and augmentation

We first preprocess the acquired skin disease dataset (36) to obtain high-quality image data. In the preprocessing step, each image is first reduced to the specified size of 450*600, and then each pixel of the image is normalized according to Equation (1). In this way, the image is easy for the network to calculate. The image preprocessing part is transformed from Figure 3A to Figure 3B.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

The dataset presents great disparities among the amounts of image data contained in various categories. Without performing certain processing steps, the prediction results will be greatly affected by this unbalanced dataset. Therefore, we must upsample the image data to obtain a balanced image dataset. First, we carry out the following basic operations on the images (except for those in the nv category): left and right mirror rotation, up and down mirror rotation, symmetric rotation, etc.; these operations can balance the images to a certain extent. The left and right mirror rotation operations mirror the original image with respect to its vertical centerline. The upper and lower mirror rotation operations mirror the original image with respect to its horizontal centerline. Symmetric rotation is an image transformation that flips the original image left and right before flipping them again in the up and down directions. After completing the basic image operations, the image data contained in different image categories are shown in Table 1. The basic image augmentation operation can be converted from Figure 3B to Figure 3C.

It can be seen from Table 1 that the numbers of images in various categories are still seriously imbalanced, so we adopt an image style transfer algorithm (37) to upsample the images. The image style transfer algorithm proposed by Ghiasi has been successfully trained on a corpus of ∼80,000 paintings. In addition, it can be generalized to previously unobserved images.

First, this paper calculates the sample size differences between nv and the other categories in the image dataset according to Equation (2) and then divides each difference by the sample size of the corresponding category to obtain the sample size "n" that needs to be randomly added to the other categories. The image to be upsampled is selected as the "content image," "*n*" images are randomly selected from the image samples of

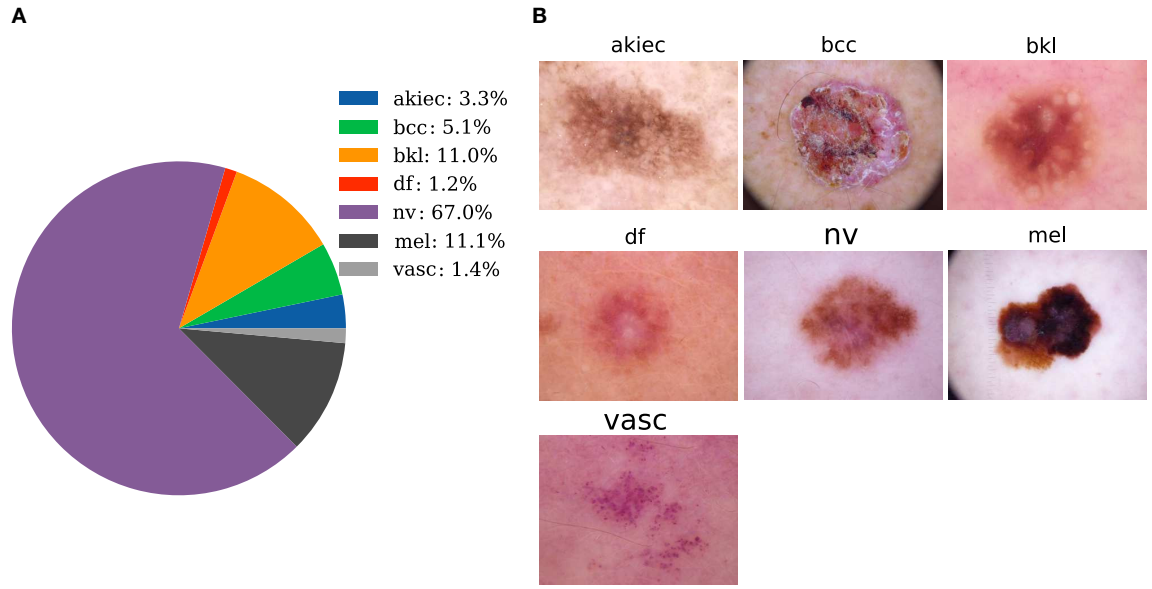


FIGURE 2
Data preparation. **(A)** The distribution of pigmented skin diseases in the specimens. **(B)** Photos of the representative pigmented skin diseases in each sampling category for clinical diagnosis.

this category as the "style images," and the "content image" and "*n*" "style images" are input into the image style transfer model in turn to obtain "*n*" upsampling images generated by the fusion of the "content image" and "style images" (the calculation process is shown in Algorithm 2). After performing image style transfer, the amount of data in each category is shown in Table 1. An example diagram of image style transfer is shown in Figures 3C–E.

$$Add_n = \frac{Num(Class\_nv) - Num(Class\_i)}{Num(Class\_i)} \quad (2)$$

In the equation, i represents the akiec, bcc, bkl, df, mel, and vasc categories; Num(Class_i) represents the data volume of the selected category. If Add_n is <1, it indicates that the data volume of this category is not very different from that of nv. In this paper, the number of data differences is randomly extracted for image style transfer.

3.3. Model building and prediction module

The base classifier of the fusion network used in this paper can consist of Inception V3, InceptionResNet V2, and Xception. The fusion part explores feature-level fusion based on deep features and classifier-level fusion based on a classification layer.

Feature-level fusion based on deep features has been proven to be an efficient fusion strategy (38–42) that can combine features extracted from N networks into a single feature vector containing more image information. Feature-level fusion techniques can be divided into parallel feature-level fusion and serial feature-level fusion based on whether the feature dimensions output by the networks are consistent. Three methods are available for realizing parallel feature fusion: summing up each feature (Equation 3); averaging each feature (Equation 4); and executing the max operation (Equation 5) for each feature. Serial feature fusion can only realize feature splicing (Equation 6) according to the channel dimension because of the inconsistency of the feature output dimensions. The classifier-level fusion method based on a classification layer can make the features extracted from *N* networks remain unchanged and perform feature splicing at the output of the classification layer.

$$F_{feature\_level\_fusion} = \sum_{i=1}^{N} F_i \quad (3)$$

$$F_{feature\_level\_fusion} = \frac{1}{N}\sum_{i=1}^{N} F_i \quad (4)$$

$$F_{feature\_level\_fusion} = max(F_1, F_2, F_3, ..., F_N) \quad (5)$$

$$F_{Decision\_level\_fusion} = Concat(F_1, F_2, F_3, ..., F_N) \quad (6)$$

When the input picture size is (Batch, 450, 600, 3), the output dimensions of Inception V3 are (Batch, 12, 17, 2048), the output dimensions of InceptionResNet V2 are (Batch, 12, 17, 1536), and the output dimensions of Xception are (Batch,

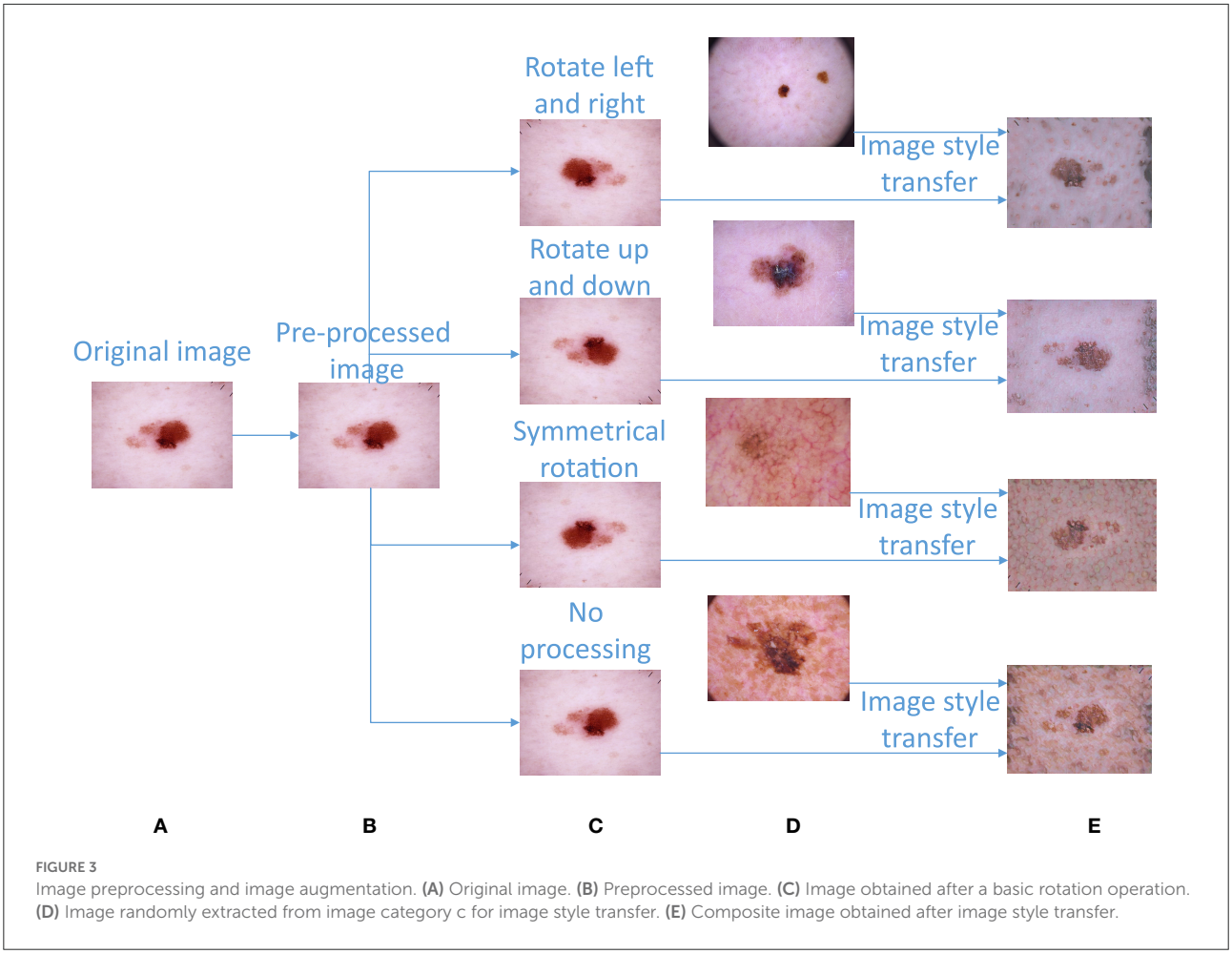


FIGURE 3
Image preprocessing and image augmentation. **(A)** Original image. **(B)** Preprocessed image. **(C)** Image obtained after a basic rotation operation. **(D)** Image randomly extracted from image category c for image style transfer. **(E)** Composite image obtained after image style transfer.

TABLE 1 Image number statistics during image preprocessing.

Category	Number of original images	Number of images after basic operations	Number of images to be added per image	Number of images after image style transfer
akiec	209	836	4	4,180
bcc	329	1,316	2	3,948
bkl	703	2,812	0.5	4,291
df	74	296	13	4,144
nv	4,291	4,291	0	4,291
mel	712	2,848	0.5	4,291
vasc	91	364	10	4,004

14, 19, 2048). In this paper, feature-level fusion based on deep features employs the output fusion results of three different networks, and the dimensions of the outputs of the three models are inconsistent. Therefore, we optimize the feature-level fusion strategy based on deep features. In the first method, the convolution layer is used to convert the feature map to achieve dimensional consistency. The dimension conversion method is shown in Equations (7) and (8), and the overall algorithm flow is shown in Figure 4A.

$$W_{out} = \frac{W_{in} - F + 2P}{S} + 1 \tag{7}$$

$$H_{out} = \frac{H_{in} - F + 2P}{S} + 1 \tag{8}$$

In the equation, W_{in} and H_{in} are the width and height of the input, F is the size of the filter, P is the padding

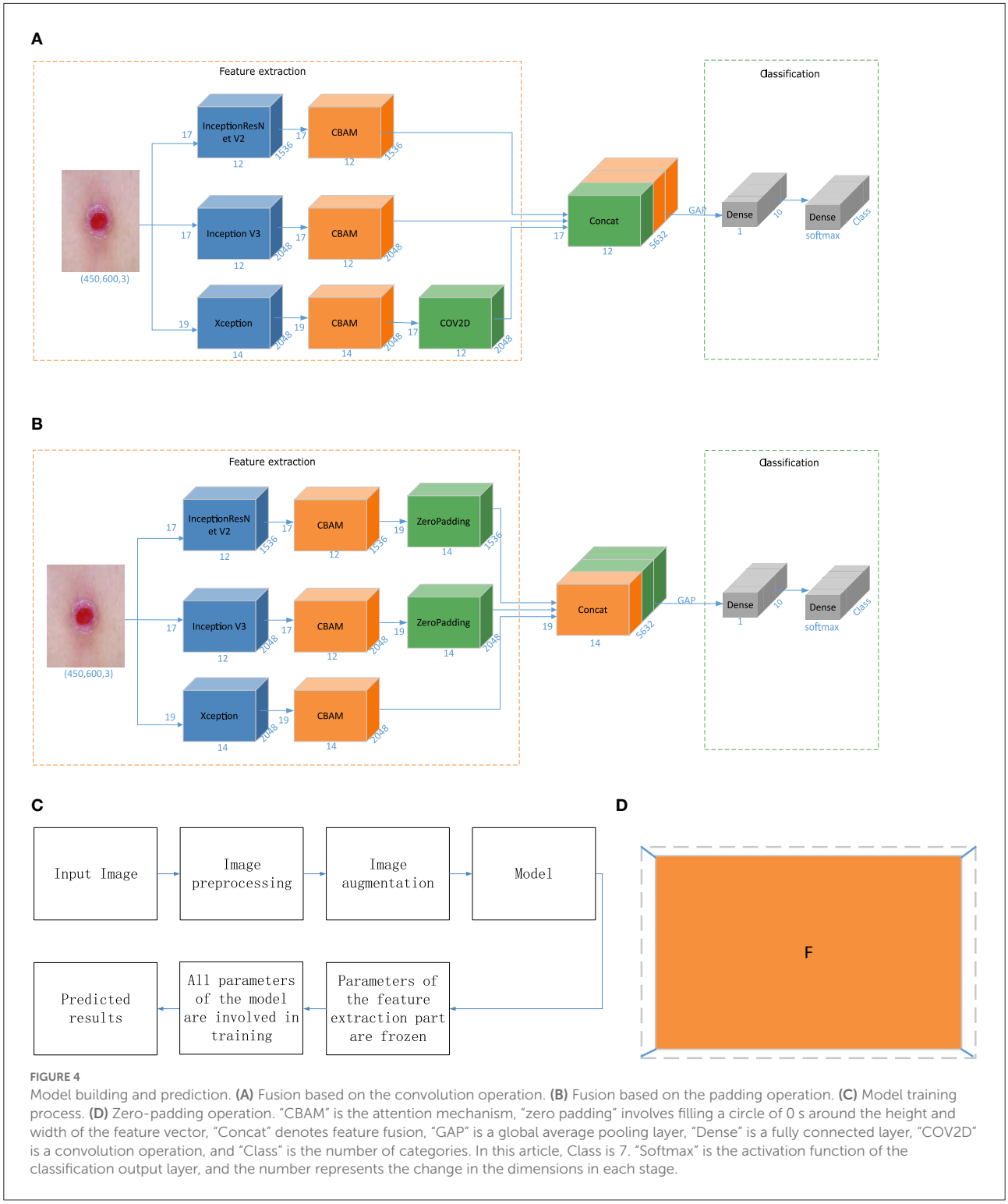


FIGURE 4
Model building and prediction. **(A)** Fusion based on the convolution operation. **(B)** Fusion based on the padding operation. **(C)** Model training process. **(D)** Zero-padding operation. "CBAM" is the attention mechanism, "zero padding" involves filling a circle of 0 s around the height and width of the feature vector, "Concat" denotes feature fusion, "GAP" is a global average pooling layer, "Dense" is a fully connected layer, "COV2D" is a convolution operation, and "Class" is the number of categories. In this article, Class is 7. "Softmax" is the activation function of the classification output layer, and the number represents the change in the dimensions in each stage.

size, S is the step size, and W_{out} and H_{out} are the final width and height, respectively. W_{in} and H_{in} are 14 and 19, and W_{out} and H_{out} are 12 and 17, respectively. Therefore, according to this equation, we set F as 3, P as 0, and S as 1. The output can realize the splicing of the three dimensions.

In the second method, in this paper, the outputs of Inception V3 and InceptionResNet V2 are surrounded by a circle of 0s

TABLE 2 Evaluation criteria.

Evaluation criteria	Equation	Meaning
Accuracy (Acc)	$\frac{TP+TN}{TN+FP+FN+TP}$	The proportion of all results correctly judged by the classification model to the total sample size
Sensitivity=Recall	$\frac{TP}{TP+FN}$	The proportion of correct model predictions among all the results whose true values are positive
Specificity	$\frac{TN}{TN+FP}$	The proportion of correct model predictions among all the results whose true values are negative
Precision	$\frac{TP}{TP+FP}$	The proportion of correct model predictions among all the results for which the predicted value of the model is positive
F1	$\frac{2*Precision*Recall}{Precision+Recall}$	Harmonic mean of precision and recall
Weighted avg	$\frac{\sum_{i=1}^{class\_num} P\_i*support\_i}{\sum_{i=1}^{class\_num} support\_i}$	The weighted average of evaluation indicators for each category, with the weight being the proportion of the sample size of each category in the total sample size. "support_i" represents the number of samples in category "i," "P_i" represents the score value of the evaluation index of category "i," and "class_num" represents the number of categories.
AUC		Area under the receiver operating characteristic (ROC) curve

TABLE 3 The influence of the data imbalance treatment scheme on the results.

Data imbalance processing		Original set	Class weight	Image rotation	Pixel
Precision	akiec	0.6515	0.557	0.7368	0.7167
	bcc	0.7364	0.6056	0.787	0.8333
	bkl	0.7014	0.7125	0.8144	0.7991
	df	0.7692	0.5714	1	0.8
	nv	0.8957	0.8801	0.8877	0.9155
	mel	0.7083	0.6589	0.7517	0.7861
	vasc	0.6	0.7037	0.75	0.8077
	Weighted avg	**0.8318**	**0.8064**	**0.8538**	**0.8748**
Recall	akiec	0.6615	0.6615	0.6462	0.6615
	bcc	0.7864	0.8252	0.8252	0.8252
	bkl	0.6727	0.7773	0.7182	0.7773
	df	0.4348	0.6957	0.4348	0.6957
	nv	0.9545	0.9612	0.9672	0.9612
	mel	0.4574	0.6099	0.4888	0.6099
	vasc	0.6429	0.75	0.75	0.75
	Weighted avg	**0.8397**	**0.8168**	**0.8597**	**0.8792**
F1	akiec	0.6565	0.688	0.6885	0.688
	bcc	0.7606	0.8293	0.8057	0.8293
	bkl	0.6868	0.788	0.7633	0.788
	df	0.5556	0.7442	0.6061	0.7442
	nv	0.9242	0.9378	0.9258	0.9378
	mel	0.5559	0.6869	0.5924	0.6869
	vasc	0.6207	0.7778	0.75	0.7778
	Weighted avg	**0.8315**	**0.8023**	**0.8508**	**0.8753**

to achieve dimensionality consistency with Xception. The zero-padding operation is shown in Figure 4D. The fusion process is shown in Figure 4B.

Classifier-level fusion is performed based on the classification layer. This paper first fuses the last convolution layer of each of the three different networks with the Convolutional Block Attention Module (CBAM), then performs global average pooling on this basis, splices a fully connected layer to obtain the final feature vector, and performs a simple splicing operation on the three feature vectors. Finally, the splicing result is input into the classification layer to output the final predicted category value, as shown in Figure 1 in the model building stage. In this way, the network outputs four values corresponding to Inception V3,

10.3389/fpubh.2022.1034772

TABLE 4 The influence of a single network model on evaluation metrics.

Single algorithm		Inception V3	InceptionResNet	Xception_No_CBAM	Xception
Acc	akiec	0.7538	0.6769	0.6923	0.7077
	bcc	0.8932	0.9223	0.8544	0.8835
	bkl	0.8045	0.7955	0.8091	0.8182
	df	0.7391	0.6522	0.8696	0.7391
	nv	**0.9679**	**0.9791**	**0.9754**	**0.9724**
	mel	0.6771	0.5964	0.6233	0.6682
	vasc	0.8928	0.8929	0.8929	0.8929
	Weighted avg	**0.9031**	**0.8987**	**0.9002**	**0.9046**
F1	akiec	0.7597	0.7273	0.7563	0.7541
	bcc	0.8762	0.9223	0.8756	0.8545
	bkl	0.8290	0.8140	0.8109	0.8353
	df	0.8293	0.7317	0.8696	0.8095
	nv	**0.9495**	**0.9460**	**0.9482**	**0.9525**
	mel	0.7438	0.7056	0.7221	0.7358
	vasc	0.8475	0.9091	0.8772	0.9091
	Weighted avg	**0.9007**	**0.8934**	**0.8961**	**0.9018**
Specificity	akiec	0.9923	0.9938	0.9954	0.9943
	bcc	0.9921	0.9958	0.9947	0.9900
	bkl	0.9832	0.9804	0.977	0.9826
	df	0.9995	0.9985	0.9985	0.9990
	nv	**0.8565**	**0.8157**	**0.8338**	**0.8595**
	mel	0.9820	0.9882	0.9871	0.9815
	vasc	0.9970	0.9990	0.9978	0.9990
	Weighted avg	**0.8994**	**0.8727**	**0.8843**	**0.9012**
AUC	akiec	0.9832	0.9805	0.9835	0.9843
	bcc	0.9938	0.9976	0.9963	0.9954
	bkl	0.9787	0.9769	0.9835	0.9813
	df	0.9910	0.9971	0.9928	0.9959
	nv	0.9802	0.9800	0.9806	0.9775
	mel	0.9651	0.9643	0.9599	0.9613
	vasc	0.9772	0.9930	0.9988	0.9870
	Weighted avg	**0.9792**	**0.9792**	**0.9799**	**0.9776**

InceptionResNet V2, Xception, and a merged output. The loss value of the network is the sum of the loss values of the four parts, but the final output is the overall output of the network.

In Figures 1, 4, "CBAM" is an attention mechanism proposed by Woo (43) in 2018. Woo applied attention to both the channel and spatial dimensions. Similar to the SENet[10], a CBAM can be embedded in most mainstream networks at present. The feature extraction capability of a network model can be improved without significantly increasing its computational complexity and number of parameters. Therefore, this paper embeds a CBAM into the feature extraction part to improve the feature extraction ability of the model and facilitate the subsequent network classification ability improvement.

Transfer learning transfers knowledge learned from a source dataset to a target dataset. Fine-tuning is a common technique for transfer learning. The target model replicates all the model designs and their parameters on the source model except the output layer, and fine-tunes these parameters based on the target dataset. The output layer of the target model, on the other hand, needs to be trained from scratch. The whole process of model building and prediction is shown in Figure 4C. First, all the parameters of the base classifier are "frozen" to prevent large planned changes in these parameters during the initial network training. Subsequently all parameters of the network model are "unfrozen" and the parameters of the entire network are fine-tuned to achieve classification of skin diseases.

TABLE 5 The influence of different fusion strategies on evaluation metrics.

Fusion network		Concat_Conv2D	Concat_Zeropadding	Concat_Dense
Acc	akiec	0.7538	0.6923	0.8154
	bcc	0.8058	0.8641	0.9417
	bkl	0.7955	0.8500	0.8409
	df	0.4348	0.6957	0.8261
	nv	0.9418	0.9612	0.9828
	mel	0.6682	0.6323	0.6502
	vasc	0.8929	0.8571	0.9286
	Weighted avg	**0.8757**	**0.8942**	**0.9201**
F1	akiec	0.7424	0.7258	0.7737
	bcc	0.8342	0.8812	0.9372
	bkl	0.7743	0.8184	0.8768
	df	0.5882	0.7619	0.8837
	nv	0.9383	0.9467	0.9572
	mel	0.6882	0.7050	0.7532
	vasc	0.7812	0.8276	0.8966
	Weighted avg	**0.8745**	**0.8914**	**0.9170**
Specificity	akiec	0.9907	0.9928	0.9902
	bcc	0.9932	0.9947	0.9963
	bkl	0.9680	0.9720	0.9905
	df	0.9995	0.9985	0.9995
	nv	0.8671	0.8595	0.8565
	mel	0.9657	0.9798	0.9904
	vasc	0.9944	0.9970	0.9980
	Weighted avg	**0.9029**	**0.9001**	**0.9013**
AUC	akiec	0.9706	0.9839	0.9912
	bcc	0.9801	0.9955	0.9984
	bkl	0.9651	0.9754	0.9899
	df	0.9589	0.9821	0.9923
	nv	0.9653	0.9741	0.9853
	mel	0.9227	0.9514	0.9722
	vasc	0.9728	0.9890	0.9765
	Weighted avg	**0.9615**	**0.9734**	**0.9852**

4. Experiment

4.1. Experimental conditions

The experimental environment includes Linux X86_64, an Nvidia Tesla V100, and 16 GB of memory. This experiment is based on Python version 3.7.9, TensorFlow version 2.3.0, and Keras version 2.4.3.

4.2. Evaluation criteria

In this study, the accuracy, recall, specificity, precision, F1, weighted AUC and AUC metrics are used to evaluate pigmented skin disease detection methods based on a fusion network. The model evaluation confusion matrix and calculation equations are shown in Table 2, respectively.

True Negatives (TNs) represent the number of cases for which the real values are negative and the model thinks they are negative.

False Positives (FPs) represent the number of cases for which the real values are negative and the model thinks they are positive.

False Negatives (FNs) represent the number of cases for which the real values are positive and the model thinks they are negative.

True Positives (TPs) represent the number of cases for which the real values are positive and the model thinks they are positive.

TABLE 6 The influence of fusion of two base classifiers on evaluation metrics.

Two network		Inception V3_Inception V3	InceptionRes Net_Inception-ResNet	Xception_ Xception	Inception V3_Inception-ResNet	Inception V3_Xception	InceptionRes-Net_Xception
Acc	akiec	0.2462	0.4923	0.5231	0.7692	0.7231	0.7538
	bcc	0.7282	0.7087	0.7864	0.8932	0.9223	0.9417
	bkl	0.5682	0.6364	0.6318	0.8500	0.8455	0.8409
	df	0.4783	0.5217	0.3913	0.7391	0.8696	0.7391
	nv	0.8069	0.8673	0.9314	0.9851	0.9761	0.9754
	mel	0.4439	0.4350	0.4036	0.6143	0.6726	0.6054
	vasc	0.6786	0.8214	0.8929	0.8929	0.9286	0.9286
	Weighted avg	**0.7124**	**0.7688**	**0.8123**	**0.9131**	**0.9151**	**0.9071**
F1	akiec	0.3596	0.5289	0.5574	0.7937	0.7520	0.7967
	bcc	0.6198	0.6759	0.7364	0.8846	0.8962	0.9194
	bkl	0.4505	0.5501	0.6347	0.8539	0.8493	0.8565
	df	0.4889	0.5581	0.5143	0.8095	0.9091	0.8293
	nv	0.8547	0.8851	0.9021	0.9555	0.9583	0.9471
	mel	0.4033	0.4491	0.4932	0.7366	0.7557	0.7124
	vasc	0.7308	0.8070	0.8772	0.9091	0.8966	0.8966
	Weighted avg	**0.7259**	**0.7726**	**0.8027**	**0.9088**	**0.9124**	**0.9027**
Specificity	akiec	0.9959	0.9943	0.9881	0.9943	0.9933	0.9954
	bcc	0.9663	0.9932	0.9811	0.9932	0.9926	0.9942
	bkl	0.8822	0.9826	0.9557	0.9826	0.9821	0.9849
	df	0.9944	0.9990	0.9985	0.9990	0.9995	0.9995
	nv	0.8353	0.8444	0.7296	0.8444	0.8761	0.8293
	mel	0.9051	0.9933	0.9708	0.9933	0.9865	0.9882
	vasc	0.9975	0.9990	0.9980	0.9990	0.9980	0.9980
	Weighted avg	**0.8643**	**0.8926**	**0.8094**	**0.8926**	**0.9130**	**0.8823**
AUC	akiec	0.946	0.9509	0.9427	0.9793	0.9807	0.9909
	bcc	0.9585	0.9772	0.9855	0.9976	0.9972	0.9977
	bkl	0.8584	0.9	0.9304	0.9849	0.9798	0.9836
	df	0.9315	0.9436	0.974	0.9975	0.9966	0.9937
	nv	0.9147	0.9239	0.9396	0.9819	0.9837	0.9818
	mel	0.8499	0.8645	0.9046	0.9682	0.9708	0.9633
	vasc	0.9920	0.9959	0.9963	0.9823	0.9921	0.9868
	Weighted avg	**0.9058**	**0.9195**	**0.9383**	**0.9816**	**0.9827**	**0.9813**

4.3. Determination of the experimental parameters

4.3.1. Test results of a single classifier

In this paper, Inception V3 and cbam fusion are used to test three data augmentation methods. The first (column 4 of Table 3) class weights are calculated by adjusting the model to include a penalty for prediction error for classes with smaller sample sizes, and the weight parameters for each class are calculated as follows.

$$Weight = \frac{n\_samples}{n\_classes * bincount(y)} \tag{9}$$

Where *n_samples* represents the total number of picture samples,*n_classes* represents the number of categories, and *bincount*(*y*) represents the sample size of each category in the training set. Weight is the weight corresponding to each category. The lower the sample size of the category, the higher its weight.

The second uses image flipping (column 5 of Table 3) to flip the category with a small sample size to flip the image left and right, invert it up and down, and flip it systematically so that the imbalance between its various categories is somewhat mitigated.

For the network model, a change in a pixel value of an image represents that this image will then change. Therefore, the third one (column 6 of Table 3) is based on the second one to achieve

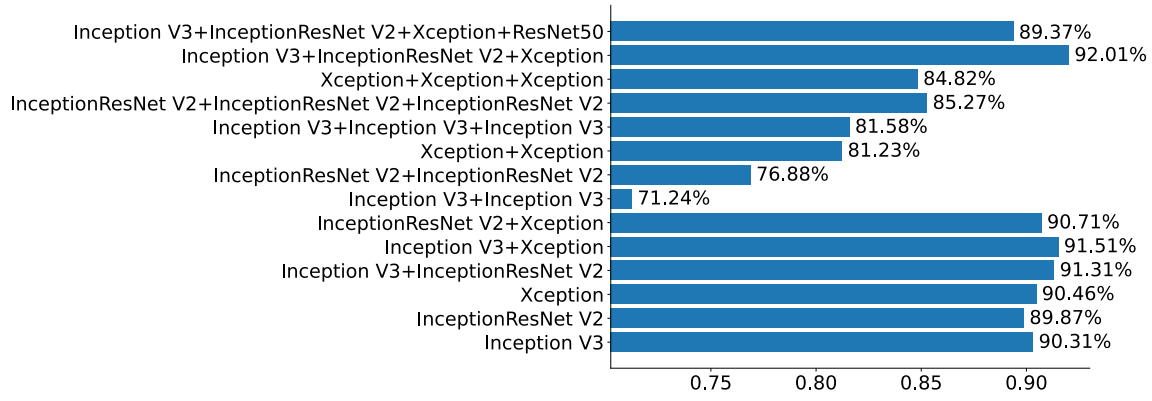


FIGURE 5
Effect of the number of classifiers on the resulting network.

a complete balance between its various categories. The interval of increasing and decreasing pixel values is first calculated by the equation, and then a random value is randomly drawn from the interval without put-back as the increasing or decreasing pixel value.

$$Pixel = \frac{differences}{2 * n\_classes} \tag{10}$$

Where *n_classes* represents the number of categories, and *differences* represents the difference between this category and the category "nv." Therefore, the interval of image increase and decrease is from 1 to Pixel.

From Table 3, it can be seen that the effect of solving the data imbalance by changing the calculation method of the model loss values decreases the correct prediction rate compared to the dataset without any change, mainly because the change of the loss values causes the model to have some bias between the categories during training. By changing the image flip compared to not making any changes, the imbalance between categories is somewhat alleviated, so the prediction accuracy is somewhat improved, but there is still some imbalance between categories. Based on the image flip, each image is randomly added or subtracted a certain pixel value to get a brand new image, thus achieving a balance between each category of the image and a certain improvement in prediction.

Therefore, in this paper, we use the image style transfer upsampling scheme to equalize the dataset. After completing dataset equalization, in the single-classifier experiment, we successively change the model module in Figure 4C into three algorithm models: "Inception V3+CBAM," "InceptionResNet V2+CBAM," "Xception," and "Xception+CBAM." The algorithm test results are shown in Table 4. It can be seen from the third to the sixth column of Table 3 and the third column of Table 4 that the effects of the original dataset, image preprocessing, pixel change and image style transfer on the detection of pigmented skin lesions based on Inception V3 are improved in order, and the accuracy of image style transfer regarding the detection of pigmented skin lesions is 4% higher than that of image preprocessing. It is proven that image style transfer is effective for the detection of pigmented skin lesions. From column 5 and column 6 of Table 4, it can be seen that the presence or absence of the attention mechanism makes some difference to the classification effect (Acc, F1, Specificity), thus proving the contribution of the attention mechanism in the classification of pigmented skin diseases. However, it can be seen from the Acc and F1 values in the table that the detection rate of the "nv" category is much higher than that of the other categories, indicating that a single model has certain anti-interference ability limitations with respect to the images generated by the algorithm.

4.3.2. Fusion test results of multiple classifiers

The detection effect of multinetwork fusion can generally strengthen the generalization ability of a model, thereby improving its detection ability. After performing dataset equalization, we first compare different fusion methods in terms of their final classification effects in multiple classifier experiments, and we test the feature-level fusion approach based on deep features and the classifier-level fusion method based on the classification layer. All three fusion strategies use Inception V3, InceptionResNet V2, and Xception as the three base classifiers. The first feature-level fusion method based on deep features reduces the dimensionality of a feature graph with a larger output through the convolution layer to realize the splicing of dimensions. The second feature-level fusion method based on deep features adds feature graphs with smaller output dimensions to larger feature graphs with the zero-padding operation. The third classifier-level fusion method based on the classification layer splices the outputs of the fully connected layers of the three base classifiers.

TABLE 7 The influence of fusion of multiple base classifiers on evaluation metrics.

Multi-network fusion		Inception V3_Inception V3_Inception V3	InceptionResNet_ InceptionResNet_ InceptionResNet	Xception_ Xception_ Xception	Inception V3_ InceptionResNet_ Xception_ ResNet50
Acc	akiec	0.6308	0.7385	0.7077	0.7846
	bcc	0.7961	0.7573	0.8058	0.8738
	bkl	0.7045	0.7318	0.7227	0.8455
	df	0.4783	0.4783	0.4348	0.7391
	nv	0.9150	0.9493	0.9493	0.9679
	mel	0.4170	0.5112	0.4709	0.5561
	vasc	0.8929	0.8214	0.8214	0.8571
	Weighted avg	**0.8158**	**0.8527**	**0.8482**	**0.8937**
F1	akiec	0.5857	0.7164	0.6765	0.7286
	bcc	0.7421	0.7464	0.8098	0.8738
	bkl	0.6610	0.7523	0.7413	0.8176
	df	0.6111	0.6111	0.5882	0.8500
	nv	0.9065	0.9228	0.9158	0.9492
	mel	0.4987	0.5891	0.5707	0.6667
	vasc	0.8333	0.8679	0.8519	0.8276
	Weighted avg	**0.8109**	**0.8468**	**0.8403**	**0.8894**
Specificity	akiec	0.9825	0.9892	0.9871	0.8860
	bcc	0.9811	0.9853	0.9900	0.9330
	bkl	0.9473	0.9736	0.9720	0.9090
	df	0.9990	0.9990	0.9995	0.8700
	nv	0.7900	0.7810	0.7492	0.9110
	mel	0.9680	0.9719	0.9775	0.7710
	vasc	0.9965	0.9990	0.9985	0.9270
	Weighted avg	**0.8485**	**0.8462**	**0.8256**	**0.8953**
AUC	akiec	0.9511	0.9769	0.9803	0.9866
	bcc	0.9820	0.9895	0.9900	0.9964
	bkl	0.9309	0.9650	0.9611	0.9726
	df	0.9505	0.9656	0.9826	0.9821
	nv	0.9413	0.9611	0.9495	0.9754
	mel	0.9070	0.9426	0.9082	0.9591
	vasc	0.9838	0.9993	0.9894	0.9971
	Weighted avg	**0.9394**	**0.9620**	**0.9502**	**0.9751**

Three kinds of fusion strategy evaluation indices are shown in Table 5. According to the data supplied by the convolution layer, the first one-dimensional characteristic figure of dimensionality reduction is generally low. The main reason for this is that adding a convolution layer results in many parameters that need to be trained. The first network loss value is large and can lead to difficult network training for reaching a more appropriate stage. As a result, the overall parameters of the network cannot achieve good results. If zero padding is used, the small-dimensional feature graph is extended, and no redundant parameter training requirement is imposed. Therefore, the output result will be consistent with the transfer learning result. The third method is to splice the output of the fully connected layer, and the final prediction index is the best option. First, the feature extraction part of the network contains the network parameters trained by ImageNet, and the features are relatively appropriate. Finally, only the parameters of the fully connected layer are added; thus, the feature extraction process of the network model does not change, and the final prediction effect is also the best.

From the weighted average of the Acc and F1 values in Tables 4, 5, it can be seen that the model training and

TABLE 8 Comparison of the results obtained in this study with those in the literature.

References	Method	Results
Sevli (32)	Custom CNN model	The accuracy on test set reaches 91.51%
Salian et al. (44)	Custom CNN model	The test accuracy is 83.15%
Pal et al. (33)	Ensemble (ResNet50, DenseNet-121, and MobileNet)	The normalized multiclass accuracy is 77.5%
Xie et al. (8)	multilevel deep ensemble (MLDE) model	The result is an average AUC of 86.5
Aldwgeri and Abubacker (35)	Ensemble[VGG, ResNet50, Inception-V3, Xception, and DenseNet-121]	Multiclass accuracy of 80.1% and mean average of 0.89 AUROC
Hard voting	Ensemble (Inception V3, InceptionResNet V2, and Xception)	The accuracy on test set reaches 91.61%
Proposed fusion network	Fusion network (Inception V3, InceptionResNet V2, and Xception)	The accuracy and AUC on the test set reach 92.01 and 95.3%, respectively

TABLE 9 Comparison of different methods on external datasets.

Method	Acc	Specificity	AUC
Kermany et al. (46)	0.934	0.94	0.988
Kaymak and Serener (47)	0.971	0.984	Not mentioned
Concat_Cov2D	**0.974**	**0.991**	**0.983**
Concat_Zeropadding	**0.975**	**0.992**	**0.983**
Concat_Dense	**0.987**	**0.996**	**0.991**

prediction steps performed by a single classifier are better than those of the two fusion strategies based on feature-level fusion. The main reason for this involves the changes in the extracted image features during feature-level fusion. Compared with the better network feature extraction ability of "ImageNet" training, the feature extraction ability of the modified network exhibits a certain decline, resulting in a decrease in the classification index based on feature-level fusion. During feature extraction, the classifier-based fusion scheme does not change the feature extraction capability of the original network based on "ImageNet." Features are learned separately through the convolution layer of each base classifier, and the results of the fully connected network (i.e., the classifier) of the base classifier are fused to obtain the final predicted category value. Based on classifier-level fusion, the output results of multiple base classifiers are fused. The generalization ability and anti-interference ability of the network are enhanced, and the model classification ability is enhanced.

4.3.3. Setting the number of fusion networks

This section mainly studies how to combine base classifiers in fusion networks to achieve the best effect for the detection of pigmented skin lesions. This paper mainly tests the effectiveness of combinations including three basic classifiers: Inception V3, InceptionResNet V2, and Xception. The fusion effects of two networks, three networks, four networks, etc. are tested. The best fusion scheme (classifier-level fusion based on the classification layer in Section 4.3.2) is adopted. Six scenarios are available regarding the fusion of two networks, as shown in the table: fusing Inception V3 with Inception V3, InceptionResNet V2 with InceptionResNet V2, Xception with Xception, Inception V3 with InceptionResNet V2, Inception V3 with Xception, and InceptionResNet V2 with Xception. Four scenarios are considered regarding the fusion of three networks, as shown in the table: the fusion of Inception V3, Inception V3, and Inception V3; the fusion of InceptionResNet V2, InceptionResNet V2, and InceptionResNet V2; the fusion of Xception, Xception, and Xception; and the fusion of Inception-V3, Inception-ResNet-V2, and Xception. The four-network case is a fusion of Inception V3, InceptionResNet V2, Xception, and ResNet50. It can be seen from Table 6 and Figure 5 that if two base classifiers are consistent in the fusion process of two networks, the classification effect will be worse than that of using one base classifier alone. In a fusion network, there must be some difference between the base classifiers; otherwise, the network easily falls into local minima during the training process. It can be seen from Table 6 that when two different base classifiers are used, the classification accuracy is greatly improved compared with that of a network containing two identical classifiers. From the values listed in Table 6, the monitoring indices of Inception V3_InceptionResNet, Inception V3_Xception, and Inception V3_InceptionResNet are better than those of single Inception V3, InceptionResNet, Xception models; It can be seen from the data in Table 7 that the fusion effect of four networks is not as good as that of three networks, thus proving that the network fusion does not guarantee that a greater number of base classifiers leads to better results. Therefore, the fusion method based on Inception V3, InceptionResNet V2, and Xception is finally selected as the network model in this paper.

To explore the performance of different network combinations in the feature extraction framework, we perform ablation experiments for each image classification configuration. The first case utilizes combinations with the same subnetwork.

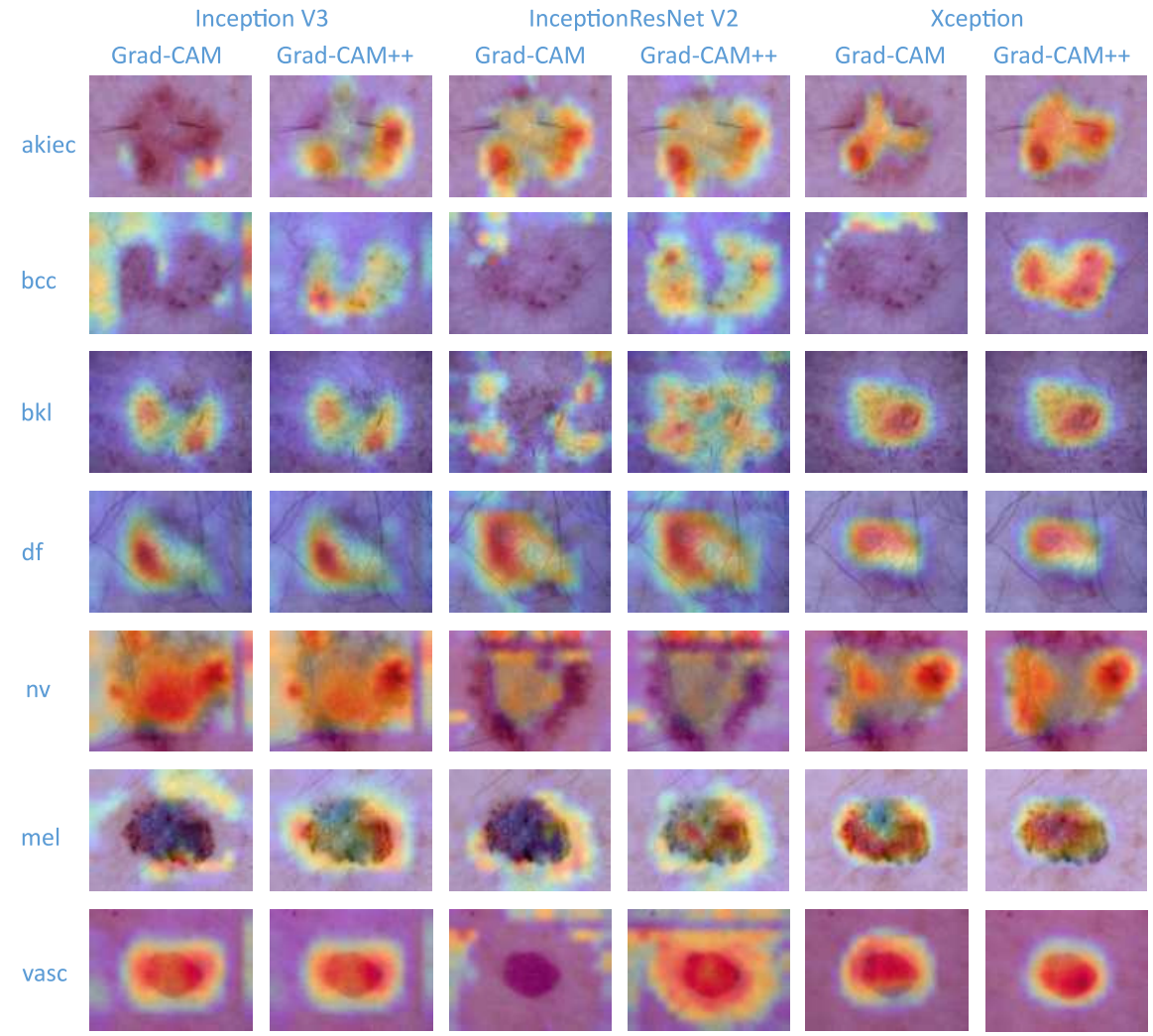


FIGURE 6
Model visualization.

With the increase in the number of networks (columns 3, 4, 6 in Table 4, 3–5 in Table 6, and 3–5 in Table 7), the classification performance declines. Therefore, it is not better to increase the number of subnetwork when they are the same. The possible reason for this finding is that overfitting easily occurs in overly complex networks, which leads to performance degradation. However, the classification performance shown in Table 7 is higher than that in Table 6. The main reason for this is that in ensemble learning, the number of general base classifiers cannot appear to be even; otherwise, the same predicted value is likely to occur, and random judgment may occur during model classification. The second was for different subnetworks. With the increase in the number of networks (columns 3, 4, 6 in Table 4, columns 6–8 in Table 6, and columns 6 in Table 7), the classification performance increases first and then decreases, indicating that increasing the number of subnetworks can improve the accuracy of pigmented skin lesion detection, but more is not always better. The overfitting of complex networks may also occur. Third, it can be seen from Table 6 that when the number of networks is the same, the performance obtained when using different subnetworks as feature extractors is better than that achieved with identical subnetworks. These results prove the feasibility of the proposed network.

4.4. Comparison of the experimental results obtained by the proposed methods

According to the test results, the comparison between this study and similar recent studies is shown in Table 8. The dataset listed in Table 8 is HAM10000, which was presented in the ISIC 2018 Challenge and is used in this study. From the evaluation indices obtained on the test set, it can be seen that the data

upsampling scheme based on image flipping and image style transfer proposed in this paper can produce the same amount of data in each category; In addition, network fusion schemes based on available data can achieve higher detection efficiency for pigmented skin lesions than hard voting fusion schemes.

4.5. Experimental expansion

In order to validate the impact of the developed fusion network on external test data, the UCSD common retinal OCT dataset (45) was collected with a total sample size of 108,309 images in four categories: Normal, Drusen, CNV, and DME. The sample sizes of the four categories are 51,140, 8,616, 37,205, and 11,348, respectively, and this paper focuses on the "limited model," i.e., 1,000 randomly selected images in each category, to compare the performance using the fusion strategies. Table 9 shows that the overall accuracies of the three fusion strategies are 97.4, 97.5, and 98.7%, respectively. Compared with the model proposed by Kermany (46), the accuracy is 93.4%, which is an average improvement of 4% points. Overall, the three fusion strategies proposed in this paper are effective.

4.6. Model interpretability

To verify the interpretable and explainable of the classifier-level fusion network based on the classification layer proposed in this paper, the visualization effect of the sample with the highest prediction probability for each category among the test set samples is shown in Figure 6. In this paper, Grad_CAM (48) and Grad_CAM++ (49) are used as visualization algorithms, and the prediction probability value of the final output category of the test model is used to visualize the fusion of the three base classifiers and the CBAM. To compare the visualization effects of the Grad_CAM and Grad_CAM++ visualization algorithms on the results of this paper and to determine the visualization effect of the final predicted probability value of the model in this paper for the fusion of each base classifier and the attention mechanism, each row in Figure 6 shows that the pictures are all derived from the same sample image. It can be seen from the results that the visualization effects of Grad_CAM++ on the three base classifiers are better than those of Grad_CAM. Grad_CAM++ can display the lesion areas of pigmented skin lesions in a good thermal map. After the image is checked by professional clinicians, the visual part of the image can show that the locations focused on by the model are similar to those yielded by human experience. The visualization effect of Xception shows that the localization area is small and that all results are contained in the lesion area, which is superior to the effects of the other two classifiers (Inception V3 and InceptionResNet V2), thus proving the more interpretable and explainable of the proposed algorithm.

5. Conclusion

A fusion network-based detection algorithm for pigmented skin lesions is proposed in this paper. Image preprocessing and image augmentation are carried out before inputting the given dataset into the network, which can solve the problem of low classification accuracy caused by the unbalanced distribution of the original data to a large extent. In this paper, various fusion strategies are used to verify the applicability of the algorithm for pigmented skin lesions. Based on a network performance comparison, we empirically find that the classification effects of the two fusion strategies based on feature-level fusion are not good according to their pigmented skin lesion results. However, the proposed fusion scheme can be applied in other application scenarios and can provide experience guidance for the corresponding model design process. Second, our algorithmic architecture (containing three fusion strategies) only covers single-modal, categorization-oriented methods. However, we also note that multimodal input data are present in medical image analyses, and the corresponding fusion schemes can be studied by extending the current framework (50–52). At the same time, two visualization algorithms are used to apply the color visualization method to make the proposed deep learning model more interpretable and explainable, and the accuracy of the developed algorithm was confirmed by comparing the results with those of related papers. In the future, we plan to test the robustness of the proposed algorithm using a hospital database of actual high definition images of pigmented skin diseases, deploy the algorithm model on servers for physicians in remote areas to diagnose pigmented skin diseases, and apply the three fusion strategies to other more medical application scenarios to validate the advantages of the algorithm.

Data availability statement

The datasets generated and analysed during the current study are available from the corresponding author upon reasonable request. All deep learning methods are implemented by using TensorFlow (https://tensorflow.google.cn/). The custom script for this study will be available at https://github.com/YHHAZ/NetworkFusion. Correspondence and requests for data materials should be addressed to LC (chlq35@126.com).

Author contributions

LW and ZA: conceptualization and writing—original draft preparation. LW, ZA, and YL: methodology. ZA, JC, QJ, HC, and QL: writing—review and editing. YL and LC: project administration. JC and QJ: data collection. LW, LC, and YL: funding acquisition. All authors read and agreed to the published version of the manuscript.

Funding

This work was supported in part by a Wuhan Medical Scientific Research Project grant to LC (WX20B25), in part by a Science and Technology Planning Project of Wuhan grant to LC (2019010701011418), in part by a Research Innovation Fund Project of Jianghan University grant to LW (211051003), and in part by Sinopharm Genomics Technology Co., Ltd. The funders were not involved with the study design; the collection, analysis, or interpretation of data; the writing of this article; or the decision to submit it for publication.

Acknowledgments

The numerical calculations in this paper were performed on the supercomputing system in the Supercomputing Center of Wuhan University.

Conflict of interest

ZA, QL, and YL are employees of Sinopharm Genomics Technology Co., Ltd.

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

References

1. National Cancer Institute, Surveillance, Epidemiology, Program ER. *Cancer Stat Facts: Melanoma of the Skin* (2021). Available online at: https://seer.cancer.gov/statfacts/html/melan.html

2. Freedberg KA, Geller AC, Miller DR, Lew RA, Koh HK. Screening for malignant melanoma: a cost-effectiveness analysis. *J Am Acad Dermatol.* (1999) 41:738–45.

3. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJ, Bottou L, Weinberger KQ, editors. *Advances in Neural Information Processing Systems*. Vol. 25. Curran Associates, Inc. (2017).

4. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *European Conference on Computer Vision*. Part 1. Zurich: Springer International Publishing (2014). p. 818–33.

5. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:14091556* (2014). doi: 10.48550/arXiv.1409.1556

6. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA (2015). p. 1–9.

7. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV (2016). p. 770–8.

8. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* Honolulu, HI (2017). p. 1492–500.

9. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT (2020). doi: 10.1109/CVPR.2018.00745

10. Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* Salt Lake City, UT (2018). p. 8697–710.

11. Tan M, Le QV. *EfficientNet: Rethinking model scaling for convolutional neural networks*. In: Chaudhuri K, Salakhutdinov R, editors. *Proceedings of the 36th International Conference on Machine Learning (ICML)* Long Beach, CA (2019). p. 6105–14.

12. Milton MAA. Automated skin lesion classification using Ensemble of deep neural networks in ISIC 2018: skin lesion analysis towards melanoma detection challenge. *arXiv preprint arXiv:190110802* (2019). doi: 10.48550/arXiv.1901.10802

13. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: results of 10 convolutional neural networks. *Comput Biol Med.* (2020) 121:103795. doi: 10.1016/j.compbiomed.2020.103795

14. Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub ZB, et al. Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access.* (2020) 8:132665–76. doi: 10.1109/ACCESS.2020.3010287

15. Narin A, Kaya C, Pamuk Z. Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. *Pattern Anal Appl.* (2021) 24:1207–20. doi: 10.48550/arXiv.2003.10849

16. Qjidaa M, Mechbal Y, Ben-Fares A, Amakdouf H, Maaroufi M, Alami B, et al. Early detection of COVID19 by deep learning transfer Model for populations in isolated rural areas. In: *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)* Fez (2020). p. 1–5.

17. Tammina S. CovidSORT: detection of novel COVID-19 in chest X-ray images by leveraging deep transfer learning models. In: Kumar A, Senatore S, Gunjan VK, editors. *Proceedings of the 2nd International Conference on Data Science, Machine Learning and Applications (ICDSMLA). Lecture Notes in Electrical Engineering.* Vol. 783. Pune: Springer (2022). p. 431–47.

18. Ai Z, Huang X, Fan Y, Feng J, Zeng F, Lu Y. DR-IIXRN : detection algorithm of diabetic retinopathy based on deep ensemble learning and attention mechanism. *Front Neuroinformatics.* (2021) 15:778552. doi: 10.3389/fninf.2021.778552

19. Sabri MA, Filali Y, El Khoukhi H, Aarab A. Skin cancer diagnosis using an improved ensemble machine learning model. In: 2020 *International Conference on Intelligent Systems and Computer Vision (ISCV)*. Fez (2020). p. 1–5.

20. Sae-Lim W, Wettayaprasit W, Aiyarak P. Convolutional neural networks using mobileNet for skin lesion classification. In: *2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE): Knowledge Evolution Towards Singularity of Man-Machine Intelligence* Chonburi (2019). p. 242–7.

21. Mohamed EH, El-Behaidy WH. Enhanced skin lesions classification using deep convolutional networks. In: *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)* Cario (2019). p. 180–8.

22. Hegde PR, Shenoy MM, Shekar BH. Comparison of machine learning algorithms for skin disease classification using color and texture features. In: *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* Bangalore (2018). p. 1825–8.

23. Hameed N, Shabut A, Hossain MA. A Computer-Aided diagnosis system for classifying prominent skin lesions using machine learning. In: *2018 10th Computer Science and Electronic Engineering (CEEC)* Colchester (2019). p. 186–91.

24. Barata C, Ruela M, Francisco M, Mendonça T, Marques JS. Two systems for the detection of melanomas in dermoscopy images using texture and color features. *IEEE Syst J*. (2013) 8:965–79. doi: 10.1109/JSYST.2013.2271540

25. Seidenari S, Pellacani G, Grana C. Pigment distribution in melanocytic lesion images: a digital parameter to be employed for computer-aided diagnosis. *Skin Res Technol*. (2005) 11:236–41. doi: 10.1111/j.0909-725X.2005.00123.x

26. Varalakshmi P, Aruna Devi V, Ezhilarasi M, Sandhiya N. Enhanced dermatoscopic skin lesion classification using machine learning techniques. In: *2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* Chennai (2021). p. 68–71.

27. Ma V, Karki MV. Skin cancer detection using machine learning techniques. In: *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)* Bangalore (2020). p. 1–5.

28. Kalwa U, Legner C, Kong T, Pandey S. Skin cancer diagnostics with an all-inclusive smartphone application. *Symmetry*. (2019) 11:790. doi: 10.3390/sym11060790

29. Gupta S, Panwar A, Mishra K. Skin disease classification using dermoscopy images through deep feature learning models and machine learning classifiers. In: *IEEE EUROCON 2021 - 19th International Conference on Smart Technologies* Lviv (2021). p. 170–4.

30. Rahman Z, Ami AM. A transfer learning based approach for skin lesion classification from imbalanced data. In: *2020 11th International Conference on Electrical and Computer Engineering (ICECE)* Dhaka (2020). p. 65–8.

31. Chaturvedi SS, Gupta K, Prasad PS. Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using mobilenet. In: Hassanien AE, Bhatnagar R, Darwish A, editors. *Advanced Machine Learning Technologies and Applications*. Vol. 1141. Cairo: Springer (2021). p. 165–76.

32. Sevli O. A deep convolutional neural network-based pigmented skin lesion classification application and experts evaluation. *Neural Comput Appl*. (2021) 33:12039–50. doi: 10.1007/s00521-021-05929-4

33. Pal A, Ray S, Garain U. Skin disease identification from dermoscopy images using deep convolutional neural network. *arXiv preprint arXiv:180709163* (2018). doi: 10.48550/arXiv.1807.09163

34. Xie Y, Zhang J, Xia Y. A multi-level deep ensemble model for skin lesion classification in dermoscopy images. *arXiv preprint arXiv:180708488* (2018). doi: 10.48550/arXiv.1807.08488

35. Aldwgeri A, Abubacker NF. Ensemble of deep convolutional neural network for skin lesion classification in dermoscopy images. In: Badioze Zaman H, Smeaton AF, Shih TK, Velastin S, Terutoshi T, Mohamad Ali N, et al. editors. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 11870. Bangi (2019). p. 214–26.

36. Tschandl P, Rosendahl C, Hararld K. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data*. (2018) 5:1–9. doi: 10.1038/sdata.2018.161

37. Ghiasi G, Lee H, Kudlur M, Dumoulin V, Shlens J. Exploring the structure of a real-time, arbitrary neural artistic stylization network. *arXiv preprint arXiv:170506830* (2017). doi: 10.48550/arXiv.1705.06830

38. Wang C, Peng G, De Baets B. Deep feature fusion through adaptive discriminative metric learning for scene recognition. *Inform Fusion*. (2020) 63:1–12. doi: 10.1016/j.inffus.2020.05.005

39. Chaib S, Liu H, Gu Y, Yao H. Deep feature fusion for VHR remote sensing scene classification. *IEEE Trans Geosci Remote Sens*. (2017) 55:4775–84. doi: 10.1109/TGRS.2017.2700322

40. Song W, Li S, Fang L, Lu T. Hyperspectral image classification with deep feature fusion network. *IEEE Trans Geosci Remote Sens*. (2018) 56:3173–84. doi: 10.1109/TGRS.2018.2794326

41. Bian X, Chen C, Sheng Y, Xu Y, Du Q. Fusing two convolutional neural networks for high-resolution scene classification. In: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Fort Worth, TX (2017). p. 3242–5.

42. Xue W, Dai X, Liu L. Remote sensing scene classification based on multi-structure deep features fusion. *IEEE Access*. (2020) 8:28746–55. doi: 10.1109/ACCESS.2020.2968771

43. Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Munich Vol. 11211 (2018). p. 3–19.

44. Salian AC, Vaze S, Singh P, Shaikh GN, Chapaneri S, Jayaswal D. Skin lesion classification using deep learning architectures. In: *2020 3rd International Conference on Communication Systems, Computing and IT Applications, CSCITA 2020 - Proceedings* Mumbai (2020).

45. Kermany D, Zhang K, Goldbaum M. Large dataset of labeled optical coherence tomography (oct) and chest x-ray images. *Mendeley Data*. (2018) 3:10–17632. doi: 10.17632/rscbjbr9sj.3

46. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. (2018) 172:1122–31. doi: 10.1016/j.cell.2018.02.010

47. Kaymak S, Serener A. Automated age-related macular degeneration and diabetic macular edema detection on oct images using deep learning. In: *2018 IEEE 14th international conference on intelligent computer communication and processing (ICCP)*. Cluj-napoca (2018). p. 265–9.

48. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks *via* gradient-based localization. In: *Proceedings of the IEEE International Conference on Computer Vision*. Venice (2017). p. 618–26.

49. Chattopadhay A, Sarkar A, Howlader P, Balasubramanian VN. Grad-CAM++: generalized gradient-based visual explanations for deep convolutional networks. In: *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Lake Tahoe, NV (2018). p. 839–47.

50. Qian X, Pei J, Zheng H, Xie X, Yan L, Zhang H, et al. Prospective assessment of breast cancer risk from multimodal multiview ultrasound images *via* clinically applicable deep learning. *Nat Biomed Eng*. (2021) 5:522–32. doi: 10.1038/s41551-021-00711-2

51. Ning W, Lei S, Yang J, Cao Y, Jiang P, Yang Q, et al. Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes *via* deep learning. *Nat Biomed Eng*. (2020) 4:1197–207. doi: 10.1038/s41551-020-00633-5

52. Mitani A, Huang A, Venugopalan S, Corrado GS, Peng L, Webster DR, et al. Detection of anaemia from retinal fundus images *via* deep learning. *Nat Biomed Eng*. (2020) 4:18–27. doi: 10.1038/s41551-019-0487-z

frontiers | Frontiers in Public Health

TYPE Opinion
PUBLISHED 28 October 2022
DOI 10.3389/fpubh.2022.1030304

OPEN ACCESS

EDITED BY
Steven Fernandes,
Creighton University, United States

REVIEWED BY
Fancun Meng,
Shantou University, China
Jayanthi T,
SRM Institute of Science and Technology, India
Raviraj P,
Visvesvaraya Technological University, India

*CORRESPONDENCE
Sujitha Juliet
sujitha@karunya.edu

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 08 September 2022
ACCEPTED 10 October 2022
PUBLISHED 28 October 2022

CITATION
Raimond K, Rao GB, Juliet S, Tamilarasi SRG, Evangelin PS and Mathew L (2022) An emerging paradigms on cervical cancer screening methods and devices for clinical trails.
Front. Public Health 10:1030304.
doi: 10.3389/fpubh.2022.1030304

An emerging paradigms on cervical cancer screening methods and devices for clinical trails

Kumudha Raimond[1], Gadudasu Babu Rao[2], Sujitha Juliet[1]*, S. Rubeena Grace Tamilarasi[3], P. S. Evangelin[3] and Limson Mathew[2]

[1]Department of Computer Science Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India, [2]Department of Mechanical Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India, [3]Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India

KEYWORDS

cervical cancer, HPV screening, VIA, cervical cytology screening test, machine learning, deep learning

Introduction

Every year around 5,70,000 women are affected with cervical cancer and over 3,11,000 women die from the disease (1). Although there are techniques of screening in various forms and types around the world, most of the knowledge or the technique does not reach the interior parts of the world, like that which are in developing countries. Most of the rural areas either lack good health care support systems or high-level screening equipment, especially when it comes to cancer screening. Most women notice changes in their body only when the symptoms are severe or close to higher rates of malignancies. Fortunately, different studies have come up with various techniques that are cost effective, simple and efficient. In the following sections, major types of screening and the subcategories of testing are described.

Cervical cancer is curable, unlike the majority of malignancies. There are two methods of prevention. First, through immunization, and second, by routine screening that can find HPV infection or aberrant cells before they become malignant (2). Though infections may be cured within 2 years, 10% of the infections may last longer than 2 years. A chronic infection raises the possibility of getting precancerous or, ultimately, aggressive cancer. However, there is a safe and effective vaccine that can stop HPV 16 and HPV 18 infections. Starting at age nine, vaccinations are preferred for young girls. If a high-grade precancerous disease manifests, it must be surgically removed before developing into cervical cancer.

Literature survey

Van Baars, studied that the primary screening with high-risk human papillomavirus (hrHPV) detection has been advocated to prevent cervical cancer. While given the chance to self-sample for hrHPV testing, women who are not already attending screening (non-responders) are more likely to participate. Dry Evalyn Brush system is as good for self-sampling compared to physician-taken sample for hrHPV detection and is highly acceptable to women. The fact that this study was conducted in a hospital setting is a drawback. Self-samples were always taken prior to practitioner smears, which is another theoretical restriction (3). Parashari et al. found that the Magnivisualizer has an increased identification rate of early malignant tumors from 60 to 95% when compared to unaided visual inspection. It also has allowed for the detection of 58 percent of low-grade dysplasia cases and 83 percent of high-grade dysplasia cases that would not have been detected by simple visual assessment. The Magnivisualizer has a poorer sensitivity for detecting low-grade dysplasias, although this may not be a severe drawback because most low-grade dysplasias tend to regress even in the absence of treatment (4).

Veena Singh et al., showed that, in resource poor environments where colposcopic services are not offered on a local level, a cost efficient, handheld instrument called magnivisualizer is a useful for identifying cervical precancerous and cancerous lesions. In comparison to Visual inspection with acetic acid (VIA), this device demonstrated higher sensitivity (83 vs. 54%) without sacrificing specificity in the detection of severe precancerous lesions of the cervix. Due to the standard of colposcopy has limited specificity, it leads to unnecessary biopsies, therefore it cannot be used as a substitute (5).

Saleh, found that, in comparison to a Pap smear, VIA is an effective screening tool because it is a simple test with a low cost and great sensitivity. It can be therefore used in low-resource locations as an alternative cervical cancer screening method. The sensitivity of Pap smear was 50.1%, specificity was 93.1%, and its negative and positive predictive values were 89.3 and 65.6%, respectively. VIA's sensitivity was 90%, specificity was 37%, and its prediction accuracy was positive. Fifty-two percent and an 81% negative predictive value. Because of the less PPV of VIA, the issue of multiple false positives, discourages the see-and-treat strategy. Although, PPV linked to incidence, the VIA test's capabilities might increase if a see-and-treat approach were used in a high incidence of cervical cancer in a high-risk area (6).

The findings of Emre Ozgu et al. suggest that TruScreen, has 86.1% of sensitivity, and can be used as a cervical cancer screening test that offers quick results without the requirement for a professional. Because it eliminates the need for pathologists and subjectivity in Pap smear interpretation, Cervical cancer screening is possible with TruScreen, particularly in nations with low socioeconomic level. The effectiveness of screening did not significantly enhance when TruScreen and HPV testing were combined (7).

Muszynski et al. performed a study where Colposcopy alone showed 61% of sensitivity and 80% specificity for identifying high-grade lesions. Zedscan and colposcope together exhibited a sensitivity of 93%−100%, and between a range of 91 and 100% negative predictive value (8). Based on the above literatures, there are certain methodologies and techniques with which cervical cancer screening is done. A detailed explanation of the various methods is discussed in methodology. From the literatures it is also observed that each of the techniques has their own advantages and disadvantages.

Methodology

There are various ways of screening, testing, and diagnosing cervical cancer. The below mentioned are mostly used for cervical cancer and these are as follows:

- Screening using Tissue Scrapping.
- Screening using Visual inspection with acetic acid (VIA).
- Screening using Devices.
- Screening using artificial intelligence (AI) and machine learning (ML) techniques.
- Screening using Mobile technology.

Screening using tissue scrapping

Cervical screening checks the health of the cervix. It helps to prevent cancer or treat them if any abnormality is found. In this method, a small portion of the cervical tissues are smeared using swab test brushes and are tested in laboratories for traces of HPV infections. There are two major methods through which this screening takes place, one of the methods is the Pap smear test and the other is HPV-DNA test. The Pap smear test is considered as the golden standard for cervical cancer screening (9).

Pap smear

Typically, a pelvic exam is performed in addition to the Pap smear (10, 11). In some circumstances, HPV test may be administered to females older than 30 in place of a Pap smear. Based on the type of test, the doctor either places the cell sample obtained from the woman' cervix onto a glass slide (conventional) or place it in a container containing a specific liquid to preserve the sample (liquid-based) (12).

Then the samples are then taken to a lab where they are examined under a microscope for cell features that might point

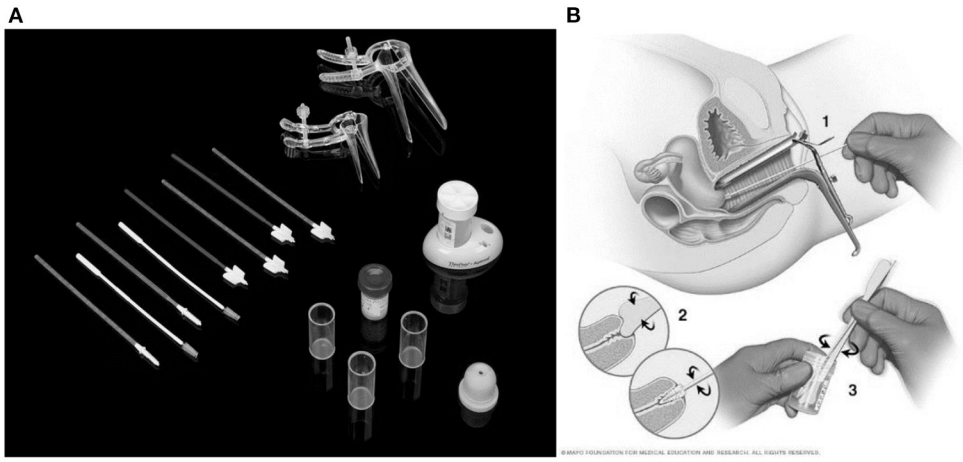


FIGURE 1
(A) The Pap smear test kit. **(B)** The procedure depiction.

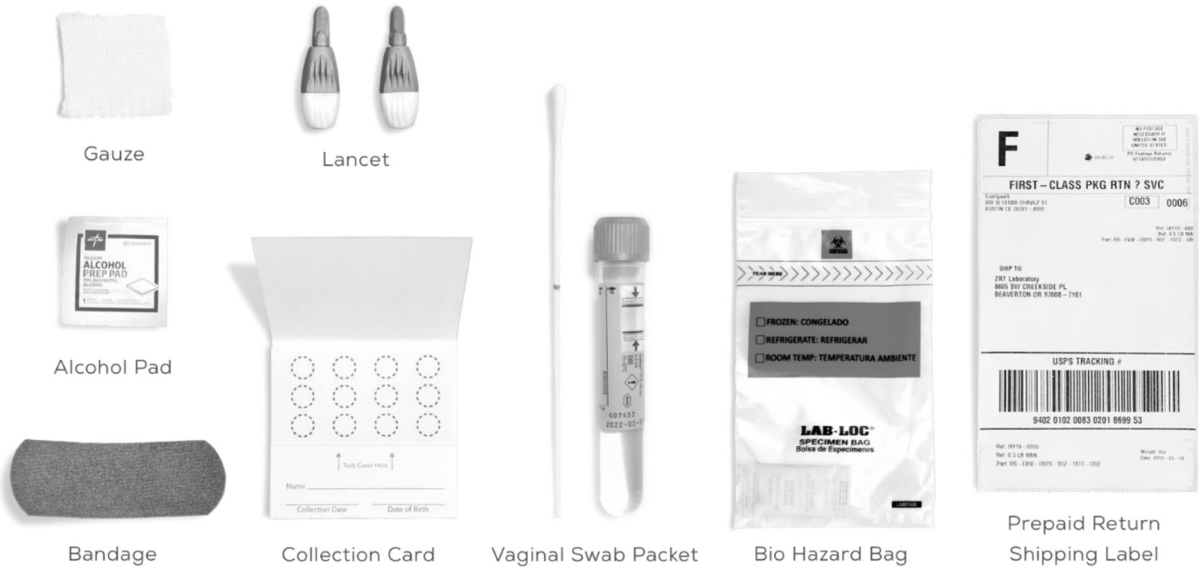


FIGURE 2
The HPV-DNA test kit.

to cervical cancer or a precancerous condition. Figure 1A shows the Pap smear test kit and Figure 1B shows the procedure depiction (13).

HPV-DNA test

It is common to have HPV infection around the genitals. Cervical cancer and other malignancies are caused by specific high-risk type of HPV. Low-risk types of HPV may cause genital warts in the vagina, cervix, and on the skin. In general, it is not advised to use the HPV-DNA test to identify low-risk HPV infections. This is because majority of low-risk lesions are physically recognizable. The medical professional inserts a device known as a speculum into the vagina, opens it slightly and gently collects the cells from the area around the cervix (14). Figure 2 shows the HPV-DNA test kit.

The cells are delivered to a lab where a microscope examination will take place. This examiner tests the cells to see if they contain genetic material (referred to as DNA) from cancer-causing HPV strains (15). To identify the exact type of HPV, further testing may be conducted. A Pap smear may be

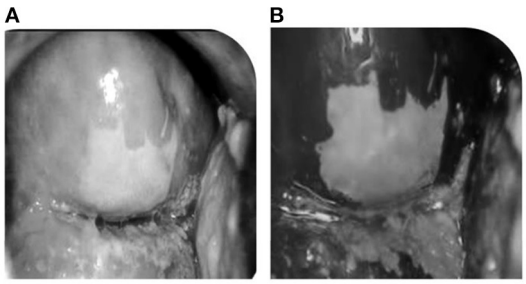


FIGURE 3
(A) Before screening with acetic acid. **(B)** After the acetic screening.

substituted with the HPV DNA test. Co-testing is the term used when they are carried out together.

Screening using VIA

VIA is a screening method in which the cervix is observed after the application of 3%−5% of acetic acid in the cervix region which results in acetowhite lesions. Figures 3A,B shows the result of before and after applying acetic acid. VIA offers the advantages of being simple to use, affordable (16), and sensitive when compared to Pap smear, and quick results assessment (17). As a result, VIA is a good way of cervical cancer screening in many regions of the world, particularly in areas with limited resources. Variations in sensitivity and specificity could be caused by a variety of factors, including the following:

- Expertise training
- Light source variation, and
- The procedure for making a 4%−5% acetic acid solution and storing it.

In poor countries with limited resources, VIA can be utilized as a mass screening method for cervical cancer. It was reported that at the low grade squamous intraepithelial lesion (LSIL) threshold, VIA was less sensitive i.e., is 86.7% which is lesser than that of clinical cytology with a sensitivity of 91.4%, but the difference was not statistically significant (18). HPV testing outperformed cytology in terms of sensitivity, but there was no significant reduction in specificity (84.2 vs. 86.6%).

In addition to VIA tests there are methods that gives importanceto a white light visual inspection of the cervix; white light enables the correct site of biopsy to be selected. The majority of rural clinics utilize a torch or a regular tungsten bulb, which misses many severe lesions. Through this study, the usage of white light is highly advisable for screening purpose. When compared to Pap smear, VIA has a high sensitivity.

FIGURE 4
The AV Magnivisualizer.

Screening using devices

From the previous studies, it is understood that a clinical test includes the collection of tissues and is slower when compared to other methods. Although the success rates of cancer screening ishigh, the scrapping method may disturb the patient's convenience. In order to be more efficient, cost effective and quick, current studies have come up with techniques that does not involve scrapping of tissues and does not infuse any discomfort. This section will discuss about the modern cervical cancer screening devices and their efficiency.

AV Magnivisualizer

It is a low-cost technology for screening uterine cervical cancer using magnivisualizer. It increases the identification rate of early malignant tumors from 60 to 95% when compared to single-handed visual inspection. It also allows the detection of 58% of low-grade dysplasia cases and 83% of high-grade dysplasia cases that would not have been detected by simple visual assessment. The magnivisualizer is highly sensitive, with a sensitivity of around 57.5% in detecting low-grade dysplasia, when compared to 75.3% of cytological evaluation (5). For higher degrees of lesions, however, the two approaches had equivalent sensitivity. The magnivisualizer had a 94.3% specificity, while cytology had a 99% specificity.

The AV Magnivisualizer, has a complete spectrum of visible light (white light) and interchangeable magnification. In Figure 4 the AV Magnivisualizer is shown.

The sole accessible light source in primary health Center outdoor situations is usually a tungsten bulb providing yellow light attached to a torch or examination light. On lesions with a pinkish mucosal background, this type of light has a masking effect. The handheld Magnivisualizer can be considered a suitable tool for identification of cervical precancerous and cancerous lesions in low-resource settings where colposcopic services are not available at the community level. However, due to its low specificity, it cannot replace colposcopy, which results in numerous needless biopsies.

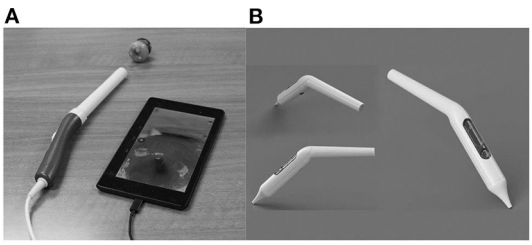


FIGURE 5
(A) 5-mega pixel POCkeT. **(B)** 2-mega pixel POCkeT.

FIGURE 6
TruScreen device.

POCkeT

Point of Care Tampon (POCkeT) is a Novel Low-Cost device that can capture images and can be used to diagnose cervical lesions. By delegating cervical cancer screening to community health workers, the portable, low-cost method has the potential to enhance access to cervical cancer screening in low-resource settings. Women who enter the screening cascade for the first time are usually not familiar with the procedure of having a speculum and are intimidated by the idea of having a cold metal object inside their bodies. This barrier was the reason that ultimately led to the conceptualization of the pocket colposcope (19).

The pocket colposcope can be inserted through the speculum to provide a close-up view of the cervix to take a picture. When the colposcope is close to the cervix, a set of high-quality pictures are obtained that are better than that of colposcopes on the market, and are both effective in cost and size. As seen in Figures 5A,B there are two versions of the pocket colposcope (20), one with a 5-mega pixel camera that can be used to obtain images *via* insertion through a speculum and one with a 2-mega pixel camera that is more slender and can be inserted into a tampon-like introducer called the Calla scope, to enable speculum-free visualization of the cervix.

TruScreen

TruScreen is a unique, proprietary Opto-Electrical technology to evaluate the tissue of the cervix. Unlike cytology, TruScreen does not only examine surface epithelial cells, it produces specific frequencies of light transmitted through the cervical tissue to identify changes in the basal and stromal layers. There are four LEDs that sequentially emit light at three wavelengths, distant red, infrared and green. Electrical measurements test the cell's resistance to current to characterize the tissue. This characterization of these tissues is called electrical impedance spectroscopy (21). As seen in Figure 6, the TruScreen system consists of a disposable Single Use Sensor (SUS), a Handheld Device (HHD), and an Intelligent Cradle (IC) that work in concert to detect and classify the cancerous and precancerous changes in the cervix.

First, many areas on the cervix are gently touched using a pen-like wand wrapped in a SUS. The SUS has electrodes and a precision lensthat interact with the cervix. During this process, it transmits and receives low-level optical and electrical information from the cervical tissue.

The signals are then analyzed by an integrated AI-enabled algorithm on the TruScreen Handheld Device, which compares them to a database of 2,000 patients from various ethnic and geographic backgrounds who have different histology diagnoses. Physicians receive immediate results from this analysis, which detects the presence of abnormal (cancerous and precancerous) cells in the cervix. In contrast to traditional Pap tests, which can take weeks or even months to provide a result in some countries, each TruScreen examination produces results in 1–2 min.

FIGURE 7
The ZedScan cervical probe.

ZedScan

ZedScan is a diagnostic gadget that makes use of an accessory to conventional colposcopy to offer an evaluation of the cervical epithelial tissue in real time. Electrical Impedance Spectroscopy (EIS) is a scientifically-proven technique to distinguish among normal, pre-cancerous and cancerous tissues (neoplasias) (8). This technique is likewise suitable for the prognosis of diverse cancers and pre-cancerous conditions. Figure 7 shows the ZedScan cervical probe.

FIGURE 8
CervAstra.

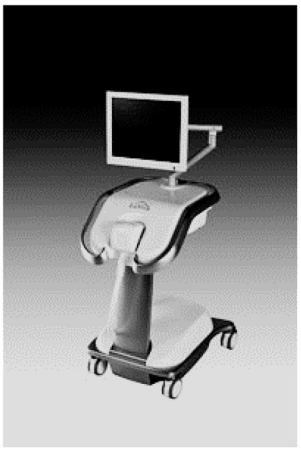


FIGURE 9
LuViva Scan device.

ZedScan makes use of EIS technique to distinguish among normal, pre-cancerous and cancerous tissue at the cervix based on electrical properties. When used along with colposcopy, ZedScan has established extra accuracy in detecting cervical disease.

CervAstra

There are many devices that detects and screens cervical cancer and Pap smear is one of the predominant ways. One of the greatest disadvantages of the Pap smear is that it takes a long time to receive the test results. In order to cater to this problem CervAstra was invented (22). CervAstra is a device to detect Cervical Cancer using a Computational Pathology platform. Figure 8 shows the device.

CervAstra analyzes Pap smear samples at the Point-of-Care to state normal or abnormal in few hours compared to a longer duration depending on the location of sample collection.

LuViva

It is a Hyperspectral Imaging Spectroscopy (HIS) technology based non-invasive scanning device that includes a base unit and a single-patient-use disposable probe. It is useful to scan the cervix with light source to detect the cancerous and pre-cancerous cells (23). Light reflected from the cervix is analyzed through a spectrometer. Figure 9 shows the LuViva Scan device.

Based on the information from the spectrometer, an image of the cervix will be generated which distinguishes the healthy from diseased tissue. The development of this technology has yielded seventeen patents.

Screening using AI and ML based applications

Numerous automatic and semi-automatic techniques have been developed as a result of the automatic analysis of colposcopy for the diagnosis of precancerous lesions. Neoplasia can be divided into several categories, and acetowhite zones can be classified as high- or low-risk, malignant or non-cancerous, normal or aberrant.

Many research works have been carried out to detect the cancer from Pap smear images and colposcopy images using ML and Deep Learning (DL) techniques. Many research works used Support Vector Machine (SVM) (24, 25), Adaptive Neuro Fuzzy Inference System (ANFIS) based classifier (25, 26), Bayesian classifier (27, 28) for cervical cancer detection and classification into cancerous or noncancerous. Many other ML techniques such K-Nearest Neighbor (KNN), Neural Networks, Adaboost classifier have been explored for the detection purpose.

By analyzing digitalized Papanicolaou-smear images with a primary training dataset, 15 different machine learning algorithms were built for the detection of cervical cancer. Almost all algorithms successfully identified the cancer cells. Although multilayer perceptrons are the highest among all the algorithms used in recent times multiple back propagation neural networks had a higher level of efficiency, whereas the other algorithms has a lower level of efficiency. The findings show that techniques based on AI can be utilized to develop tools for widespread cervical cancer screening (29).

Using colposcopy images, the traditional methods based on image processing and machine learning produced good results. However, these techniques require manual skill for feature extraction. The features can be automatically extracted from the data by deep learning. Apart from conventional ML techniques, existing DL architectures such as LeNet, VGG16/19, ResNet, MobileNet, Long Short-Term Memory (LSTM) and many proposed convolutional neural networks (CNN) have been used for the classification purpose.

The three most prominent Deep CNNs (ResNet-50, MobileNet, and NasNet) have been configured for training to

create linearly separable image feature descriptors in which the collected deep features are used to train a KNN classifier (30). In Ref. (31), the researcher has applied DL techniques to a new dataset acquired using smartphone, hand-held device and colposcope.

A hybrid method for the classification of cervical cells using deep learning-based segmentation and an ensemble-based classifier is proposed in Ref. (32) on a publicly available dataset. The model had an average accuracy and AUC of 99.7% and 0.996 for two-class classification and 75.55% and 0.909 for four-class classification, respectively.

Gated recurrent units (GRU) is applied in Ref. (33) to build a clinical event prediction model based on recurrent neural network (RNN). The results demonstrate that RNN-2-DT has a superior predictive effect compared to traditional models that directly predict clinical events.

Although AI is an advantage of the present digital era it also has major processing problems that might not give a complete solution for cancer screening techniques. It can be used for primary level of screening through which the presence of cervical cancer can be monitored. The future awaits for more development in AI algorithms that would facilitate in secondary level of cervical cancer screening.

Screening using mobile

From the different methodologies and studies done this far, it also required to explore the techniques based on mobile screening. With a growing technology based on mobile and smartphone, there are two major mobile based techniques in order to screen cervical cancer through smartphones, these are discussed in detail as follows.

Gynocular

The monocular colposcope called the Gynocular as shown in Figure 10 is a device that has optical capabilities when compared to basic colposcopes. This device screens the cancer using high resolution lens and is almost a smaller version of the traditional colposcope (34).

Using this device diagnostic forecasts from distant assessment were revealed to be equivalent to estimates from actual colposcopy evaluation for the diagnosis of CIN2+ lesions.

Mobile ODT

This device uses a method called the Enhanced Visual Assessment (EVA) Colpo, which is made up of a Mobile phone attached with magnifying lenses, a number of rechargeable batteries and LEDs for illumination. Mobile ODT has also developed a mobile app that retains patient information, preserve cervical pictures, and keeps track of biopsies along with other clinical findings (35).

Figure 11 shows the Mobile ODT device, this device has also proved to reduce false positive as well as false negative rates when compared to Pap tests, and the AI created in-house has been demonstrating good results.

FIGURE 10
Gynocular.

FIGURE 11
Mobile ODT.

Smartscopy

The ideology of smartscopy is that, instead of using an external flash and a device separately, the smartphones are used. This method is done after the application of a 3% solution of acetic acid to the cervix for 1 min, once the application is over a gynecologist inspects the cervix using Smartscopy with the activated flash mode pictures of the cervix (36). The recorded prominent areas revealed abnormal epithelium. Subsequently, the smartscopy findings, and the histological diagnosis was evaluated and was relatively successful. Although results were successful by using the iPhone 5S to inspect the uterine cervix for cervical cytology is welcoming, it might not be always always useful for screening cervical cancer.

Conclusion

This paper intended to study various techniques that were found to be successful, simple and cost effective when it came to screening cervical cancer. In equipments like digital colposcope and LuViva which are high in cost and sensitive in hardware are difficult to be moved to rural areas. This is a disadvantage caused due to whichtraditional ways have to be followed in rural areas. Though the knowledge of how different techniques and methods have been of great use in cervical cancer screening was studied. In future, studies may have to come up methods where all of the possible screening methods are put under one roof. A cost-effective method with the application of Machine learning techniques collaborated with a mobile application would be of great use, such a device would be both cost, time efficient compared to the effectiveness of other devices. When such applications become a reality, it would be of great use in remote sectors of many developing countries.

Author contributions

KR carried out the conceptualization of the research idea and supervised the study. GR carried out the analysis of the screening methods and verified the manuscript. SJ validated the systematic review. ST did a detailed survey on the screening methods for cervical cancer and drafted the manuscript. PE carried out the literature survey of screening techniques and methods. LM performed the study and drafted the manuscript. All authors checked and confirmed the manuscript finally.

Funding

We are gratefully acknowledging the financial support for this work by Department of Science and Technology-Biomedical Device and Technology Development (DST-BDTD) [Grant No:TDP/BDTD/30/2021(G)].

Conflict of interest

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

Publisher's note

References

1. Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *Lancet Glob Health*. (2020) 8:e191–203. doi: 10.1016/S2214-109X(19)30482-6

2. Harper DM, Tariq M, Alhawli A, Syed N, Patel MR, Resnicow K. Predictors of screening for cervical and colorectal cancer in women 50-65 years old in a multi-ethnic population. *Prevent Med Rep*. (2021) 22:101375. doi: 10.1016/j.pmedr.2021.101375

3. Van Baars R, Bosgraaf RP, terHarmsel BW, Melchers WJ, Quint WG, Bekkers RL. Dry storage and transport of a cervicovaginal self-sample by use of the Evalyn brush, providing reliable human papillomavirus detection combined with comfort for women. *J Clin Microbiol*. (2012) 50:3937–43. doi: 10.1128/JCM.01506-12

4. Parashari A, Singh V, Sehgal A, Satyanarayana L, Sodhani P, Gupta MM. Low-cost technology for screening uterine cervical cancer. *Bull World Health Organ*. (2000) 78:964–7.

5. Parashari A, Singh V, Sehgal A, Mehrotra R. AV Magnivisualizer: a low-cost screening technology for early detection of precancerous and early cancerous lesions of the uterine cervix. *BMJ Innov*. (2015) 1:99–102. doi: 10.1136/bmjinnov-2015-000053

6. Saleh HS. Can visual inspection with acetic acid be used as an alternative to Pap smear in screening cervical cancer? *Middle East Fertil Soc J*. (2014) 19:187–91. doi: 10.1016/j.mefs.2013.10.003

7. Özgü E, Yildiz Y, Özgü BS, Öz M, Danişman N, Güngör T. Efficacy of a real time optoelectronic device (TruScreen^{TM}) in detecting cervical intraepithelial pathologies: a prospective observational study. *J Turk Ger Gynecol Assoc*. (2015) 16:41–4. doi: 10.5152/jtgga.2015.15199

8. Muszynski C, Dupont E, Vaysse B, Lanta S, Tidy J, Sergent F, et al. The impact of using electrical impedance spectroscopy (ZedScan) on the performance of colposcopy in diagnosing high grade squamous lesions of the cervix. *J Gynecol Obstet Hum Reprod*. (2017) 46:669–73. doi: 10.1016/j.jogoh.2017.08.007

9. *HPV and Pap Testing, what is cervical cancer screening, what the ways of tissue scrapping*. Available online at: https://www.cancer.gov/types/cervical/pap-hpv-testing-fact-sheet (accessed September 18, 2022).

10. *Pap smear test procedure depiction*. Available online at: https://collection.sciencemuseumgroup.org.uk/objects/co8614235/liquid-based-cytology-kit-for-cervical-screening-cervical-screening-kit (accessed September 14, 2022).

11. *Pap smear test procedure depiction*. Available online at: https://www.mayoclinic.org/tests-procedures/pap-smear/about/pac-20394841 (accessed September 22, 2022).

12. Johnson NL, Head KJ, Scott SF, Zimet GD. Persistent disparities in cervical cancer screening uptake: knowledge and socio demographic determinants of Papanicolaou and human papillomavirus testing among women in the United States. *Public Health Rep.* (2020) 135:483–91. doi: 10.1177/0033354920925094

13. Zhao S, Huang L, Basu P, Domingo EJ, Supakarapongkul W, Ling WY, et al. Cervical cancer burden, status of implementation and challenges of cervical cancer screening in Association of Southeast Asian Nations (ASEAN) countries. *Cancer Lett.* (2022) 525:22–32. doi: 10.1016/j.canlet.2021.10.036

14. *HPV Testing, what is hpv vaccination, what the ways of acetic acid application*. Available online at: https://www.cancer.gov/about-cancer/causes-prevention/risk/infectious-agents/hpv-and-cancer (accessed September 15, 2022).

15. *HPV DNA test procedure depiction*. Available online at: https://www.everlywell.com/products/std-test-female/ (accessed September 19, 2022).

16. Poli U, Bidinger P, Gowrishankar S. Visual inspection with acetic acid (VIA) screening program: 7 years experience in early detection of cervical cancer and pre-cancers in rural South India. *Indian J Commun Med.* (2015) 40:203–7. doi: 10.4103/0970-0218.158873

17. Sankaranarayanan R. Screening for cancer in low- and middle-income countries. *Ann Glob Health.* (2014) 80:412–7. doi: 10.1016/j.aogh.2014.09.014

18. Rani SA, Rama K. Comparative analysis of visual inspection with acetic acid and Lugol's iodine and Liquiprep^{TM} in cervical cancer screening with cervical biopsy as gold standard. *IOSR J Dent Med Sci.* (2016) 15:54–62. doi: 10.9790/0853-150735462

19. Mueller JL, Asma E, Lam CT, Krieger MS, Gallagher JE, Erkanli A, et al. International image concordance study to compare a point of care tampon colposcope to a standard-of-care colposcope. *J Low Genit Tract Dis.* (2017) 21:112. doi: 10.1097/LGT.0000000000000306

20. Lam CT, Krieger MS, Gallagher JE, Asma B, Muasher LC, Schmitt JW, et al. Design of a novel low cost point of care tampon (POCkeT) colposcope for use in resource limited settings. *PLoS ONE.* (2015) 10:e0135869. doi: 10.1371/journal.pone.0135869

21. Salazar-Campos JE, González-Enciso A, Diaz-Molina R, Lara-Hernández ME, Coronel-Martínez J, Pérez-Plasencia C, et al. Cervicouterine cancer screening–TruScreen^{TM} vs. conventional cytology: pilot study. *J Cytol.* (2018) 35:143. doi: 10.4103/JOC.JOC_111_17

22. Roy A, Das BR. Artificial intelligence in clinical diagnostics–an Indian perspective. *Int J Clin Diagn Res*. (2019) 7:1–10.

23. Thekkek N, Richards-Kortum R. Optical imaging for cervical cancer detection: solutions for a continuing global problem. *Nat Rev Cancer.* (2008) 8:725–31. doi: 10.1038/nrc2462

24. Nehra S, Raheja JL. Detection of cervical cancer using GLCM and Support Vector Machines. In: *2018 6th Edition of International Conference on Wireless Networks & Embedded Systems (WECON)*. Rajpura. (2018), p. 49–53. doi: 10.1109/WECON.2018.8782065

25. Xu T, Zhang H, Xin C, Kim E, Long LR, Xue Z, et al. Multi-feature based benchmark for cervical dysplasia classification evaluation. *Pattern Recognit.* (2017) 63:468–75. doi: 10.1016/j.patcog.2016.09.027

26. Karthiga Jaya B, Senthil Kumar S. Image registration based cervical cancer detection and segmentation using ANFIS classifier. *Asian Pac J Cancer Prevent.* (2018) 19:3203–9. doi: 10.31557/APJCP.2018.19.11.3203

27. Ponnusamy S, Samikannu R, Venkatachary SK, Sukumar S, Ravi R. Computer aided innovation method for detection and classification of cervical cancer using ANFIS classifier. *J Ambient Intell Humaniz Comput.* (2020) 12:6231–40. doi: 10.1007/s12652-020-02191-9

28. RamaPraba PS, Ranganathan H. Computerized lesion detection in colposcopy cervix images based on statistical features using Bayes classifier. *Adv Intell Soft Comput*. (2012) 132:597–604. doi: 10.1007/978-3-642-27443-5_69

29. Gupta R, Sarwar A, Sharma V. Screening of cervical cancer by artificial intelligence based analysis of digitized papanicolaou-smear images. *Int J Contemp Med Res.* (2017) 4:1108–13.

30. Pal A, Xue Z, Befano B, Rodriguez AC, Long LR, Schiffman M, et al. Deep metric learning for cervical image classification. *IEEE Access.* (2021) 9:53266–75. doi: 10.1109/ACCESS.2021.3069346

31. Xue Z, Guo P, Desai KT, Pal A, Ajenifuja KO, Adepiti CA, et al. A deep clustering method for analyzing uterine cervix images across imaging devices. *In: 34th International Symposium on Computer-Based Medical Systems (CBMS)*. Aveiro. (2021), p. 527–32. doi: 10.1109/CBMS52027.2021.00085

32. Sabeena K, Gopakumar C. A hybrid model for efficient cervical cell classification. *Biomed Signal Process Control.* (2022) 72(Part A):103288. doi: 10.1016/j.bspc.2021.103288

33. Yan Y, Zhao K, Cao J, Ma H. Prediction research of cervical cancer clinical events based on recurrent neural network. *Procedia Comput Sci.* (2021) 183:221–9. doi: 10.1016/j.procs.2021.02.052

34. Nessa A, Wistrand C, Begum SA, Thuresson M, Shemer I, Thorsell M, et al. Evaluation of stationary colposcope and the gynocular, by the Swede score systematic colposcopic system in VIA positive women: a crossover randomized trial. *Int J Gynecol Cancer.* (2014) 24:339–45. doi: 10.1097/IGC.0000000000000042

35. Kundrod KA, Smith CA, Hunt B, Schwarz RA, Schmeler K, Richards-Kortum R. Advances in technologies for cervical cancer detection in low-resource settings. *Expert Rev Mol Diagn.* (2019) 19:695–714. doi: 10.1080/14737159.2019.1648213

36. Rahatgaonkar V, Uchale P, Oka G. Comparative Study Of Smart Scope® visual screening test with naked eye visual screening and Pap test. *Asian Pac J Cancer Prevent.* (2020) 21:3509. doi: 10.31557/APJCP.2020.21.12.3509

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 04 November 2022
DOI 10.3389/fpubh.2022.1046296

OPEN ACCESS

EDITED BY
Yu-Dong Zhang,
University of Leicester,
United Kingdom

REVIEWED BY
Mazin Mohammed,
University of Anbar, Iraq
Amir Faisal,
Sumatra Institute of
Technology, Indonesia
Priti Bansal,
Netaji Subhas University of
Technology, India

*CORRESPONDENCE
Muhammad Attique Khan
attique.khan@hitecuni.edu.pk
Jaehyuk Cha
chajh@hanyang.ac.kr

SPECIALTY SECTION
This article was submitted to
Digital Public Health,
a section of the journal
Frontiers in Public Health

RECEIVED 16 September 2022
ACCEPTED 12 October 2022
PUBLISHED 04 November 2022

CITATION
Hamza A, Attique Khan M, Wang S-H, Alhaisoni M, Alharbi M, Hussein HS, Alshazly H, Kim YJ and Cha J (2022) COVID-19 classification using chest X-ray images based on fusion-assisted deep Bayesian optimization and Grad-CAM visualization.
Front. Public Health 10:1046296.
doi: 10.3389/fpubh.2022.1046296

COVID-19 classification using chest X-ray images based on fusion-assisted deep Bayesian optimization and Grad-CAM visualization

Ameer Hamza[1], Muhammad Attique Khan[1]*, Shui-Hua Wang[2], Majed Alhaisoni[3], Meshal Alharbi[4], Hany S. Hussein[5,6], Hammam Alshazly[7], Ye Jin Kim[8] and Jaehyuk Cha[8]*

[1]Department of Computer Science, HITEC University, Taxila, Pakistan, [2]Department of Mathematics, University of Leicester, Leicester, United Kingdom, [3]Computer Sciences Department, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia, [4]Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia, [5]Electrical Engineering Department, College of Engineering, King Khalid University, Abha, Saudi Arabia, [6]Electrical Engineering Department, Faculty of Engineering, Aswan University, Aswan, Egypt, [7]Faculty of Computers and Information, South Valley University, Qena, Egypt, [8]Department of Computer Science, Hanyang University, Seoul, South Korea

The COVID-19 virus's rapid global spread has caused millions of illnesses and deaths. As a result, it has disastrous consequences for people's lives, public health, and the global economy. Clinical studies have revealed a link between the severity of COVID-19 cases and the amount of virus present in infected people's lungs. Imaging techniques such as computed tomography (CT) and chest x-rays can detect COVID-19 (CXR). Manual inspection of these images is a difficult process, so computerized techniques are widely used. Deep convolutional neural networks (DCNNs) are a type of machine learning that is frequently used in computer vision applications, particularly in medical imaging, to detect and classify infected regions. These techniques can assist medical personnel in the detection of patients with COVID-19. In this article, a Bayesian optimized DCNN and explainable AI-based framework is proposed for the classification of COVID-19 from the chest X-ray images. The proposed method starts with a multi-filter contrast enhancement technique that increases the visibility of the infected part. Two pre-trained deep models, namely, EfficientNet-B0 and MobileNet-V2, are fine-tuned according to the target classes and then trained by employing Bayesian optimization (BO). Through BO, hyperparameters have been selected instead of static initialization. Features are extracted from the trained model and fused using a slicing-based serial fusion approach. The fused features are classified using machine learning classifiers for the final classification. Moreover, visualization is performed using a Grad-CAM that highlights the infected part in the image. Three publically available COVID-19 datasets are used for the experimental process to obtain improved accuracies of 98.8, 97.9, and 99.4%, respectively.

KEYWORDS

corona virus, multi-filters contrast enhancement, deep learning, Bayesian optimization, hyperparameters, fusion

Introduction

The coronavirus has recently spread throughout the world as a new infection. Coronavirus is typically spread by animals or humans (1, 2). It is discovered to be transmitted by bats as a result of animal transmission. Coronavirus also replicates in the human body with several other common coronaviruses, including 229E: Alpha, NL63: Alpha, OC43: Beta, and HKU1: Beta (3). The coronavirus disease outbreak was dubbed the coronavirus global pandemic or COVID-19 pandemic by the World Health Organization (WHO) in March 2020 (4). The disease known as COVID-19 is caused by a virus (SARS-CoV-2). Lung diseases range in severity from a common cold to a potentially fatal illness. Coronavirus illnesses were frequently accompanied by respiratory system diagnoses. Individuals may occasionally contract minor, self-limiting infections with severe consequences, such as influenza. Symptoms of respiratory problems, fatigue, and a sore throat include fever, cough, and breathing difficulties (5). The majority of researchers have emphasized the need for COVID-19-specific diagnostic methods, medications, or vaccinations to prevent its spread (6). Because of its higher sensitivity and specificity in terms of observations, the reverse transcription-polymerase chain reaction (RT-PCR) is the current gold standard for diagnosing COVID-19 (7).

Visual indicators could be used as an alternative strategy for quickly screening infected individuals (8). This infection's most prevalent symptom is respiratory sickness. For chest radiography, images (X-rays of the chest) are thought to be the most reliable visual signal. Radiologists examine these images physically to identify visual patterns that indicate the presence of COVID-19 (9). Even though traditional diagnosis has improved over time, it is still vulnerable to medical staff errors. It is also more expensive because each patient requires a diagnostic test kit. Medical-based imaging procedures, such as CXR and CT scans, are much faster, safer, and more widely available for screening (10). For COVID-19 screening, CXR image screening is superior to CT scans because it is more accessible and less expensive (11, 12). However, it may take some time to manually diagnose the virus using X-ray scans. If there is little or no prior knowledge and expertise about the infection and its characteristics, it may result in several inaccuracies and human-made mistakes. As a result, there is a compelling need to automate such operations on a large scale, and it should be accessible to all, so that treatment can become more effective, precise, and timely (13).

Previous research has used computer vision (CV) and artificial intelligence (AI) methods involving deep learning (DL) algorithms; specifically, CNNs have been validated as a realistic method for analyzing medical images (14, 15). A deep learning technique called a convolutional neural network was previously utilized to accurately identify pneumonia in CXR images of a patient's chest (16–18). The researchers introduced several CNN models for classification tasks, including ResNet50 (19), AlexNet (20), InceptionV3 (21), and a few others (22). Computer vision researchers have used pre-trained deep learning models in medical imaging, particularly for COVID-19 diagnosis and classification (23, 24).

Loey et al. (25) presented a Bayesian-based optimization DCNN model to classify coronavirus illness by using CXR images. The presented approach tuned the hyperparameters of DCNN models and extracted the high-level features. The data used in the experimental process were large in size and achieved 96% accuracy. This approach is limited by its high computing time due to the Bayesian optimization because it takes too much iteration during the training process. Yoo et al. (26) employed a hybrid technique model on CXR images by classifying the coronavirus using a decision tree classifier and deep learning. The created method achieved 95% accuracy. Wang et al. (27) designed a deep learning model-based transfer learning approach to identify the coronavirus. CXR images were utilized for this method. COVID-19 and healthy images were 565,537, respectively. The created deep learning technique gained 96.7% accuracy. They extracted high-level features and ML-based classifiers to create an efficient technique for improving the sensitivity of DCNN models. Chowdhury et al. (28) implemented a novel framework based on a CNN. They used a multiclass dataset that included COVID-19, pneumonia, and the healthy class. They constructed a CNN in the parallel pipeline and supplied crucial elements for the classification method. The suggested approach attained an accuracy of 96.9%, which was superior to the current techniques. Khan and Aslam (29) utilized the DCNN networks such as ResNet, DenseNet, and VGGNet and performed transfer learning concepts for training the models on the Chest X-ray dataset. The dataset includes 195 COVID-19 images and 862 normal images. On the selected dataset, the provided method achieved an accuracy of 99%. Che Azemin et al. (30) designed a ResNet CNN based on a deep learning algorithm to diagnose COVID-19 from CXR images. They considered the binary class problem—COVID-19 and healthy classes. The selected dataset was utilized for the training of the CNN model through transfer learning. The trained model achieves 72% accuracy, which is higher than the recent methods. Khan et al. (31) presented a DL and explainable AI-based framework for COVID-19 classification from CXR images. Transfer learning was utilized to train pre-trained deep models on enhanced images, and features were merged for greater information. Following that, the Whale–Elephant herding method is used to choose the best features, which are then classified using the ELM classifier. Few other techniques such as meta-classifier with deep learning approach for COVID-19 classification (32), novel CNN approach called CNN-COVID (33), optimization algorithm called novel crow swarm (34), and multi-agent deep reinforcement learning (35).

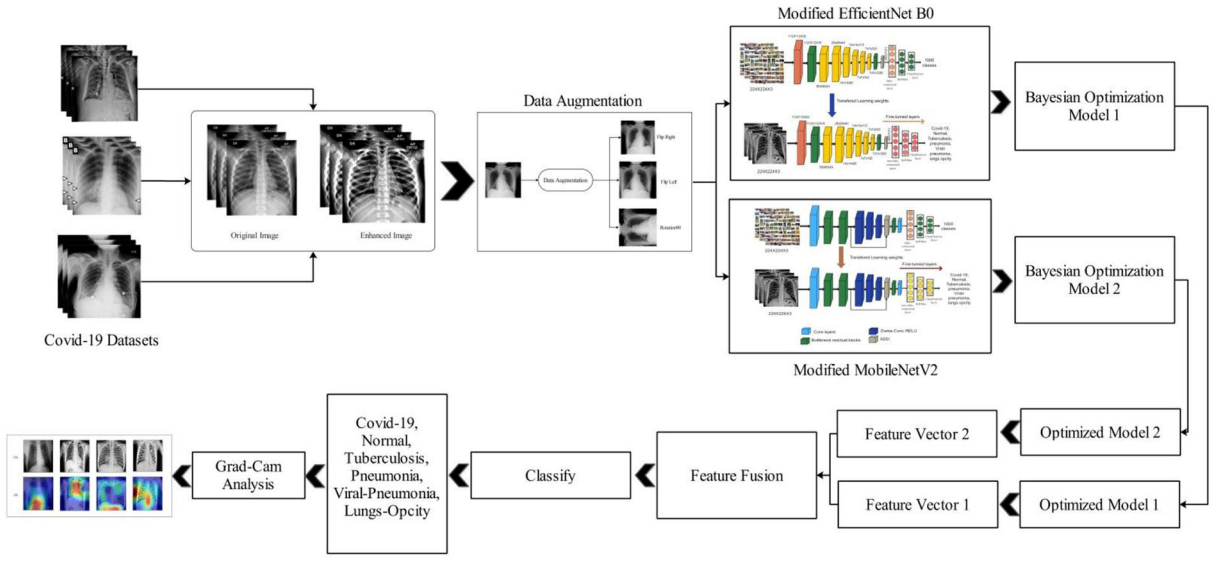


FIGURE 1
Proposed classification architecture for COVID-19 utilizing deep transfer learning and Bayesian optimization.

The models in the preceding studies were retrained using the transfer learning concept, which involves freezing the weights of a few layers to save computational time. They also used fixed hyperparameters like learning rate, momentum, mini-batch size, epoch count, etc. When there is a lot of variation in the results due to different hyperparameter values, this method is inefficient. In this work, we proposed a multimodal Bayesian hyperparameter optimization method for the training of deep learning models for COVID-19 classification. Moreover, an explainable AI-based diagnosis has been performed. Our major participation in this work is as follows:

- A multi-filter fusion-based hybrid technique is proposed for contrast enhancement that increases the local and global information of an image.
- Bayesian optimization is employed on deep learning models for the optimization of hyperparameters that helps in the better training of selected data.
- High-level features are extracted by both models and fused by a novel slicing-based serial fusion.
- Grad-CAM visualization is performed on the final classification, resulting in the colored visualization of the COVID-19, pneumonia, and tuberculosis-infected regions.

The manuscript is organized as follows. The proposed methodology such as multi-filters fusion-based hybrid contrast enhancement technique, Bayesian optimization of hyperparameters of DCCN models, deep transfer learning, feature extraction, fusion, and Grad-CAM for explainable AI, is presented in Section Proposed methodology. The findings of the proposed approach are shown in Section Experimental results and analysis, and Section Conclusion presents the conclusion.

Proposed methodology

The proposed methodology for the COVID-19 classification and explainable AI-based diagnosis is presented here. In the proposed method, multi-filter contrast enhancement and deep transfer learning with Bayesian optimization are employed. Figure 1 shows the proposed architecture based on Bayesian optimization and features fusion for COVID-19 classification. This figure illustrates that, in the first phase, data augmentation is performed on the selected datasets using a multi-filter contrast enhancement method and a few additional filters. Two pre-trained models, namely, EfficientNet-B0 and MobileNet-V2, are modified and trained using deep transfer learning and optimized the hyperparameters by employing Bayesian optimization. Features are extracted from both optimized models and fusion is performed by utilizing a new method named, slicing-based serial fusion. Finally, the samples are subjected to Grad-CAM analysis in order to pinpoint the source of the infection.

Contrast enhancement

Enhancing contrast is one of the most important and useful steps to enhance the vital objects in the images (36). Another goal of this step is to enhance the overall image quality. Medical image identification and interpretation mainly rely

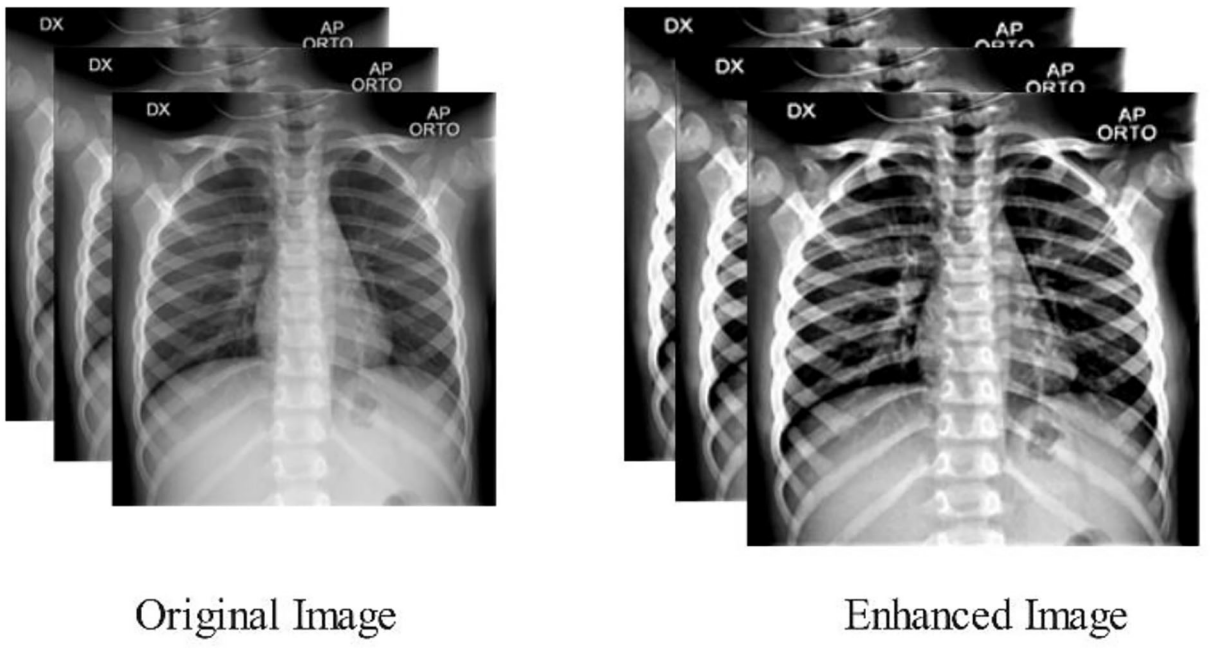


FIGURE 2
Samples enhanced images of multi-filters fusion technique.

on image enhancement methods (37). When segmentation is performed, poor contrast is always detected incorrectly. In the classification process, the enhanced images can extract more important features than the feature extraction through the original images. In this work, the images of the selected datasets have low contrast and poor quality. These problems may lead us to misclassification. Therefore, we designed a multi-filter technique by utilizing the fusion of different filters. First, top-hat and bottom-hat filtering are implemented and combined with information. After that, intensity values are adjusted by using a mathematical formula.

Consider, the COVID-19 datasets $\mathbb{D}$ having k images $\mathbb{D} \in \mathbb{R}^k$, where each image represented by $\mathcal{T}^k(h,w)$ and $(h,w) \in \mathbb{R}$. Each sample has resized into $\mathcal{N} \times \mathcal{M} = 224$. Suppose that the kernel σ with a value of 13. The top-hat is based on $\bullet$ opening operation and the bottom-hat is based on $\blacksquare$ closing operation. So, the top-hat and bottom-hat filtering are derived as:

$$\mathcal{T}_{top}(h,w) = \mathcal{T}^k(h,w) - \left(k^k(h,w) \circ \sigma\right) \quad (1)$$

$$\mathcal{T}_{bottom}(h,w) = \left(k^k(h,w) \blacksquare \sigma\right) - \mathcal{T}^n(h,w) \quad (2)$$

$$f'(h,w) = \mathcal{T}^n(h,w) + \mathcal{T}_{top}(h,w) - \mathcal{T}_{bottom}(h,w) \quad (3)$$

where $f'(h,w)$ represents the fused image of top-bottom filtering. In the next step, the adjust filter is employed on the resultant images from the top-hat and bottom-hat filters. Adjust filter boosts an image's lightness by transforming the points of the input pixels' intensities to new ones, with the mean amount of data absorbed in the low and high intensities being about 1.5%. The symbol p is the pixel value of the image, the *gamma* (γ) is a variable, which evaluates the form of the procedure among the coordinating coefficients (q,f) and (r,e).

$$\mathcal{A}dj^k(h,w) = \left(\frac{p-q}{r-q}\right)^{\gamma}(e-f) + e \quad (4)$$

$$F^k(h,w) = f'(h,w) + \mathcal{A}dj^k(h,w) \quad (5)$$

where $F^k(h,w)$ is the final enhanced image, visually illustrated in Figure 2.

Dataset collection and description

This study adopts an experimental technique that makes use of three publicly accessible datasets: COVID-GAN and COVID-Net small chest x-ray (https://www.kaggle.com/yash612/covidnet-mini-and-gan-enerated-chest-xray), COVID-19 radiography (https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database). CXR (pneumonia, COVID-19, TB) (https://www.kaggle.com/datasets/jtiptj/chest-xray-pneumoniacovid19tuberculosis). There are three classes in COVID-GAN and COVID-Net small chest x-ray datasets. COVID-19 radiography and CXR (pneumonia, COVID-19, and tuberculosis) consist of four classes. The original images are

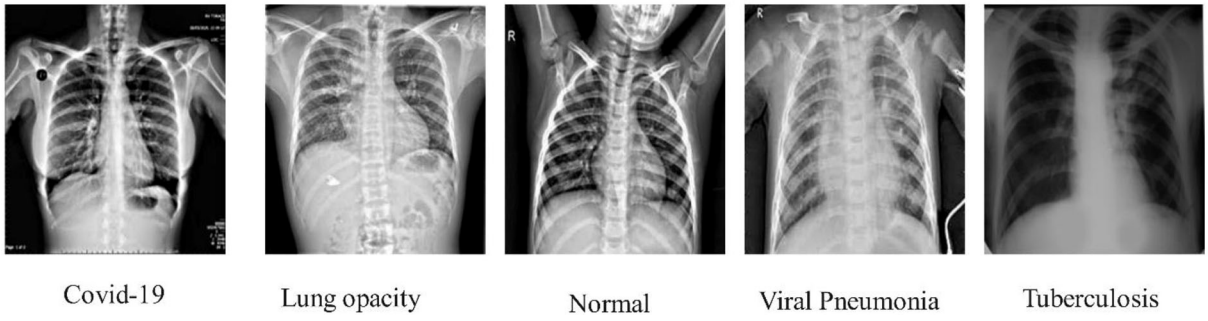


FIGURE 3
sCXR instances for the classification of COVID-19 and other infections.

TABLE 1 Complete explanation of selected datasets.

Classes	Original images	Augmented images	Training/testing images
COVID-19 radiography database			
COVID-19	3,616	6,000	3,000/3,000
Lung opacity	6,012	6,000	3,000/3,000
Normal	10,192	6,000	3,000/3,000
Viral pneumonia	1,345	6,000	3,000/3,000
COVID-GAN and COVID-Net mini–Chest X-ray			
Corona	461	6,000	3,000/3,000
Normal	1,575	6,000	3,000/3,000
Pneumonia	4,481	6,000	3,000/3,000
CXR (pneumonia, COVID-19, tuberculosis)			
COVID-19	566	6,000	3,000/3,000
Normal	1,575	6,000	3,000/3,000
Pneumonia	4,265	6,000	3,000/3,000
Tuberculosis	491	6,000	3,000/3,000

shown in Figure 3. These datasets are highly imbalanced as shown in Table 1. For balancing the dataset, we set 6,000 images in each class for all the datasets by utilizing data augmentation. Using the augmented dataset, 50% of images have been utilized for the training, while the rest of the 50% were used for the testing. In the data augmentation process, three primary functions are used: flip-left, rotate 90, and flip-right. The augmented images are visually shown in Figure 4.

EfficientNet deep features

The EfficientNet model, which ranks among the top models, achieved 84.4% accuracy on ImageNet for the classification task with a parameter size of 5.3 M (38). Deep learning architectures are intended to find simple but efficient solutions. By uniformly increasing depth, breadth, and resolution while reducing model size, EfficientNet outperforms competitor state-of-the-art models. The first step in the compound scaling technique is to find a grid that identifies whether distinct scaling dimensions of the baseline network connect to one another within the limits of a constrained set of resources. The optimal scaling factor for height, breadth, and resolution may be determined using this procedure. These coefficients are then added to the original network to make the final network the appropriate size (17).

The main building block for EfficientNet-B0 is the asymmetrical bottleneck MB Conv. Blocks in MB Conv consist of an expansion layer followed by a compression layer. Later, it was possible to connect bottlenecks directly while connecting a much smaller number of channels. When compared to conventional layers, the computational cost of this design's deep separable convolutions is around k^2, where k is the kernel size that determines the width and height of the 2D convolution window (39). In this work, we utilized the EfficientNet-B0 model for the features extraction. The model was originally trained on 1,000 classes and accepts the input size of 224 × 224 × 3. We fine-tuned the FC Layer with the new FC Layer which consists of COVID-19 classes. The updated model was trained by utilizing deep transfer learning and BO. The detail of Bayesian optimization (BO) is provided below. The objective of BO was to find the best hyperparameters for EfficientNet-B0 which gives the minimum error rate and increase the accuracy. The hyperparameters are selected dynamically *via* BO. The high-level features extracted from the average global pooling layer after the model has been trained on selected COVID-19 datasets and obtained a feature vector of size N × 1,280. Visually, the process of fine-tuning and deep transfer learning is shown in Figure 5.

MobileNet-V2 deep features

MobileNet-V2 employs depth-wise separable convolutions (DSCs) for portability and to solve the problem of data loss

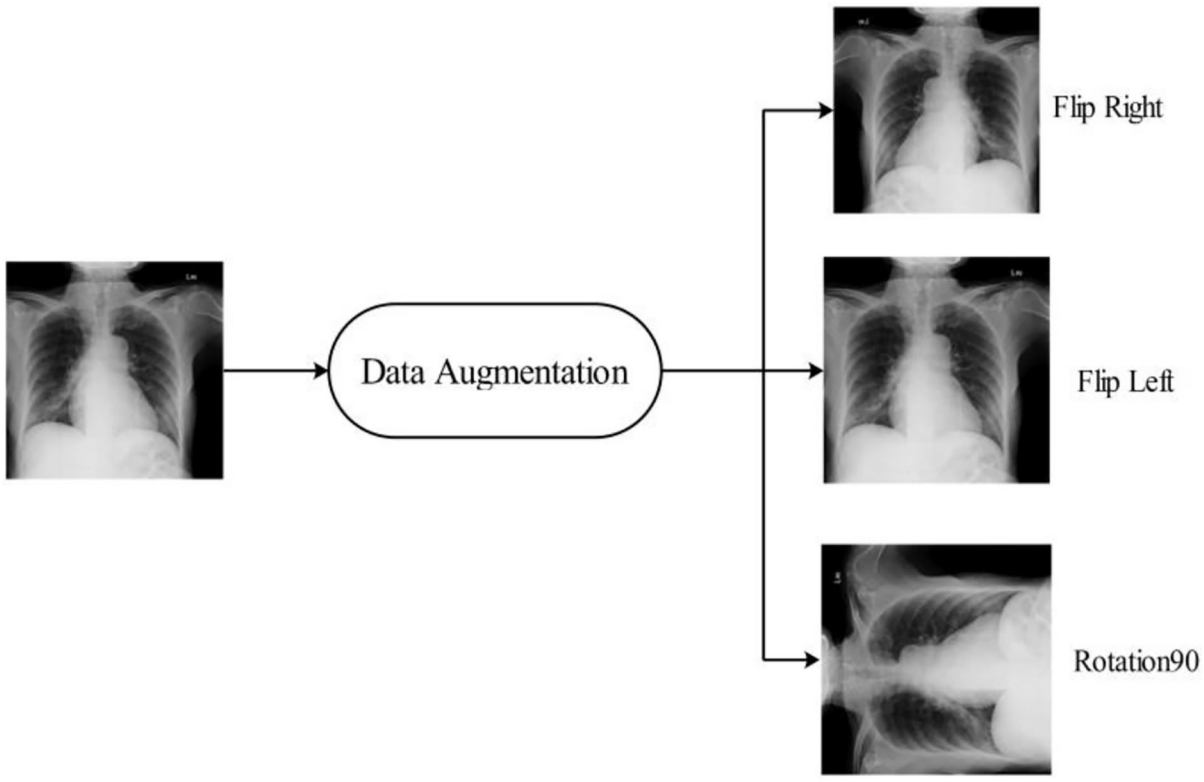


FIGURE 4
Sample images after data augmentation.

in non-linear layers inside convolution blocks. MobileNet-V2 has 5.3 million parameter values (40). The building block of MobileNet-V2 is shown in Figure 6. We used the MobileNet-V2 model in our proposed work for deep feature extraction. The model was pre-trained on the ImageNet dataset, which has 1,000 classes, and it takes input sizes of 224 × 224 × 3. The FC Layer was replaced with a new FC layer. As described in Section Hyperparameters optimization using BO, the updated model was trained using deep transfer learning, and the hyperparameters were optimized using BO. The trained model was utilized for the feature extraction. The activation is performed on global average pool (GAP) layer and retrieved features have a dimension of $N \times 1280$. Visually, the process of deep transfer learning is shown in Figure 6.

Hyperparameters optimization using BO

When using deep learning architectures, we need to adjust all of the hyperparameters in order to obtain classification accuracy. The selection of hyperparameters has a significant impact on the accuracy of the correct prediction (41). The goal of optimizing hyperparameters is to choose the values that get the best validation results. The hyperparameter optimization is calculated as:

$$x^* = argmin\, f\,(x) \tag{6}$$

where $f\,(x)$ is the objective score to minimize error rate when compared to the validation set, and x is the set of hyperparameters with a value in the domain where hyperparameter optimization evaluation is more expensive. It takes longer to train and is nearly impossible to achieve by hand with deep neural network models with many hyperparameters. BO has been used in simulations and machine learning models. To improve model performance, computer vision-based approaches use feed-forward network architectures to adjust hyperparameters. It simplifies the time-taking task of optimizing a number of parameters (42). Deep learning models need particular hyperparameter (HP) tuning. These parameters can be manually or automatically set. Although manual optimization produces adequate results, it is highly dependent on expertise and lacks consistency, making it less than ideal. HPs can be modified automatically by random and grid searches; however, some ineffectual sites may be unavoidable due to the inability to gain knowledge from previous searches. BO has garnered a lot of attention in parameter modification because of its distinct

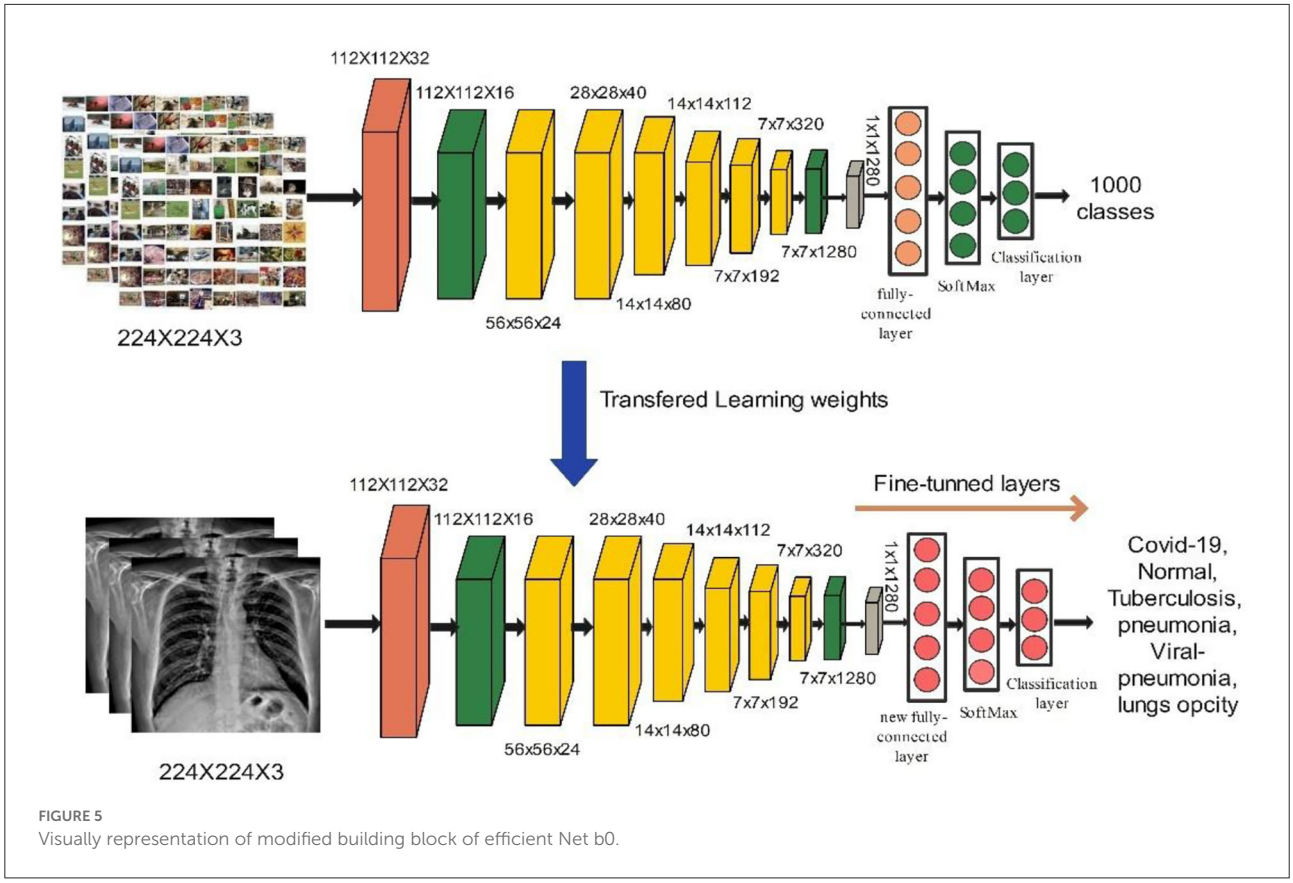


FIGURE 5
Visually representation of modified building block of efficient Net b0.

advantages. BO differs from other approaches in that it takes into account historical parameter information by updating the prior with Gaussian progress (GP). Also, BO has a very low number of iterations and a very fast convergence time. The BO method may also avoid local optimality when dealing with non-convex situations. The strong convergence and robustness of BO make it an excellent choice for optimizing HPs (43, 44). In our work, we utilized Bayesian optimization for a deep convolutional neural network to optimize the hyperparameters for achieving the minimum error of models. Section Depth, learning rate, momentum, and L2Regularization are the optimization parameters. The ranges of these parameters are shown in Table 2.

Proposed feature fusion

Feature fusion is an important step in which multi-directional information is combined to get a better output. As shown in Figure 1, features are extracted from two pre-trained models; therefore, fusion is important to combine the only important information (36). We proposed a novel feature fusion technique called slicing-based serial fusion in our study.

Consider, the first vector $\mathcal{V}_N^{k1}$, which has a dimension of $N \times 1280$, and the size of second vector $\mathcal{V}_N^{k2}$, which also has a dimension of $N \times 1280$ and is obtained by selected models EfficientNet-B0 and MobileNet-V2, respectively. Suppose $\mathcal{V}_N^{k3}$ is fused feature vector having dimension $N \times \boldsymbol{K}$. We selected a mid-point based on any from the selected vectors, which are computed as follows:

$$m = \frac{N}{2} \tag{7}$$

where m represents the mid-point of vector and N represents the total number of images used for feature extraction. Based on the m-value, $\mathcal{V}_N^{k1}$ and $\mathcal{V}_N^{k2}$ are divided into slices. The slices equation is calculated as:

$$\mathcal{V}_N^{11} = \mathcal{V}_N^{k1}\left(f_1 f_2 f_3 \ldots f_m\right) \tag{8}$$
$$\mathcal{V}_N^{12} = \mathcal{V}_N^{k1}\left(f_{m+1} \ldots f_{end}\right) \tag{9}$$
$$\mathcal{V}_N^{21} = \mathcal{V}_N^{k2}\left(f_1 f_2 f_3 \ldots f_m\right) \tag{10}$$
$$\mathcal{V}_N^{22} = \mathcal{V}_N^{k2}\left(f_{m+1} \ldots f_{end}\right) \tag{11}$$

where $\mathcal{V}_N^{11}$ and $\mathcal{V}_N^{12}$ represent the slicing that contains half and half of the features of vector $\mathcal{V}_N^{k1}$, and $\mathcal{V}_N^{21}$ and $\mathcal{V}_N^{22}$ represent the slicing that contains half and half of the features of vector $\mathcal{V}_N^{k2}$. The structure of slicing vectors is visually shown in Figure 7.

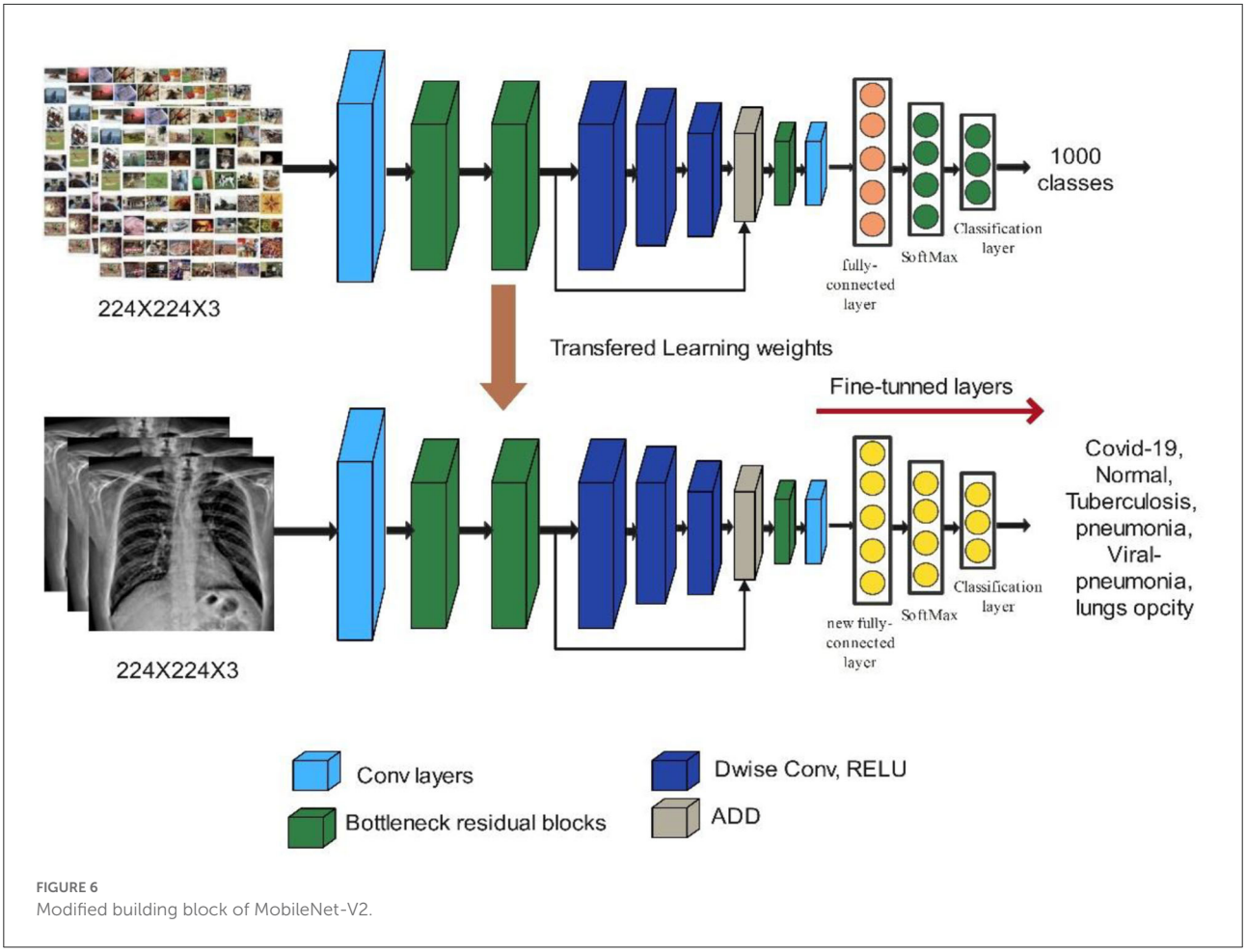


FIGURE 6
Modified building block of MobileNet-V2.

TABLE 2 Hyperparameters ranges for Bayesian optimization.

Hyperparameters	Ranges
Section depth	(1, 3)
Learning rate	[0.001, 1]
Momentum	[0.8, 0.98]
L2Regularization	$[1e^{-10}, 1e^{-2}]$

After slicing both vectors, the information is aggregated in the initial fused vector $\mathcal{V}_{fused}$. The sliced vectors are fused in this sequence $\mathcal{V}_N^{11}, \mathcal{V}_N^{21}, \mathcal{V}_N^{12}, \mathcal{V}_N^{22}$, respectively.

$$\mathcal{V}_{fused} = \begin{pmatrix} \mathcal{V}_N^{11} \\ \mathcal{V}_N^{21} \\ \mathcal{V}_N^{12} \\ \mathcal{V}_N^{22} \end{pmatrix}_{K \times N} \tag{12}$$

The output vector $\mathcal{V}_{fused}$ are attained with dimensions $N \times 2560$ but these features are mixed with each other by utilizing the slicing technique. In the next phase, features are refined further using a Kurtosis-based function. We tried to select the important features in the fused vector using this function.

$$Kr = \frac{\mu_4}{\sigma^4} \tag{13}$$

$$\mu_4 = \frac{E\left[(\mathcal{V} - \mu)^4\right]}{\sigma^4} \tag{14}$$

$$\sigma^2 = \sqrt{E\left[(\mathcal{V} - \mu)^2\right]}, \sigma = \sqrt{\sigma^2} \tag{15}$$

$$Fusion = \begin{cases} \mathcal{V}_N^{k3} & for \quad \mathcal{V}_{fused} \geq Kr \\ Ignore, & Elsewhere \end{cases} \tag{16}$$

Based on this equation, we obtained a final vector having dimension $N \times 1422$. This resultant vector is fed to machine learning classifiers for final classification.

Experimental results and analysis

For the experimental process, the datasets are split 50:50, indicating that 50% of the images are used to train the models and the remaining 50% are utilized for the testing process. The entire experimental process is carried out using

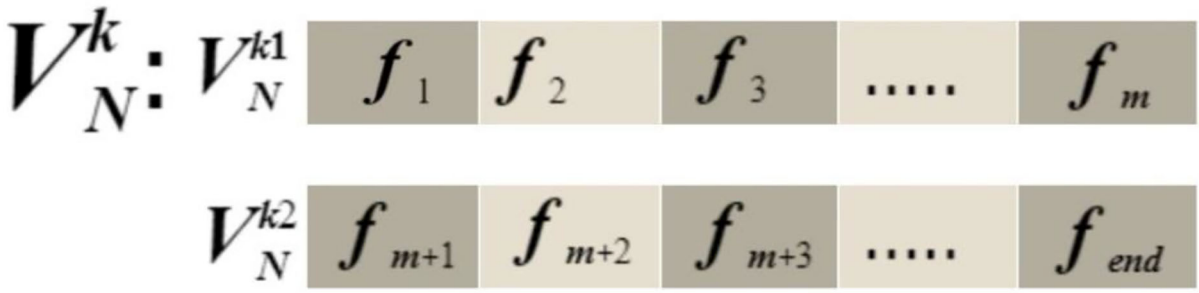


FIGURE 7
Structure of slicing vectors.

TABLE 3 Classification accuracy of modified EfficientNet-B0 Bayesian optimization features on COVID-19 radiography dataset.

Classifiers	Sensitivity	Precision	FPR	F1-score	Accuracy	Time
QSVM	**97.12**	**97.15**	**0.0075**	**97.13**	**97.2**	**19.996**
CSVM	97.22	97.22	0.0075	97.22	97.2	21.042
M G SVM	97.02	97.05	0.01	97.03	97.0	21.943
C G SVM	96.70	96.75	0.01	96.72	96.08	21.283
ESD	96.75	96.77	0.01	96.75	96.9	82.864
LSVM	97.05	97.07	0.10	97.05	97.0	18.952
SVM Kernel	96.92	96.95	0.0075	96.93	96.9	688.63
LRK	96.55	96.57	0.035	96.55	96.5	212.31
LD	96.55	96.57	0.010	96.55	96.5	18.935
WNN	95.52	96.57	0.015	96.04	96.5	18.004

Bold represent best values.

10-fold cross-validation. The static hyperparameters that are used during the training of deep models are epochs and mini-batch sizes having values 200 and 16, respectively. Moreover, the initial learning rate, stochastic gradient descent, momentum, L2Regularization, and section depth are optimized by utilizing Bayesian optimization. Multiple classifiers are used in this work for the classification results, including a support vector machine, wide neural network, ensemble subspace discriminant, and linear regression kernel. The classifier's performance parameters are sensitivity, precision, false positive rate, F1-score, accuracy, and computation time. Moreover, Grad-CAM analysis is conducted for further verification of the infected COVID-19 region in the image. All the simulations are conducted in MATLAB2022a executing on a workstation from MSI's GL75 Leopard series equipped with an 8 GB NVIDIA GTX graphics card, 512 SSD, and an Intel Core i7 10th generation processor.

COVID-19 radiography database results

Modified EfficientNet-B0 features

In this experiment, features are extracted from modified EfficientNet-B0. This model was trained through BO and transfer learning on the augmented dataset. Table 3 shows the classification accuracy of this updated model on the COVID-19 radiography dataset. In this table, it is noted that the QSVM classifier has a higher accuracy of 97.2% than the other classifiers listed. This classifier has a sensitivity rate of 97.12%, a precision rate of 97.15%, and an F1-score of 97.13%. Additionally, these values are determined for the remaining classifiers. During the classification process, the computation time of all classifiers is also recorded, with the wide neural network consuming the least time 18.004 (s) and the SVM kernel classifier taking the most time (688.63) (s).

Modified MobileNet-V2 features

From this experiment, the modified MobileNet-V2 model is fine-tuned and trained using BO on the COVID-19 radiography dataset. Table 4 shows the classification results of this experiment. From this table, the ESD classifier has an accuracy of 94.2% that is better than the other classifiers, listed in this table. This classifier has a 94.25% sensitivity rate, 94.28% precision rate, and F1-score is 94.28%. The numerical outcomes support the conclusion that the ESD outperforms the other classifiers. These values are also generated for the experiment's

TABLE 4 Classification accuracy of modified MobileNet-v2 Bayesian optimization features on COVID-19 radiography dataset.

Classifiers	Sensitivity	Precision	FPR	F1-score	Accuracy	Time
QSVM	92.85	92.92	0.022	92.88	92.8	28.351
CSVM	92.20	92.27	0.025	92.23	92.2	33.41
M G SVM	91.05	91.20	0.027	91.12	91.0	33.386
C G SVM	87.52	88.90	0.04	88.20	87.5	50.183
ESD	**94.25**	**94.32**	**0.02**	**94.28**	**94.2**	**106.56**
LSVM	93.75	93.80	0.022	93.77	93.8	20.353
SVM Kernel	94.25	94.32	0.017	94.28	94.2	89.6
LRK	90.45	90.62	0.032	90.53	90.8	127.41
LD	83.92	87.05	0.052	85.45	83.9	15.743
WNN	91.37	91.42	0.027	91.39	91.5	42.71

Bold represent best values.

TABLE 5 Classification results of proposed slicing-based serial fusion technique on COVID-19 radiography database.

Classifiers	Sensitivity	Precision	FPR	F1-score	Accuracy	Time
QSVM	**98.57**	**98.75**	**0.004**	**98.75**	**98.8**	**42.023**
CSVM	98.56	98.68	0.004	98.68	98.7	44.317
M G SVM	97.14	97.46	0.008	97.43	97.4	52.454
C G SVM	97.32	97.60	0.008	97.60	97.6	69.348
ED	98.15	98.45	0.005	98.47	98.5	454.86
LSVM	98.25	98.48	0.005	98.42	98.5	38.901
SVM Kernel	98.35	98.56	0.006	98.51	98.6	95.00
LRK	97.69	97.73	0.007	97.88	97.9	147.65
LD	94.56	94.62	0.017	97.65	94.7	42.13
WNN	97.82	97.90	0.006	98.07	98.1	48.073

Bold represent best values.

remaining classifiers. In this experiment, the amount of time for each classifier is noted and the linear discriminant classifier required the least amount of time of 15.743 s. In contrast, the LRK classifier was executed in 127.41 s, the highest of all the classifiers.

Proposed fused results

In this experiment, the proposed fusion approach is opted and fused the features of both optimized models. The fused vector is passed to the classifiers, which yielded the best accuracy of 98.8% on QSVM, which is higher than in experiments 1 and 2. Table 5 shows the detailed results of this experiment. QSVM has a sensitivity rate of 98.57%, a precision rate of 98.57%, and an F1-score of 98.57%. In addition, a QSVM confusion matrix is shown in Figure 8. This statistic indicates that the correct prediction rate for each class exceeds 97%. Also observed is the computing time of each classifier, with the linear SVM classifier executing faster than the others. This classifier's execution time is 38.901 s, while the longest execution time is 454.86 seconds (s). Comparing Tables 3, 4, it is observed that the fusion process improves accuracy, but time is increased due to the addition of extra features.

COVID-GAN and COVID-Net mini chest X-ray dataset

In this section, the results of the COVID-GAN and COVID-Net Mini Chest X-Ray dataset have been presented. In the first phase, features are extracted from the modified EfficientNet-B0 model. This model was trained through BO and transfer learning on the augmented dataset. Table 6 shows the classification accuracy of this model and obtained the 97.4% on Cubic SVM. The sensitivity, precision, and F1-score are 97.13, 97.26, and 97.15%, respectively. The classification computational time for all classifiers in this phase experiment is also recorded; the Quadratic SVM classifier has the shortest execution time of 14.318 (s) and the longest execution time of 128.74 (s). In the next phase, features are extracted through the modified MobileNet-V2 model. Features are passed to the classifiers and obtained the maximum accuracy of 93.5% on Cubic SVM, as shown in Table 7. This table shows that the sensitivity rate for

10.3389/fpubh.2022.1046296

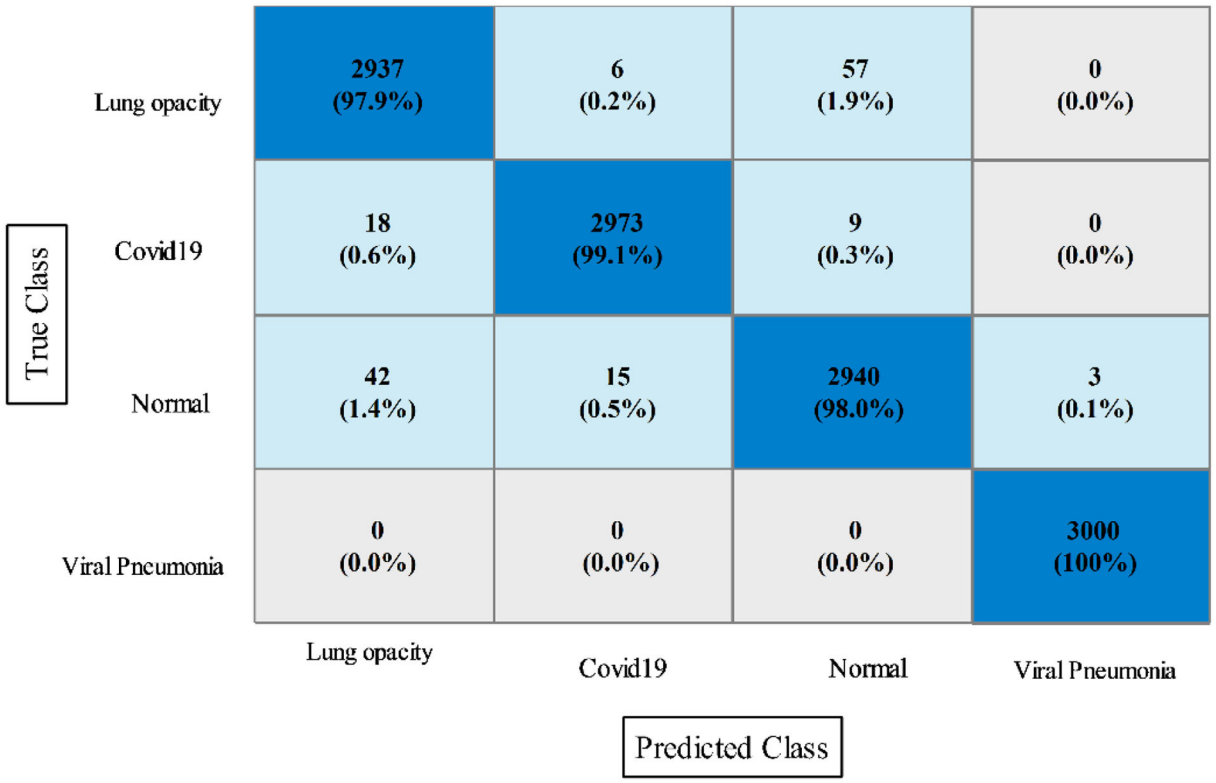


FIGURE 8
Confusion matrix of quadratic SVM utilizing the proposed slicing-based serial fusion technique on the COVID-19 radiography dataset.

TABLE 6 Proposed modified EfficientNet-B0 Bayesian optimized features results on COVID-GAN and COVID-Net mini chest X-ray dataset.

Classifiers	Sensitivity	Precision	FPR	F1-score	Accuracy	Time
QSVM	97.13	97.16	0.01	97.14	97.1	14.318
CSVM	**97.13**	**97.26**	**0.01**	**97.15**	**97.4**	**16.054**
M G SVM	96.95	96.96	0.01	96.95	96.9	18.33
C G SVM	95.36	95.53	0.02	95.44	95.4	18.085
ESD	96.70	96.73	0.01	96.71	96.7	121.690
LSVM	96.50	96.56	0.01	96.52	96.5	15.486
SVM Kernel	96.60	96.66	0.01	96.62	96.6	347.63
LRK	96.00	96.06	0.01	96.02	96.0	128.74
LD	95.30	95.36	0.02	95.32	95.0	15.587
WNN	96.76	96.76	0.01	96.76	96.8	59.033

Bold represent best values.

Cubic SVM is 93.35%, the precision rate is 93.46%, and the F1-score is also 93.47%. All classifiers' processing times are also noted down, and it is noted that modified EfficientNet-Bo features work better than modified MobileNet-V2 features.

In the final step, fusion is performed using the proposed approach, and results are shown in Table 8. According to the data in this table, the Cubic SVM classifier has the highest accuracy of 97.9%, which is higher than the previous two steps (Tables 6, 7). The sensitivity and precision rates are also improved—97.26 and 97.36%, respectively. A confusion matrix, as shown in Figure 9, can be used to confirm the performance of CSVM. In comparison to the previous two experiments on this dataset, accuracy improves significantly after the fusion of features of both optimized trained models. Also, it is noted that the time is increased after the proposed fusion step.

TABLE 7 Proposed method modified MobileNet-V2 utilizing Bayesian optimization results on COVID-GAN and COVID-Net mini chest X-ray dataset.

Classifiers	Sensitivity	Precision	FPR	F1-score	Accuracy	Time
QSVM	93.43	93.44	0.032	93.37	93.4	21.75
CSVM	**93.35**	**93.46**	**0.032**	**93.47**	**93.5**	**19.246**
M G SVM	92.25	92.54	0.037	92.43	92.5	23.361
C G SVM	91.30	76.26	0.060	79.77	91.3	17.924
ESD	93.03	92.13	0.034	92.94	92.9	148.03
LSVM	93.16	93.24	0.034	93.09	93.1	15.974
SVM Kernel	92.37	92.41	0.038	92.28	92.4	396.83
LRK	91.82	91.83	0.040	91.73	91.7	140.18
LD	86.20	86.46	0.069	85.98	86.2	24.760
WNN	92.52	92.68	0.036	92.66	92.7	95.177

Bold represent best values.

TABLE 8 Proposed slicing-based serial fusion results on COVID-GAN and COVID-Net mini chest X-ray dataset.

Classifiers	Sensitivity	Precision	FPR	F1-score	Accuracy	Time
QSVM	97.42	97.56	0.012	97.59	97.6	36.804
CSVM	**97.26**	**97.36**	**0.010**	**97.89**	**97.9**	**39.492**
M G SVM	96.99	97.08	0.014	97.03	97.0	48.462
C G SVM	94.98	95.01	0.016	95.03	95.5	43.458
ESD	97.12	97.46	0.012	97.01	97.5	339.12
LSVM	96.91	96.97	0.013	97.00	97.1	34.862
SVM Kernel	96.95	96.99	0.012	97.02	97.3	432.8
LRK	95.97	96.01	0.015	96.04	96.5	169.25
LD	93.01	93.06	0.019	93.05	93.9	87.107
WNN	96.96	97.04	0.011	97.04	97.6	37.795

Bold represent best values.

Chest X-ray (pneumonia, COVID-19, and tuberculosis)

The results of this dataset are shown in Tables 9–11. Table 9 shows the results after the feature extraction through a modified EfficientNet-B0 model that was trained through BO. For these features, quadratic SVM gives a better accuracy of 98.80%. The linear SVM has the least execution time of 21.394 (s), whereas the SVM Kernel classifier has the highest execution time is 3,497.8 (s). Table 10 shows the classification results of modified MobileNet-V2 features. In this table, the QSVM obtained the best accuracy of 97.0%. The F1-score is 97.05, the sensitivity rate is 96.02, and the precision rate is 96.6%. The computational time of this model is a little high than the modified Efficientnet-B0 features. Finally, fusion is improved, and the outcomes are shown in Table 11. From this table, the wide neural network classifier has the highest accuracy of 99.4%. Other measures of this classifier are calculated as well, including an F1-score of 99.38%, a sensitivity rate of 99.63%, and a precision rate of 99.79%. The confusion matrix of this classifier is also shown in Figure 10. Based on this figure, the sensitivity rate can be verified. After the fusion process, computational time increases but significantly improves accuracy.

Grad-CAM visualization and comparison

Grad-CAM is a CAM generalization that offers a localization map on the image based on the selected layer. In our work, we utilized global average pooling convolutional (GAP) feature maps that are directly fed into SoftMax (45). Grad-CAM needs to acquire a localization map that discriminates based on social status. Grad-CAM $\mathcal{D}^{c}_{GRAD-CAM} \in \mathbb{R}^{m\times n}$ in deep convolutional neural networks, after a convolutional layer has been trained, its feature mappings β are used to calculate the layer's gradient of g^c. Weights α^c_k are calculated using global average pooled interpretations of these gradients.

$$\alpha^c_k = \frac{1}{Z}\sum_i \sum_j \frac{\partial g^c}{\partial \beta^k_{ij}} \tag{17}$$

where weights are represented by α^c_k that defines the feature map k for a specific class c and serves as a partial linearization of

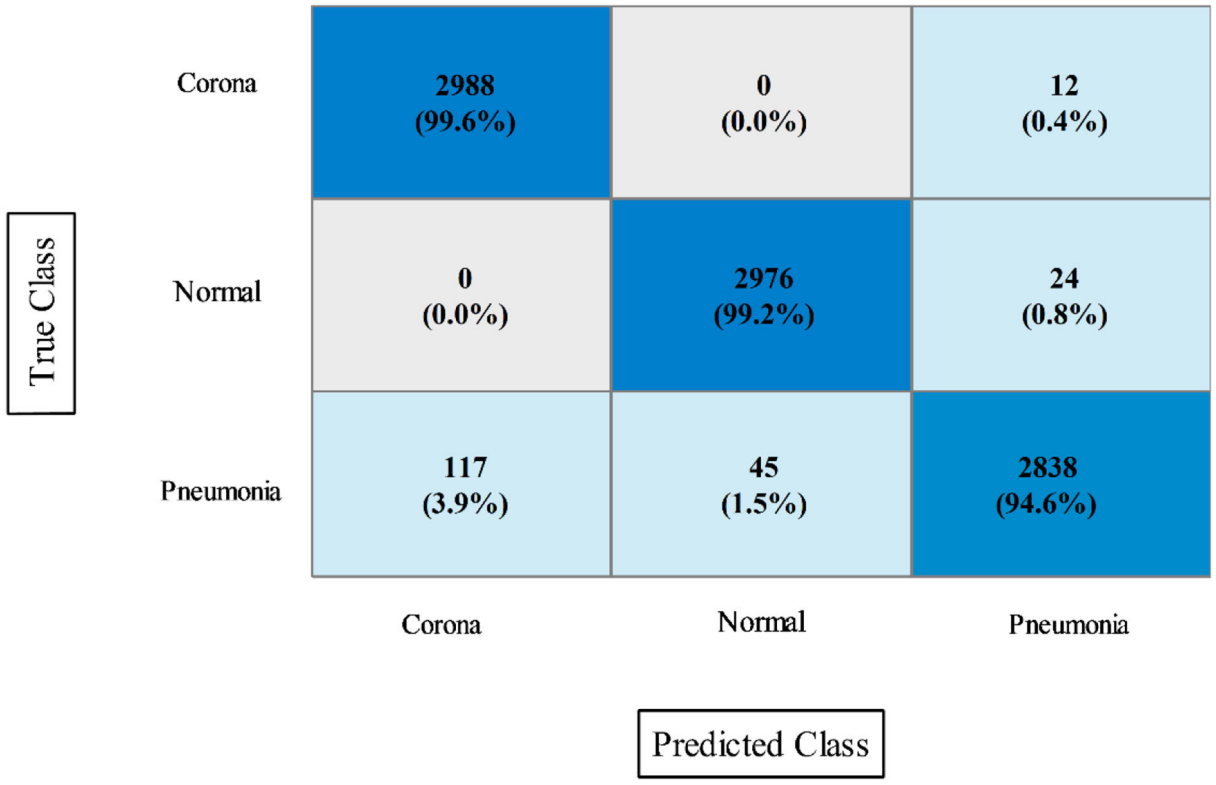


FIGURE 9
Confusion matrix of CSVM after proposed features fusion for COVID-GAN and COVID-Net mini chest X-ray dataset.

TABLE 9 Proposed modified EfficientNet-B0 Bayesian optimized features results on chest X-ray dataset.

Classifiers	Sensitivity	Precision	FPR	F1-score	Accuracy	Time
QSVM	**98.65**	**98.77**	**0.005**	**98.77**	**98.80**	**25.560**
CSVM	97.45	97.67	0.006	97.67	97.99	24.728
M G SVM	97.24	97.67	0.006	97.67	97.70	34.803
C G SVM	98.49	95.57	0.005	97.04	98.01	28.913
ESD	98.15	98.67	0.005	98.67	98.07	149.33
LSVM	98.14	98.72	0.005	98.72	98.70	21.394
SVM Kernel	98.65	98.73	0.005	98.05	98.07	3,497.8
LRK	98.23	98.32	0.005	98.32	98.30	290.05
LD	98.35	98.57	0.005	98.57	98.60	22.809
WNN	96.30	96.45	0.016	96.60	96.60	22.083

Bold represent best values.

the deep network downstream of β. It is not necessary for g^c to be a class score; alternatively, it might be anything that can be triggered in a different way. Our Grad-CAM heat map, like CAM, is a weighted combination of feature maps, but we then refine the findings using a ReLU:

$$\mathcal{D}^{c}_{GRAD-CAM} = RELU\left(\sum_{k} \alpha^{c}_{k}\beta^{k}\right) \tag{18}$$

This $\mathcal{D}^{c}_{GRAD-CAM}$ generates the primitive heat map that is normalized for visualization. In our article, we utilized the Grad-CAM for the analysis of the selected models. The Grad-CAM creates a heap map of the infected area of the lungs in the CXR images. A few resultant samples are shown in Figure 11. Finally, Table 12 shows a comprehensive evaluation of many computerized methods. This table contains a large number of newly introduced strategies that all use deep learning

TABLE 10 Proposed modified MobileNet-V2 Bayesian optimized features results on chest X-ray dataset.

Classifiers	Sensitivity	Precision	FPR	F1-score	Accuracy	Time
QSVM	**96.02**	**96.6**	**0.009**	**96.05**	**97.0**	**23.671**
CSVM	96.52	96.68	0.011	96.67	96.7	38.324
M G SVM	95.42	95.85	0.013	95.84	95.9	41.99
C G SVM	96.35	96.45	0.011	96.45	96.5	36.792
ESD	96.72	96.81	0.010	96.85	96.9	114.75
LSVM	96.58	96.83	0.010	96.82	96.9	30.087
SVM Kernel	96.69	96.75	0.010	96.82	96.8	878.01
LRK	96.06	96.21	0.012	96.40	96.4	276.76
LD	95.65	95.80	0.014	95.79	95.8	15.669
WNN	96.28	96.34	0.012	96.35	96.4	19.3

Bold represent best values.

TABLE 11 Proposed slicing-based serial feature fusion results on chest X-ray dataset.

Classifiers	Sensitivity	Precision	FPR	F1-score	Accuracy	Time
QSVM	98.90	99.23	0.002	99.1	99.4	54.639
CSVM	99.19	99.25	0.002	99.32	99.4	52.913
M G SVM	98.16	98.39	0.005	98.37	98.3	75.546
C G SVM	99.13	99.17	0.002	99.20	99.2	63.507
ESD	99.04	99.07	0.002	99.33	99.3	379.4
LSVM	99.21	99.25	0.002	99.35	99.4	43.302
SVM Kernel	99.26	99.34	0.0019	99.43	99.4	805.69
LRK	99.18	99.26	0.002	99.17	99.2	315.67
LD	98.56	98.65	0.003	98.95	99.0	116.59
WNN	**99.63**	**99.79**	**0.002**	**99.38**	**99.4**	**33.657**

Bold represent best values.

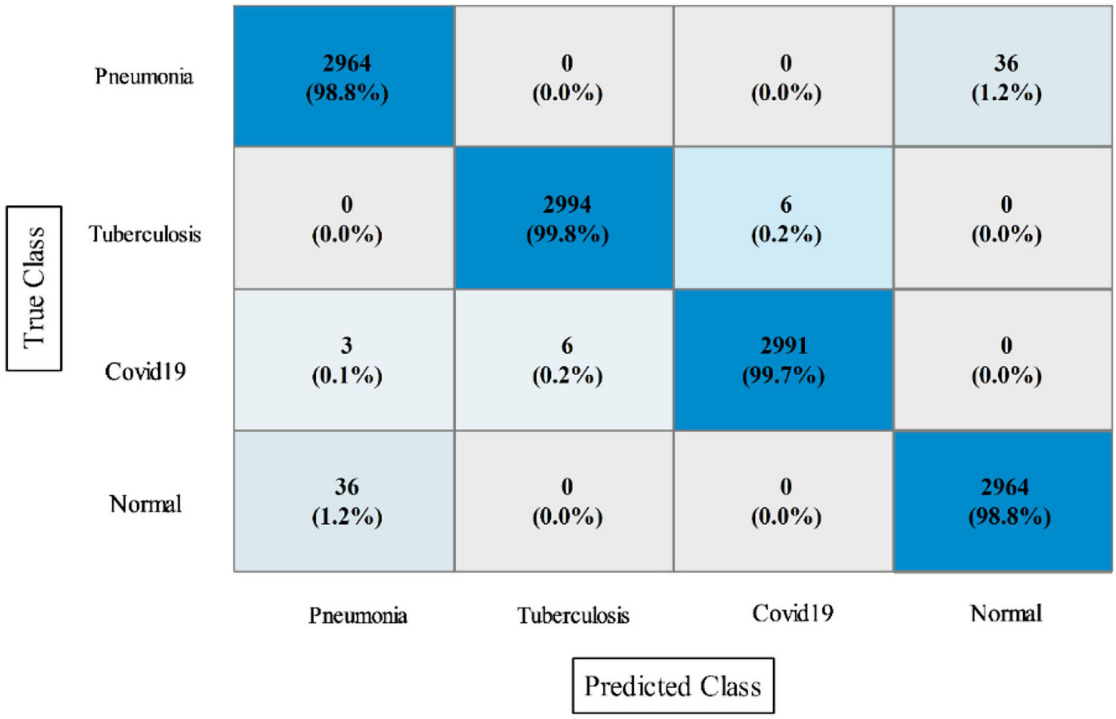


FIGURE 10
Confusion matrix of WNN for the proposed fusion on the chest X-ray dataset.

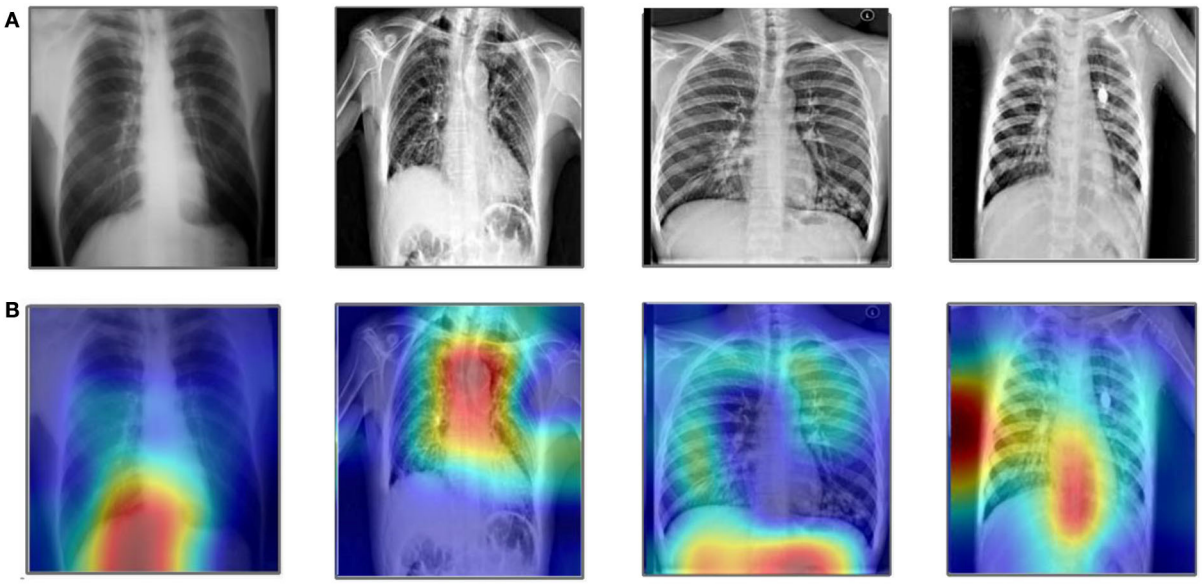


FIGURE 11
Sample images of Grad-Cam-based analysis. **(A)** Original images and **(B)** Grad-CAM analysis.

TABLE 12 Comparison of the proposed method to existing techniques.

Sr. No	References	Year	Method	Accuracy (%)
1	(47)	2020	Detection of COVID19 infection by using deep features and Bayesian optimization	98.97
2	(48)	2022	COVID19 diagnosis using CNN architected and Bayesian Optimization	96.29
3	(46)	2022	Diagnosis of Corona virus on CT images using bayes optimized DCNN and ML	99.3
Proposed			DCNN Bayesian optimization with slicing	98.8
			Serial fusion method	97.9
				99.4

Bold represent best values.

and Bayesian optimization concepts. Recently, the maximum accuracy has reached 99.3% by (46). In contrast, the suggested framework obtained a high degree of accuracy, as shown in Table 12. This shows the improvement of the proposed method.

Conclusion

This article presents an automated COVID-19 classification technique based on the hyperparameter optimization of pre-trained deep learning models *via* BO. Initially, contrast is increased to improve the visual quality of the input images, which are later used to train selected pre-trained models. Transfer learning is used to fine-tune and train both models. BO was used to optimize the hyperparameters of selected pre-trained models during training. Following that, features are extracted and fused using a proposed slicing-based approach. Three publicly available datasets were used in the experiment, and the accuracy was higher than with previous techniques. Based on the findings, we concluded that the proposed contrast-enhanced approach improved training capability, allowing for the later extraction of important features. Furthermore, the BO-based hyperparameters selection trained selected models more effectively than static initialization. Furthermore, the proposed fusion method improved classification accuracy. The computational time of the classification accuracy that was increased after the fusion process is the work's limitation. In future, feature selection methods will be prioritized in order to reduce the dimension of fused data.

Data availability statement

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

Author contributions

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

Funding

This work was supported by the Human Resources Program in Energy Technology of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resources from the Ministry of Trade, Industry & Energy, Republic of Korea (No. 20204010600090).

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for supporting this work through Research Groups Project under Grant No. RGP.2/16/43. The authors also like to thank HITEC University.

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.

References

1. Alshazly H, Linse C, Barth E, Martinetz T. Explainable COVID-19 detection using chest CT scans and deep learning. *Sensors.* (2021) 21:455. doi: 10.3390/s21020455

2. Shui-Hua W, Khan MA, Govindaraj V, Fernandes SL, Zhu Z, Yu-Dong Z. Deep rank-based average pooling network for COVID-19 recognition. *Comp Mater Cont.* (2022) 2797–2813. doi: 10.32604/cmc.2022.020140

3. Bhuyan HK, Chakraborty C, Shelke Y, Pani SK. COVID-19 diagnosis system by deep learning approaches. *Expert Syst.* (2022) 39:e12776. doi: 10.1111/exsy.12776

4. Subramanian N, Elharrouss O, Al-Maadeed S, Chowdhury M. A review of deep learning-based detection methods for COVID-19. *Comput Biol Med.* (2022) 143:105233. doi: 10.1016/j.compbiomed.2022.105233

5. Kogilavani SV, Prabhu J, Sandhiya R, Kumar MS, Subramaniam U, Karthick A, et al. COVID-19 detection based on lung Ct scan using deep learning techniques. *Comput Math Methods Med.* (2022) 2022:7672196. doi: 10.1155/2022/7672196

6. Krull A, Vičar T, Prakash M, Lalit M, Jug F. Probabilistic Noise2Void: unsupervised content-aware denoising. *Front Comput Sci.* (2020) 2:5. doi: 10.3389/fcomp.2020.00005

7. Jahanshahi AA, Dinani MM, Madavani AN Li J, Zhang SX. The distress of Iranian adults during the Covid-19 pandemic - More distressed than the Chinese and with different predictors. *Brain Behav Immun.* (2020) 87:124–5. doi: 10.1016/j.bbi.2020.04.081

8. World Health Organization. *Weekly Epidemiological Update on COVID-19.* (2021). Available online at: https://www.who.int/; https://www.who.int/publications/m (accessed September 14, 2022).

9. Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology.* (2020) 295:202–7. doi: 10.1148/radiol.2020200230

10. Alshazly H, Linse C, Abdalla M, Barth E, Martinetz T. COVID-Nets: deep CNN architectures for detecting COVID-19 using chest CT scans. *PeerJ Comp Sci.* (2021) 7:e655. doi: 10.7717/peerj-cs.655

11. Zhang J, Xie Y, Pang G, Liao Z, Verjans J, Li W, et al. Viral pneumonia screening on chest X-ray images using confidence-aware anomaly detection. *IEEE Trans Med Imaging.* (2020) 40:879–90. doi: 10.1109/TMI.2020.3040950

12. Oh Y, Park S, Chul Ye K. Deep learning COVID-19 features on CXR using limited training data sets. *IEEE Trans Med Imaging.* (2020) 39:2688–700. doi: 10.1109/TMI.2020.2993291

13. Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging.* (2016) 35:1285–98. doi: 10.1109/TMI.2016.2528162

14. Nogues I, Yao J, Mollura D, Summers RM. *Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning.* IEEE (2016).

15. Kini AS, Gopal Reddy AN, Kaur M, Satheesh S, Singh J, Martinetz T, et al. Ensemble deep learning and internet of things-based automated COVID-19 diagnosis framework. *Contrast Media Mol Imaging* (2022). doi: 10.1155/2022/7377502

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

17. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436–44. doi: 10.1038/nature14539

18. Lakhani P, Sundaram BJR. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology.* (2017) 284:574–82. doi: 10.1148/radiol.2017162326

19. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2016). p. 770–8.

20. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. *Int J Comput Vis.* (2015) 115:211–52. doi: 10.1007/s11263-015-0816-y

21. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2016). p. 2818–26.

22. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. *arXiv.* (2016) 1–13. doi: 10.48550/arXiv.1602.07360

23. Hamza A, Attique Khan M, Wang S, Alqahtani A, Alsubai S, Binbusayyis A, et al. COVID19 Classification using Chest X-Ray images: a framework of CNN-LSTM and improved max value moth flame optimization. *Front Public Health.* 2795:1–20. doi: 10.3389/fpubh.2022.948205

24. Syed HH, Khan MA, Tariq U, Armghan A, Alenezi F, Khan JA, et al.. A rapid artificial intelligence-based computer-aided diagnosis system for COVID-19 classification from CT images. *Behav Neurol.* (2021). doi: 10.1155/2021/2560388

25. Loey M, El-Sappagh S, Mirjalili S. Bayesian-based optimized deep learning model to detect COVID-19 patients using chest X-ray image data. *Comput Biol Med.* (2022) 142:05213. doi: 10.1016/j.compbiomed.2022.105213

26. Yoo SH, Geng H, Chiu TL Yu SK, Cho DC, Heo J, et al. Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging. *Front Med.* (2020) 7:427. doi: 10.3389/fmed.2020.00427

27. Wang D, Mo J, Zhou G, Xu L, Liu Y. An efficient mixture of deep and machine learning models for COVID-19 diagnosis in chest X-ray images. *PLoS ONE.* (2020) 15:e0242535. doi: 10.1371/journal.pone.0242535

28. Chowdhury NK, Rahman MM, Kabir MA. PDCOVIDNet: a parallel-dilated convolutional neural network architecture for detecting COVID-19 from chest X-ray images. *Health Inf Sci Syst.* (2020) 8:27. doi: 10.1007/s13755-020-00119-3

29. Khan IU, Aslam N. A deep-learning-based framework for automated diagnosis of COVID-19 using X-ray Images. *Information.* (2020) 11:419. doi: 10.3390/info11090419

30. Che Azemin MZ, Hassan R, Mohd Tamrin MI, Md Ali AM. COVID-19 deep learning prediction model using publicly available radiologist-adjudicated chest X-ray images as training data: preliminary findings. *Int J Biomed Imaging.* (2020) 2020:8828855. doi: 10.1155/2020/8828855

31. Khan MA, Azhar M, Ibrar K, Alqahtani A, Alsubai S, Binbusayyis A, et al. COVID-19 classification from chest X-ray images: a framework of deep explainable artificial intelligence. *Comp Intell Neurosci.* (2022). doi: 10.1155/2022/4254631

32. Ravi V, Narasimhan H, Chakraborty C, Pham TD. Deep learning-based meta-classifier approach for COVID-19 classification using CT scan and chest X-ray images. *Multimedia Syst.* (2022) 28:1401–15. doi: 10.1007/s00530-021-00826-1

33. De Sousa PM, Carneiro PC, Oliveira MM, Pereira GM, da Costa Junior CA, de Moura LV, et al. COVID-19 classification in X-ray chest images using a new convolutional neural network: CNN-COVID. *Res Biomed Eng.* (2022) 38:87–97. doi: 10.1007/s42600-020-00120-5

34. Mohammed MA, Al-Khateeb B, Yousif M, Mostafa SA, Kadry S, Abdulkareem KH, et al.. Novel crow swarm optimization algorithm and selection approach for optimal deep learning COVID-19 diagnostic model. *Comp Intell Neurosci.* (2022). doi: 10.1155/2022/1307944

35. Allioui H, Mohammed MA, Benameur N, Al-Khateeb B, Abdulkareem KH, Garcia-Zapirain B, et al. A multi-agent deep reinforcement learning approach for enhancement of COVID-19 CT image segmentation. *J Pers Med.* (2022) 12:309. doi: 10.3390/jpm12020309

36. Khan MA, Alhaisoni M, Tariq U, Hussain N, Majid A, Damaševičius R, et al. COVID-19 case recognition from chest CT images by deep learning, entropy-controlled firefly optimization, and parallel feature fusion. *Sensors.* (2021) 21:7286. doi: 10.3390/s21217286

37. Kaur R, Kaur S. Comparison of contrast enhancement techniques for medical image. In: *2016 Conference on Emerging Devices and Smart Systems (ICEDSS).* (2016). p. 155—9.

38. Atila Ü, Uçar M, Akyol K, Uçar E. Plant leaf disease classification using EfficientNet deep learning model. *Ecol Inform.* (2021) 61:101182. doi: 10.1016/j.ecoinf.2020.101182

39. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2018). p. 4510–20.

40. Dong K, Zhou C, Ruan Y, Li Y. MobileNetV2 model for image classification. In: *2020 2nd International Conference on Information Technology and Computer Application (ITCA).* (2020). p. 476–80.

41. Victoria AH, Maragatham G. Automatic tuning of hyperparameters using Bayesian optimization. *Evol Syst.* (2021) 12:217–23. doi: 10.1007/s12530-020-09345-2

42. Joy TT, Rana S, Gupta S, Venkatesh S. Hyperparameter tuning for big data using Bayesian optimisation. In: *2016 23rd International Conference on Pattern Recognition (ICPR).* (2016). p. 2574—9.

43. Abbasimehr H, Paki R. Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization. *Chao, Solit Fract.* (2021) 142:110511. doi: 10.1016/j.chaos.2020.110511

44. Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. *Adv Neural Inf Process Syst.* (2012) 25:1–16.

45. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2016). p. 2921—9.

46. Canayaz M, Sehribanoglu S, Özdag R, Demir M. COVID-19 diagnosis on CT images with Bayes optimization-based deep neural networks and machine learning algorithms. *Neural Comp Appl.* (2022) 34:5349–65. doi: 10.1007/s00521-022-07052-4

47. Nour M, Cömert Z, Polat K. A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization. *Appl Soft Comput.* (2020) 97:106580. doi: 10.1016/j.asoc.2020.106580

48. Aslan MF, Sabanci K, Durdu A, Unlersen MF. COVID-19 diagnosis using state-of-the-art CNN architecture features and Bayesian Optimization. *Comp Biol Med.* (2022) 105244. doi: 10.1016/j.compbiomed.2022.105244

frontiers | Frontiers in Public Health

TYPE Opinion
PUBLISHED 01 December 2022
DOI 10.3389/fpubh.2022.1076472

OPEN ACCESS

EDITED BY
Steven Fernandes,
Creighton University, United States

REVIEWED BY
Belfin Robinson,
University of North Carolina at Chapel Hill, United States

*CORRESPONDENCE
J. Anitha
anitha_j@karunya.edu

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 21 October 2022
ACCEPTED 14 November 2022
PUBLISHED 01 December 2022

CITATION
Francis A, Pandian IA and Anitha J (2022) A boon to aged society: Early diagnosis of Alzheimer's disease–An opinion.
Front. Public Health 10:1076472.
doi: 10.3389/fpubh.2022.1076472

A boon to aged society: Early diagnosis of Alzheimer's disease–An opinion

Ambily Francis[1,2], Immanuel Alex Pandian[1] and J. Anitha[3]*

[1]Department of Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India, [2]Department of Electronics and Communication Engineering, Sahrdaya College of Engineering and Technology, Kodakara, India, [3]Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, India

KEYWORDS

deep learning, pre-trained network, mild cognitive impairment (MCI), features, cognitively normal (CN), Alzheimer's disease

Introduction

Alzheimer's disease is a highly terrible condition for both the victims and their loved ones to endure. It is a degenerative neurological condition that steadily worsens until the patient is no longer able to do daily tasks. Even though there is no complete cure for the illness, several medications can slow its progression. After the age of 60, every person's brain, notably the hippocampus, begins to shrink (1, 2). But, for some unknown reasons, the shrinking rate will be higher for some people. In this context, the four cognitive stages of the human brain are classified as cognitively normal (CN), mild cognitive impairment convertible (MCIc), mild cognitive impairment non-convertible (MCInc), and Alzheimer's disease (AD) (3, 4). The stages are differentiated in terms of the shrinkage rate of the hippocampus. The hippocampus frequently shrinks as people age. Alzheimer's disease is clearly shown by the hippocampus's unusually rapid rate of atrophy. Cerebrospinal fluid (CSF) will occupy the gray matter area as the hippocampus shrinks. As a result, gray matter falls short of the level needed to maintain proper function. Therefore, the increasing CSF volume is yet another sign of Alzheimer's disease. The healthy brain and Alzheimer's disease affected brain are given in Figure 1 (*Credit: Adapted from illustration by Stacy Jannis/Alzheimer's Association, Link: https://www.alz.org/alzheimers-dementia/what-is-alzheimers/brain_tour/credits). The main determinants for classifying various stages of cognition are shrinkage of the hippocampus, growth of the brain's ventricles, and the change from gray matter to CSF (5, 6). The normal aging stage that will not lead to AD is known as mild cognitive impairment non-convertible stage. In this stage memory problems due to normal aging can be identified. However, Alzheimer's disease will never develop from this stage. The early stage of Alzheimer's disease with a significant rate of brain shrinking is known as the mild cognitive impairment convertible stage. The patients with mild cognitive impairment convertible stage will become Alzheimer's patients within years. In the field of early detection of AD, many algorithms address the classification of AD and CN. However, if the system accurately distinguishes between MCIc and MCInc, early detection of AD can be asserted. So the classification between MCIc and MCInc deserves more attention in this field.

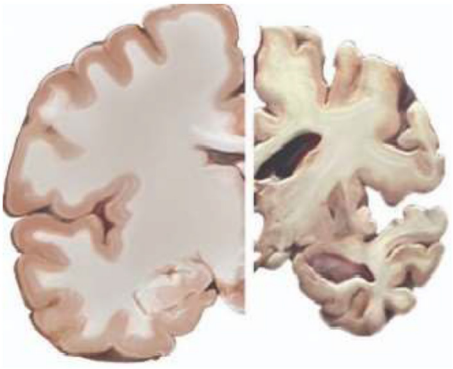


FIGURE 1
Healthy brain **(Left)** and Alzheimer's affected brain **(Right)**[*].

Clinical practices of early diagnosis of Alzheimer's disease

To better comprehend the aberrant state of the human brain, doctors are running clinical trials. Clinical diagnoses are made using anecdotal evidence. The most often used technique for making a clinical diagnosis is the mini-mental state examination. Patients who are undergoing a mini-mental state assessment fill out a questionnaire designed to reflect any brain disorders, such as Alzheimer's disease. For the accurate diagnosis of disease, many conversations with clinicians are necessary.

For the screening, diagnosis, and subsequent management of Alzheimer's disease patients, doctors must be able to quickly and accurately identify the symptoms and pathology of the disease that are associated with Alzheimer's disease. Additionally, it enables patients and others who are caring for them to make necessary lifestyle modifications that may prolong the maintenance of their quality of life. Unfortunately, identifying Alzheimer's disease in its early stages in clinical practice can be difficult and is hampered by several obstacles, including time restrictions on clinicians, difficulty accurately diagnosing Alzheimer's pathology, and the fact that patients and healthcare professionals frequently write off symptoms as being a normal part of aging (7).

Computer based early diagnosis of Alzheimer's disease

Clinical exams and MRI scan reports may not be able to detect the early stage of AD in patients. If the early illness indications are discovered before the victim notices any symptoms, early detection of Alzheimer's disease gives a great possibility of prompt treatment. The researchers should integrate and evaluate the necessary data for the crucial improvement of medical diagnostic systems. Numerous contributions by researchers addressed the challenges related to the early detection of AD. In recent studies, MRI images are evaluated and processed to find signs of Alzheimer's disease in its early stages (8, 9). Doctors can use the successfully developed system as a clear channel of communication to identify AD in its early stages.

Machine learning techniques, both supervised and unsupervised, are applied to the analysis of medical images. Using supervised learning algorithms, labeled data is categorized using domain knowledge from an expert and pertinent feature data. The unsupervised learning algorithms work on unlabeled data. The supervised and unsupervised deep learning models are capable of automatically creating feature extractors and extracting discriminative features from train data. The rest of this section is structured as follows. Early diagnosis of AD using conventional feature extraction methods are deliberated in section Early diagnosis based on conventional feature extraction methods. Section Early diagnosis of Alzheimer's disease based on deep learning techniques describes related works on early diagnosis of AD using deep learning techniques. In section Early diagnosis of Alzheimer's disease based on pre-trained models, related works for early diagnosis of AD based on pre-trained models are discussed. The basic working flow of three computer-based methods for the early diagnosis of Alzheimer's disease are portrayed in Figure 2.

Early diagnosis based on conventional feature extraction methods

Feature descriptors play an important role in medical image analysis. Medical image processing applications that explicitly use feature descriptors include disease diagnosis (10, 11), medical joint photographic experts group image steganography (12), object recognition, and segmentation. The two types of feature descriptors are global feature descriptors and local feature descriptors. Global features are those that apply to the entire image. Local features are provided by patches of an image. For any application involving the analysis of medical images using conventional feature extraction methods, global features are insufficient as they cannot provide the spatial information of images. For early Alzheimer's disease identification, many works suggested novel local feature descriptors.

Texture features are highly important low-level features that give significant details about specific regions of medical images. The pixel intensities of image patches are used to obtain the texture features. Numerous texture descriptors, including the local binary pattern, the scale-invariant feature transform, and the speed- up robust features, are used to interpret MRI images. The well-known descriptor, local binary pattern (LBP) is proposed for describing texture qualities (13).

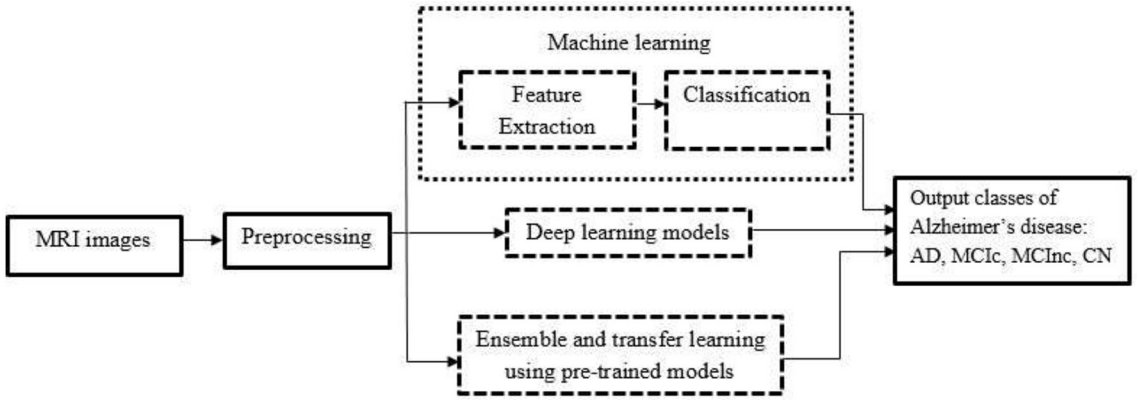


FIGURE 2
Basic working flow of computer-based methods.

A high classification accuracy is achieved at the expense of a significant calculation time by using texture features taken from an elliptical neighborhood (14). Two-dimensional and three-dimensional advanced local binary patterns are combined to obtain accurate multi-class Alzheimer's disease prediction (15). The computational time taken for the processing of high-dimension features is the major limitation of this method. The textural features are identified by considering the voxel neighbors of MRI image (16–19).

Scale-invariant feature transform (SIFT) is another interesting local feature descriptor. The scale invariance and rotation invariance are the important characteristics of SIFT. Both the frequency domain and the spatial domain are utilized for extracting SIFT features. SIFT is effectively used for the early detection of AD (20–23). According to the discussions in the aforementioned articles, the benefit is that SIFT did not alter with variations in illumination. However, the gradients of each voxel along the path must be calculated. Hence the computation must be regarded as intensive.

The Hessian matrix serves as the foundation for the Speeded Up Robust Features (SURF) which shares many characteristics with SIFT (24). SURF performs admirably in computer vision applications and has been researched for use in medical image applications (5, 25, 26). The primary feature of SURF is the enhanced speed that the integral filter produces. As previously stated, one key sign of AD is the increasing CSF volume in the hippocampus. CSF volume computation is used to carry out early identification of AD (5). The CSF volume computation is done based on the number of CSF voxels in the hippocampal area. Circular harmonic functions (CHF) offers great image description independent of scale, position, and illumination for the early diagnosis of AD (27, 28).

After preprocessing, MRI images are sampled with nine volumes of interest in brain regions related to AD. Intensity and texture features are extracted from the interest of volume. The support vector machine classifier trains on the features. Non-linear registration systems can be made better even when the method offers tolerable accuracy (29). Clinical and texture characteristics are integrated to identify the mild cognitive impairment convertible stage (MCIc). The key benefit is that MCI (Mild cognitive impairment) and AD have been classified using the entire brain's MRI texture with binary logistic regression (30). A feature selection technique that makes use of a multivariate general linear model is proposed for the MCIc vs. CN classification. The modest intensity fluctuations from CN to MCIc are produced with the use of a general linear model. Additionally, multivariate adaptive regression splines are utilized as a classifier (31).

The orientation surrounding each voxel of eight local regions of MRI image including white matter and gray matter is calculated using a 3D local directional pattern. The algorithm is less susceptible to illumination and noise (11). The features obtained from many modalities are combined using multi- kernel SVM. In this method hyper graph-based regularization is used for AD vs. MCI classification. The method provides superior classification accuracy than existing multi-modality strategies, according to the results. The algorithm's primary flaw is that all hyperedge weights are set to 1 without considering various hyperedges (32). It is possible to identify Alzheimer's disease more quickly by analyzing data sets of medical records using the machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, Naive Bayes and variants of support vector machine to identify Alzheimer's infection (33).

Early diagnosis of Alzheimer's disease based on deep learning techniques

In recent years, deep learning methods have been prominent in many medical sectors. For efficient and precise categorization, deep learning techniques will abstract important features from the images. The methods aid medical professionals in early disease diagnosis. These methods enhance researchers' abilities to engage in studies that examine medical image approaches for disease diagnosis. Shortly, deep learning may replace traditional medical diagnostic techniques because of its learning capabilities and cost-effectiveness. A sizable number of convolutional neural network (CNN) models are combined to successfully anticipate various stages of AD. The classification performance is good as a result of the integration of Bagging, Boosting, and Random Forest algorithms (34). The review (35) listed and examined the most recent studies in the area of early Alzheimer's disease diagnosis using deep learning algorithms. Using 2D MRI slices, the work (36) proposes a CNN-based method for extracting discriminatory characteristics from structural MRI, intending to categorize Alzheimer's disease and healthy people.

The algorithm (37) uses functional MRI images in addition to medical data like age, gender, and genetic information. Using stacked autoencoders and functional MRI time-series data or correlation coefficient data, the deep neural network is trained. The research (38) reviews a number of deep learning based algorithms used in the early identification of Alzheimer's disease. The approach also offers a flexible foundation for repeatable testing. The strengths of hessian matrix and local binary pattern are utilized with the use of convolutional neural network (4). The algorithm provides best classification accuracy for AD vs. CN. But the Classification accuracy of MCIc vs. MCInc can be improved.

Early diagnosis of Alzheimer's disease based on pre-trained models

Transfer learning and ensemble learning are typically expressed in computer vision by employing pre-trained models. The innovative approaches for dealing with the issue of training data and test data having different distributions include transfer learning and ensemble learning. Pre-trained models can be used as the foundation for new models for the early diagnosis of AD. Using a neural network model, transfer learning is utilized to improve the precision and computing efficiency of image categorization. In transfer learning models, pre-trained neural networks are used as feature extractors, and the output of the pre-trained network is given to a new classification layer that is trained on data particular to the task. By integrating many previously trained models with new classification layers, ensemble transfer learning models (39, 40) are utilized to increase performance at the cost of increased model complexity. Without much computational complexity, the squeeze and excitation network affects the channel attention process by strengthening and weakening each feature channel. The algorithm (41) illustrates that non-biomedical pre-trained models, such as ResNet, learn cross-domain characteristics that enable the model to extract critical low-level properties from MRI scans to improve classification accuracy. The suggested method guarantees effective data augmentation before learning. The augmentation helps to regularize the model.

A 3D multi-channel feature maps based on Voxception-Resnet is created for the classification of AD and CN (42). Data augmentation is done before feature map creation. The dataset used in this algorithm is diffusion MRI images. The work (43) employs the VGG-16 pre-trained model, a non-biomedical model, to learn cross-domain characteristics and boost accuracy. The method achieves exceptional three-class classification accuracy and offers a mathematical model based on VGG-16 transfer learning. The method (44) uses 3D DenseNet to learn both hippocampus segmentation and classification features that based on deep CNN. Alzheimer's checking web application is proposed based on transfer learning of pre-trained models such as VGG19 (45).

Discussions

As discussed in the previous sections, AD develops gradually that takes years for symptoms to physically manifest in a patient, thus clinical approaches for an early diagnosis of the disease are insufficient. Physical symptoms of AD in its early stages will resemble those of a typical aged person in many ways. Clinical approaches are not very reliable for separating mild cognitive impairment into convertible and non-convertible stages in the context of early diagnosis.Early diagnosis is made using MRI scans using image processing techniques. The early diagnosis of AD utilizing MRI and Positron Emission Tomography (PET) image modalities is made possible by numerous algorithms (46, 47). Algorithms that need manual feature extraction to take a lot of time and have high computational complexity. Additionally, hardware implementation of such a system necessitates an extremely complicated system. Also, these kinds of systems only provide very poor MCIc vs. MCInc classification accuracy. The distinction between phases MCIc and MCInc is very difficult because of very slight voxel changes. Thus, the difference between MCIc and MCInc images may be described by a high-end feature description.

For the early diagnosis of AD, deep learning algorithms play a significant role. Numerous algorithms are developed for deep learning-based early diagnosis. Reduced computational complexity, less calculation time, low dimension features and good differentiation capability between MCIc vs. MCInc should be the goal

of the computer-based diagnostic system. The pre-trained models successfully adapt the MRI images to predict the early phase of AD (48). The high-dimension features of pre-trained models cause complexity in the physical implementations later. But with pre-trained models, MCIc vs. MCInc classification accuracy is fairly good at the cost of computational complexity, processing time, and high dimension features.

Many investigations tried to classify AD vs. CN, and AD vs. CN vs. MCI (not specifically MCIc or MCInc). But only less focus is given to the classification of MCI vs. MCInc. Even though early Alzheimer's disease detection has been the subject of countless research, it is still challenging to identify the specific traits that can detect the disease in its earliest stages (49). The most salient classification for the early identification of AD is MCIc vs. MCInc. The objective of the upcoming efforts is to increase the classification precision of MCIc vs. MCInc. Due to extremely slight voxel changes, the differentiation between phases of MCIc and MCInc is exceedingly laborious. According to the published researches, MCIc vs. MCInc categorization accuracy ranges from 55 to 75% (38, 50, 51). Further study is necessary due to the low categorization accuracy for MCIc vs. MCInc. Unique learning strategies that discriminate between mild cognitive impairment convertible and nonconvertible stages will accelerate the classification accuracy of MCIc vs. MCInc. The reliable, fully automated system for the early diagnosis of AD can be a boon to the aged society in the near future.

Author contributions

AF, IP, and JA contributed to the study conception and design, literature review, interpretation, and manuscript preparation. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors express their sincere gratitude to the Karunya Institute of Technology and Sciences for the research facilities that have been extended.

Conflict of interest

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

Publisher's note

References

1. Bethlehem RA, Seidlitz J, White SR, Vogel JW, Anderson KM, Adamson C, et al. Brain charts for the human lifespan. *Nature.* (2022) 604:525–33. doi: 10.1038/s41586-022-04554-y

2. Harris TC, de Rooij R, Kuhl E. The shrinking brain: cerebral atrophy following traumatic brain injury. *Ann Biomed Eng.* (2019) 47:1941–59. doi: 10.1007/s10439-018-02148-2

3. Salvatore C, Cerasa A, Battista P, Gilardi MC, Quattrone A, Castiglioni I, et al. Magnetic resonance imaging biomarkers for the early diagnosis of alzheimer's disease: a machine learning approach. *Front Neurosci.* (2015) 9:307. doi: 10.3389/fnins.2015.00307

4. Francis A, Pandian IA. Early detection of alzheimer's disease using local binary pattern and convolutional neural network. *Multimed Tools Appl.* (2021) 80:29585–600. doi: 10.1007/s11042-021-11161-y

5. Ben Ahmed O, Benois-Pineau J, Allard M, Ben Amar C, Catheline G. Classification of alzheimer's disease subjects from mri using hippocampal visual features. *Multimed Tools Appl.* (2015) 74:1249–66. doi: 10.1007/s11042-014-2123-y

6. Diogo VS, Ferreira HA, Prata D. Early diagnosis of alzheimer's disease using machine learning: a multi-diagnostic, generalizable approach. *Alzheimer's Res.* (2022) 14:1–21. doi: 10.1186/s13195-022-01047-y

7. Porsteinsson A, Isaacson R, Knox S, Sabbagh M, Rubino I. Diagnosis of early alzheimer's disease: Clinical practice in 2021. *J Prevent Alzheimer's Dis.* (2021) 8:371–86. doi: 10.14283/jpad.2021.23

8. Singh N, Soni N, Kapoor A, et al. Automated detection of alzheimer disease using mri images and deep neural networks-a review. *arXiv.* arXiv:2209.11282. (2022). doi: 10.48550/arXiv.2209.11282

9. Arafa DA, Moustafa HED, Ali-Eldin AM, Ali HA. Early detection of alzheimer's disease based on the state-of-the-art deep learning approach: a comprehensive survey. *Multimed Tools Appl.* (2022) 1–42. doi: 10.1007/s11042-022-11925-0

10. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med.* (2019) 25:65–9. doi: 10.1038/s41591-018-0268-3

11. Yan S, Song C, Zheng B. 3d local directional patterns for early diagnosis of Alzheimer's disease. *J Eng.* (2019) 2019:530–5. doi: 10.1049/joe.2018.9412

12. Liao X, Yin J, Guo S, Li X, Sangaiah AK. Medical jpeg image steganography based on preserving inter-block dependencies. *Comp Electrical Eng.* (2018) 67:320–9. doi: 10.1016/j.compeleceng.2017.08.020

13. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell.* (2002) 24:971–87. doi: 10.1109/TPAMI.2002.1017623

14. Nanni L, Lumini A, Brahnam S. Local binary patterns variants as texture descriptors for medical image analysis. *Artif Intell Med.* (2010) 49:117–25. doi: 10.1016/j.artmed.2010.02.006

15. Sarwinda D, Bustamam A. Detection of alzheimer's disease using advanced local binary pattern from hippocampus and whole brain of mr images. In: *2016 International Joint Conference on Neural Networks (IJCNN)* (IEEE) (2016) p. 5051–6. doi: 10.1109/IJCNN.2016.7727865

16. Oliver A, Llado' X, Freixenet J, Mart'i J. False positive reduction in mammographic mass detection using local binary patterns. In: *International*

Conference on Medical Image Computing and Computer-Assisted Intervention (Springer) (2007) p. 286–293. doi: 10.1007/978-3-540-75757-3_35

17. Unay D, Ekin A, Jasinschi R. Medical image search and retrieval using local binary patterns and klt feature points. In: *2008 15th IEEE International Conference on Image Processing* (IEEE) (2008) p. 997–1000. doi: 10.1109/ICIP.2008.4711925

18. Chang CW, Ho CC, Chen JH. Adhd classification by a texture analysis of anatomical brain mri data. *Front Syst Neurosci.* (2012) 6:66. doi: 10.3389/fnsys.2012.00066

19. Oppedal K, Engan K, Aarsland D, Beyer M, Tysnes OB, Eftestøl T. Using local binary pattern to classify dementia in MRI. In: *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)* 235 (IEEE) (2012) p. 594–7. doi: 10.1109/ISBI.2012.6235618

20. Yang M, Yuan Y, Li X, Yan P. Medical image segmentation using descriptive image features. *BMVC (Citeseer).* (2011) 1–11. doi: 10.5244/C.25.94

21. Castellani U, Rossato E, Murino V, Bellani M, Rambaldelli G, Perlini C, et al. Classification of schizophrenia using feature-based morphometry. *J Neural Transm.* (2012) 119:395–404. doi: 10.1007/s00702-011-0693-7

22. Mizotin M, Benois-Pineau J, Allard M, Catheline G. Feature-based brain mri retrieval for alzheimer disease diagnosis. In: *2012 19th IEEE International Conference on Image Processing* (IEEE) (2012) p. 1241–4. doi: 10.1109/ICIP.2012.6467091

23. Mondal P, Mukhopadhyay J, Sural S, Bhattacharyya PP. 3d-sift feature based brain atlas generation: An application to early diagnosis of alzheimer's disease. In: *2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)* (IEEE) (2014) p. 342–347. doi: 10.1109/MedCom.2014.7006030

24. Bay H, Tuytelaars T, Gool LV. Surf: Speeded up robust features. In: *European conference on computer vision* (Springer) (2006) p. 404–417. doi: 10.1007/11744023_32

25. Sargent D, Chen CI, Tsai CM, Wang YF, Koppel D. Feature detector and descriptor for medical images. *Med Imag 2009: Image Proc (SPIE).* (2009) 7259:991–8. doi: 10.1117/12.811210

26. Lecron F, Benjelloun M, Mahmoudi S. Descriptive image feature for object detection in medical images. In: *International Conference Image Analysis and Recognition* (Springer) (2012) p. 331–338. doi: 10.1007/978-3-642-31298-4_39

27. Sorgi L, Cimminiello N, Neri A. Keypoints selection in the gauss laguerre transformed domain. *BMVC.* (2006) 539–547. doi: 10.5244/C.20.56

28. Sorokin DV, Mizotin MM, Krylov AS. Gauss-laguerre keypoints extraction using fast hermite projection method. In: *International Conference Image Analysis and Recognition* (Springer) (2011) p. 284–293. doi: 10.1007/978-3-642-21593-3_29

29. Chincarini A, Bosco P, Gemme G, Esposito M, Rei L, Squarcia S, et al. Automatic temporal lobe atrophy assessment in prodromal ad: Data from the descripa study. *Alzheimer's & Dementia.* (2014) 10:456–67. doi: 10.1016/j.jalz.2013.05.1774

30. Luk CC, Ishaque A, Khan M, Ta D, Chenji S, Yang YH, et al. Alzheimer's disease: 3-dimensional mri texture for prediction of conversion from mild cognitive impairment. *Alzheimer's & Dementia.* (2018) 10:755–63. doi: 10.1016/j.dadm.2018.09.002

31. Çevik A, Weber GW, Eyüboğlu BM, Oguz KK. Voxel-mars: a method for early detection of alzheimer's disease by classification of structural brain mri. *Ann Operations Res.* (2017) 258:31–57. doi: 10.1007/s10479-017-2405-7

32. Shao W, Peng Y, Zu C, Wang M, Zhang D, Initiative ADN, et al. Hypergraph based multi-task feature selection for multimodal classification of alzheimer's disease. *Computerized Medical Imag Graphics.* (2020) 80:101663. doi: 10.1016/j.compmedimag.2019.101663

33. Kishore P, Kumari CU, Kumar M, Pavani T. Detection and analysis of alzheimer's disease using various machine learning algorithms. *Materials Today: proceedings.* (2021) 45:1502–8. doi: 10.1016/j.matpr.2020.07.645

34. Pan D, Zeng A, Jia L, Huang Y, Frizzell T, Song X. Early detection of alzheimer's disease using magnetic resonance imaging: a novel approach combining convolutional neural networks and ensemble learning. *Front Neurosci.* (2020) 14:259. doi: 10.3389/fnins.2020.00259

35. Al-Shoukry S, Rassem TH, Makbol NM. Alzheimer's diseases detection by using deep learning algorithms: a mini-review. *IEEE Access.* (2020) 8:77131–41. doi: 10.1109/ACCESS.2020.2989396

36. Al-Khuzaie FE, Bayat O, Duru AD. Diagnosis of alzheimer disease using 2d mri slices by convolutional neural network. *Appl Bionics Biomech.* (2021) 2021. doi: 10.1155/2021/6690539

37. Ju R, Hu C, Li Q, et al. Early diagnosis of alzheimer's disease based on resting-state brain networks and deep learning. *IEEE/ACM Trans. Comput. Biol. Bioinform.* (2017) 16:244–57. doi: 10.1109/TCBB.2017.2776910

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

39. Fathi S, Ahmadi M, Dehnad A. Early diagnosis of alzheimer's disease based on deep learning: a systematic review. *Comput Biol Med.* (2022) 105634. doi: 10.1016/j.compbiomed.2022.105634

40. Francis A, Pandian IA. Early detection of alzheimer's disease using ensemble of pre-trained models. In: *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)* (IEEE). (2021) p. 692–696. doi: 10.1109/ICAIS50930.2021.9395988

41. Valliani A, Soni A. Deep residual nets for improved alzheimer's diagnosis. In: *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics.* (2017) p. 615–615. doi: 10.1145/3107411.3108224

42. McCrackin L. Early detection of alzheimer's disease using deep learning. In: *Canadian Conference on Artificial Intelligence* (Springer) (2018) p. 355–359. doi: 10.1007/978-3-319-89656-4_40

43. Jain R, Jain N, Aggarwal A, Hemanth DJ. Convolutional neural network based alzheimer's disease classification from magnetic resonance brain images. *Cogn Syst Res.* (2019) 57:147–59. doi: 10.1016/j.cogsys.2018.12.015

44. Liu M, Li F, Yan H, Wang K, Ma Y, Shen L, et al. A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in alzheimer's disease. *Neuroimage.* (2020) 208:116459. doi: 10.1016/j.neuroimage.2019.116459

45. Helaly HA, Badawy M, Haikal AY. Deep learning approach for early detection of alzheimer's disease. *Cognitive Computat.* (2021) 1–17.

46. Lu D, Popuri K, Ding GW, Balachandar R, Beg MF. Multimodal and multiscale deep neural networks for the early diagnosis of alzheimer's disease using structural mr and fdg-pet images. *Sci Rep.* (2018) 8:1–13. doi: 10.1038/s41598-018-22871-z

47. Sherin A, Rajeswari R. Computer-aided diagnosis system for alzheimer's disease using positron emission tomography images. *Interdisciplinary Sci.* (2021) 13:433–42. doi: 10.1007/s12539-020-00409-0

48. Shanmugam JV, Duraisamy B, Simon BC, Bhaskaran P. Alzheimer's disease classification using pre-trained deep networks. *Biomed Signal Process Control.* (2022) 71:103217. doi: 10.1016/j.bspc.2021.103217

49. Kavitha C, Mani V, Srividhya S, Khalaf OI, Romero CAT. Early-stage alzheimer's disease prediction using machine learning models. *Front Public Health.* (2022) 10. doi: 10.3389/fpubh.2022.853294

50. Oh K, Chung YC, Kim KW, Kim WS, Oh IS. Classification and visualization of alzheimer's disease using volumetric convolutional neural network and transfer learning. *Sci Rep.* (2019) 9:1–16. doi: 10.1038/s41598-019-54548-6

51. Huang Y, Xu J, Zhou Y, Tong T, Zhuang X, (ADNI) ADNI. Diagnosis of alzheimer's disease via multi-modality 3d convolutional neural network. *Front Neurosci.* (2019) 13:509. doi: 10.3389/fnins.2019.00509

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 30 January 2023
DOI 10.3389/fpubh.2023.1109236

OPEN ACCESS

EDITED BY
Steven Fernandes,
Creighton University, United States

REVIEWED BY
Jabez Christopher,
Birla Institute of Technology and Science, India
Feras Alattar,
National University of Science and Technology, Oman
Kameswara Buragapu,
Gandhi Institute of Technology and Management (GITAM), India

*CORRESPONDENCE
Chuan-Yu Chang
✉ chuanyu@yuntech.edu.tw

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 27 November 2022
ACCEPTED 04 January 2023
PUBLISHED 30 January 2023

CITATION
Rajinikanth V, Vincent PMDR, Srinivasan K, Ananth Prabhu G and Chang C-Y (2023) A framework to distinguish healthy/cancer renal CT images using the fused deep features. *Front. Public Health* 11:1109236. doi: 10.3389/fpubh.2023.1109236

A framework to distinguish healthy/cancer renal CT images using the fused deep features

Venkatesan Rajinikanth[1], P. M. Durai Raj Vincent[2], Kathiravan Srinivasan[3], G. Ananth Prabhu[4] and Chuan-Yu Chang[5,6]*

[1]Division of Research and Innovation, Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS, Chennai, Tamil Nadu, India, [2]School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India, [3]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India, [4]Department of Computer Science Engineering, Sahyadri College of Engineering and Management, Mangaluru, India, [5]Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan, [6]Service Systems Technology Center, Industrial Technology Research Institute, Hsinchu, Taiwan

Introduction: Cancer happening rates in humankind are gradually rising due to a variety of reasons, and sensible detection and management are essential to decrease the disease rates. The kidney is one of the vital organs in human physiology, and cancer in the kidney is a medical emergency and needs accurate diagnosis and well-organized management.

Methods: The proposed work aims to develop a framework to classify renal computed tomography (CT) images into healthy/cancer classes using pre-trained deep-learning schemes. To improve the detection accuracy, this work suggests a threshold filter-based pre-processing scheme, which helps in removing the artefact in the CT slices to achieve better detection. The various stages of this scheme involve: (i) Image collection, resizing, and artefact removal, (ii) Deep features extraction, (iii) Feature reduction and fusion, and (iv) Binary classification using five-fold cross-validation.

Results and discussion: This experimental investigation is executed separately for: (i) CT slices with the artefact and (ii) CT slices without the artefact. As a result of the experimental outcome of this study, the K-Nearest Neighbor (KNN) classifier is able to achieve 100% detection accuracy by using the pre-processed CT slices. Therefore, this scheme can be considered for the purpose of examining clinical grade renal CT images, as it is clinically significant.

KEYWORDS

kidney cancer, renal CT slices, deep learning, KNN classifier, validation

1. Introduction

It is becoming increasingly apparent that infectious and acute syndromes are rising worldwide. Appropriate clinical procedures are necessary for detecting and treating these diseases as early as possible. Untreated diseases will likely result in various problems, including death, and they may also burden the healthcare system substantially. It should be noted that acute diseases are usually more severe than infectious diseases. Compared with infectious diseases, acute diseases will also lead to death in individuals. According to the current literature, cancer is a severe acute disease that accounts for a substantial number of deaths worldwide and has been identified as a disease that causes many deaths as well (1–3).

A report published by the World Health Organization (WHO) in 2020 shows that cancer was the leading cause of death worldwide in 2020 and is expected to continue in that way.[1] Several studies have indicated that, in the year 2020, approximately 10 million people will have died worldwide from various cancer-related causes, including cancer of the internal and external organs. According to this report, lung and colon cancer are the leading causes of death worldwide.

The Global Cancer Observatory (GCO) report lists several cancer cases in various body organs.[2] This report lists cancer in organs based on its occurrence rate, cancer in the kidney is listed as the 14th most dangerous disease, and untreated renal cancer will lead to death. This report also confirms that, in 2020, the number of cancer patients increased to 431,288. This report also confirms that nearly 430,000 new cases will be diagnosed in 2020 alone. According to the disease prediction by GCO, kidney (renal) cancer is severe, and its occurrence rate is gradually rising due to various causes. Early recognition and management are compulsory to cure the disease completely using appropriate medications. Kidney cancer (KC) is commonly assessed by automatic methods using a chosen medical imaging dataset (renal CT images), and the achieved results are analyzed and recorded for further investigation.

The earlier studies in the literature confirm that renal CT (RCT)-based kidney detection is a recommended procedure to precisely detect kidney abnormality during the disease screening process. Usually, the RCT is collected as a three-dimensional (3D) image, and then, a 3D to 2D conversion is employed to reduce the computation complexity during the RCT analysis (4, 5). The axial-plane 2D slices are commonly adopted in the literature, and it helps to provide the necessary information about abdominal conditions, including kidney health. Hence, this study also considered the axial-plane 2D RCT slices to examine the KC. Before implementing the detection task, every image is resized to a recommended dimension.

The ultimate task of this investigation is to prepare a disease detection structure to accurately identify the KC using the RCT images with the help of the chosen deep-learning scheme. To achieve better detection accuracy, this study implemented a preprocessing image procedure to treat the raw renal CT using a threshold filter approach discussed in earlier research. In the earlier studies, this arrangement is considered to strip the skull region from the brain MRI slices (6, 7) and to remove the artifact in lung CT slices (8–10). A similar procedure is adopted in this study to remove the artifact in RCT slices to improve the visibility of the kidney section. The proposed cancer detection framework consists of the following phases:

i. Image resizing and artifact removal using threshold filter.
ii. Deep feature extraction using chosen pre-trained methods.
iii. Dual-deep feature generation using serially concatenated deep features.
iv. Binary classification and verification using a 5-fold cross-validation.

The merit of the computerized scheme depends on its explainability and robustness, and hence, this study considered a framework that is very simple and robust (11). This scheme considered MATLAB for initial image processing, and the developed framework is implemented using PYTHON. The experimental exploration is separately implemented using (i) RCT with the artifact and (ii) RCT without the artifact, and the achieved performance values are compared. This approves that the classification accuracy realized with the artifact-removed RCT is better than the raw RCT. Furthermore, this study employs pre-trained schemes, such as VGG16, VGG19, ResNet50, ResNet101, DenseNet121, and DenseNet201, to obtain better detection in the considered task. The results authorize that the outcome achieved with VGG19 and DenseNet121 is better for the chosen RCT, and hence, the proposed scheme is implemented using deep features of (i) VGG19 and (ii) DenseNet121 and serially concatenated features of VGG19 and DenseNet121 after a 50% dropout. The deep feature-based classification helps accomplish an accuracy of 100% with the RCT without the artifact. This confirms that the proposed framework is clinically noteworthy and can be considered to identify the KC from the RCT collected from actual patients.

The key contributions of this framework include the following:

i. Threshold filter-supported preprocessing is executed to eliminate artifacts in RCT.
ii. Implementation of the proven deep-learning schemes to detect the KC using RCT.
iii. Implementation of serially concatenated deep features to enhance the KC detection accuracy.

This study is divided into the following sections: Section 2 presents the context, Section 3 illustrates the methodology, and Sections 4 and 5 discuss the results and conclusions.

2. Related studies

Computerized disease screening and diagnosis is one of the recent advancements, adopted in a variety of hospitals and disease screening laboratories to reduce the diagnostic burden of doctors and lab technicians. The increased disease occurrence rates need a faster and more accurate system to detect the disease using chosen medical data. The bio-image-supported disease screening is a common and widely adopted procedure to verify the condition of the internal organs. Furthermore, the bio-image-supported methods support accurate disease information compared with other medical modalities, and hence, these methods are widely employed to screen patients suffering due to cancer.

Kidney cancer is one of the acute diseases and ranked 14th based on the year 2020 reports of the WHO and GCO. Appropriate diagnosis and treatment will help the patient to recover from the disease. Due to its importance, a number of computerized schemes are discussed by the researchers to distinguish the KC using RCT pictures. Table 1 summarizes a few chosen KC detection procedures found in the literature.

Along with the above-considered studies, the research by Abdelrahman and Viriri (21) presents a detailed survey on traditional and deep-learning segmentation of the abnormal fragment in the kidney in RCT images. The research by Wang et al. (22) also presents a thorough evaluation of the deep-learning-supported scheme for biomedical image examination, including the RCT. These studies authorize the need for a well-organized methodology to detect

1 https://www.who.int/health-topics/cancer

2 https://gco.iarc.fr/

TABLE 1 Summary of the renal CT image examination methods.

References	Procedure implemented	Outcome
Alzu'bi et al. (12)	This study presents a new database of RCT images, which has been created using VGG16 code and ResNet50 support, which is used to detect the KC	97% accuracy
Xu et al. (13)	As a result of the implementation of ResNet50 and ResNet101, the cropped RCT images have been classified into the following two categories: healthy and cancerous	>82% accuracy
Amiri et a l. (14)	The execution of the machine learning scheme with a radiomics feature is discussed in order to detect kidney abnormalities by using RCT slices to perform the machine learning	94% accuracy
Miskin et al. (15)	With the application of machine learning techniques based on the cropped RCT images, the detection of benign and malignant cystic renal masses can be accomplished	93% specificity
Shehata et al. (16)	An innovative computer-assisted diagnosis system is proposed for examining kidney cancer in cropped RCT slices using a novel comprehensive renal cancer computer-assisted diagnosis scheme	89.6% accuracy
Nikpanah et al. (17)	A deep-learning-supported technique based on multi-phasic MRI is presented as an example of the execution of the technique for detecting the clear cell renal cell carcinoma	81% accuracy
Heller et al. (18)	With the use of the KiTS19 challenge benchmark 3D RCT images, we are able to segment the abnormal kidney region using 3D U-Net	Dice value of 97.4 and 0.85.1% is achieved for kidney and tumor, respectively
Bhandari et al. (19)	The present study discusses the detection of low/high-grade renal cell cancers from RCT images in detail	82 to 96% Area Under Curve (AUC) is present.
Islam et al. (20)	The work presented here implemented the VGG16, InceptionV3, and resNet50 using RCT slices to detect kidney abnormalities in multi-classes, with the VGG16 presenting a better detection metric compared to the others	98.2% accuracy

abnormality from the chosen medical image. Hence, in this research, a framework based on deep learning is proposed to detect the KC from the axial-plane RCT slices accurately.

3. Methodology

Using a binary classifier, this research division demonstrates how RCT slices are classified into healthy and cancerous classes in an axial plane. When the patient visits the nephrologist to verify the condition of the kidney, a recommended clinical protocol will be followed by the doctor to examine the kidney and its condition, and based on the observations/disease symptoms, the nephrologist recommends a bio-imaging-based examination to get the complete information about the kidney. When the patient undergoes a CT scan, it will provide a 3D picture of the abdominal region, which is then converted into 2D to reduce the computational complexity. Furthermore, the personal verification of the kidney section from the bio-image needs a 2D picture printed on a specialized film. A similar procedure is executed when a computer-supported diagnosis is implemented.

From the data collection to the decision-making process, the proposed scheme is depicted in Figure 1. A number of procedures are involved in the proposed scheme, including image collection and preprocessing for improved detection accuracy, feature extraction utilizing a selected deep-learning technique, feature reduction, and serial feature concatenation to produce the fused feature vector, binary classification *via* 5-fold cross-validation, and verification of the proposed scheme's performance on the basis of the results obtained. In this study, the fused feature vector is constructed by integrating the deep features of VGG19 and DenseNet121. In addition, based on the computation of performance measures, the merit of the proposed scheme is confirmed based on the evaluation of these features to determine the classification performance of SoftMax and other binary classifiers.

3.1. Image database

This study considered the axial-plane RCT slices provided by Islam et al. (21). This dataset consists of both the axial-plane and coronal-plane images with categories, such as cyst, stone, cancer, and healthy. In this study, only the healthy and cancer axial-plane images alone are considered for the examination. To have a balanced database, this study considered 2,680 images (1,340 healthy class and 1,340 cancer class). Before implementing the classification task, every image is resized to 224 × 224 × 1 pixels. The proposed detection task is implemented using the RCT with and without the artifact, and the obtained results are separately examined and verified. Figure 2 represents the trial imageries considered in this study, and the number of images considered in this study is depicted in Table 2. In this study, 80% of images are considered to train the deep-learning scheme, 10% of images are considered for validation, and the remaining 10% of images are used to test the performance of the scheme with a 5-fold cross-validation with individual and fused features.

3.1.1. Artifact removal

The merit of the automatic medical image examination procedure depends mainly on the image database considered during the experimental investigation. The earlier studies in the research verify that the images without the artifact help in achieving a better accuracy compared with the images with the artifact (23). This study implements a threshold filter-supported method to remove the artifact from the chosen RCT, and this task is executed using MATLAB software as discussed in (24). In this process, the threshold value (Th), which separates the image into a processed artifact, is identified manually. When an appropriate Th is obtained, it is implemented to divide the raw test image into two sections as shown in Figure 3. Figure 3A shows the raw RCT, and Figures 3B, C shows the processed picture and the removed artifact. This task depends on the threshold level of the image, and it is shown in Figure 3D. The original histogram (red) depicts the pixel distribution of the raw RCT, the green histogram depicts the pixel distribution of the processed

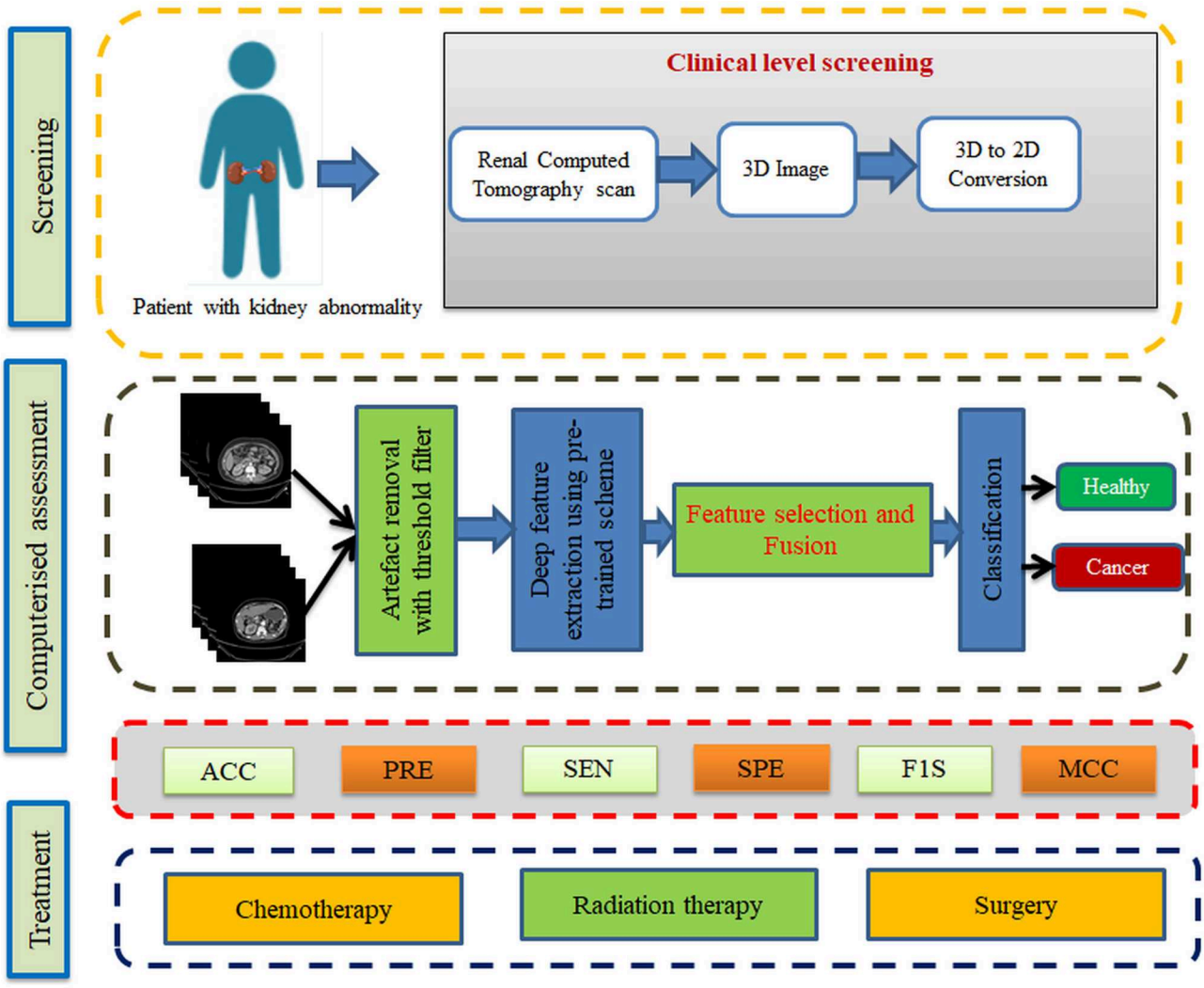


FIGURE 1
Kidney cancer detection framework.

image, and the remaining section (blue) shows the pixel value of the artifact.

3.2. Deep-learning model

Recently, pre-trained and customized deep-learning procedures have been widely implemented in various data analytic tasks due to their performance, ease of implementation, and significance. Compared to the traditional and machine-learning schemes, the deep-learning procedures efficiently provide a better result on moderate and large datasets. Furthermore, most of these methods can be practically implementable in a chosen hardware system, improving its performance (25–27).

Researchers have recently widely employed pre-trained models to achieve better results during medical image examination tasks. The proposed research study also implements well-known pre-training procedures, such as VGG16, VGG19, ResNet50, ResNet101, DenseNet121, and DenseNet201, to examine the KC in RCT slices. The complete evidence concerning the preferred schemes can be found in the literature (28–32), and in this study, these schemes are considered along with chosen binary classifiers. The following initial parameters are assigned for these models: learning rate = 1×10^{-5}, training with linear dropout rate (LDR), Adam optimization, ReLu activation, total iteration = 2000, total epochs = 150, and classification with a SoftMax unit using a 5-fold cross-validation.

Before implementing the developed scheme, an image augmentation procedure is implemented to increase the learning capability of the chosen deep-learning systems. The augmentation process involves the horizontal and vertical flip, an angle-based rotation, and zoom-in and zoom-out. This helps the system to learn better about the features of the image.

3.3. Feature vector generation and classification

Each deep-learning procedure implemented in this study provides a deep feature vector of dimension 1 x 1 x 1,000, which is then used to authenticate the merit of the classifiers. The feature

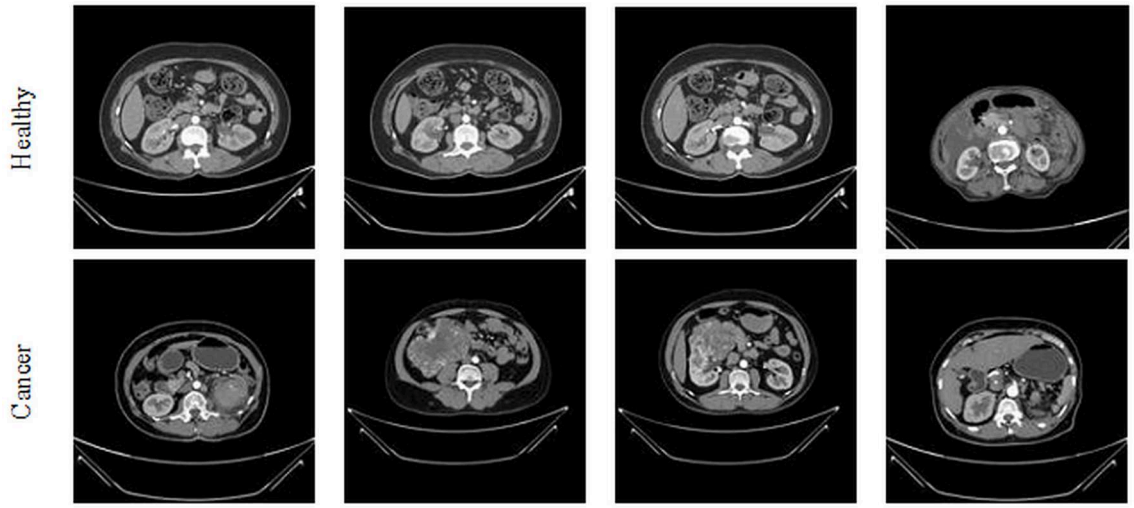


FIGURE 2
Sample axial-plane test images of renal CT slices.

TABLE 2 Dataset considered to verify the proposed framework.

Class	Dimension	Number of images			
		Total	Training (80%)	Validation (10%)	Testing (10%)
Healthy	224 × 224 × 1	1,340	1,072	134	134
Cancer	224 × 224 × 1	1,340	1,072	134	134

vector after a 50% dropout will offer a reduced feature vector of dimension 1 x 1 x 500, which is the concatenated deep feature with similar reduced features to achieve a fused feature vector of dimension 1 x 1 x 1,000, which helps in achieving a better classification accuracy during the RCT-based KC detection task. The total dimension of these features is 1 x 1 x 1,000, which is then reduced to 1 x 1 x 500 using a 50% dropout, and from this, the fused feature vector is obtained. The feature vectors of this system are depicted in Equations (1)–(3) (33, 34):

$$DLF_{VGG19\ (1\times1\times1000)} = VGG19_{1,1}, VGG19_{1,2}, ..., VGG19_{(1,1000)} \quad (1)$$

$$DLF_{DenseNet121\ (1\times1\times1000)} = DN_{1,1}, DN_{1,2}, ..., DN_{(1,1000)} \quad (2)$$

$$DLF_{VGG+DN\ (1\times1\times1000)} = VGG + DN_{1,1}, VGG + DN_{1,2}, ..., VGG + DN_{(1,1000)} \quad (3)$$

where DLF = deep-learning features, VGG = VGG19, and DN = DenseNet121.

3.4. Performance metric computation

Performance metrics obtained during the classification task are used to verify the merit of the proposed scheme. To begin with, the measures, such as true-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN), are computed from the confusion matrix presented in Equations, which are then used to implement these values into mathematical expressions. From Equations (4) to (9), the necessary measures, such as accuracy (ACC), precision (PRE), sensitivity (SEN), specificity (SPE), F1-score (F1S), and Matthews correlation coefficient (MCC), are calculated. In contrast to the binary classification task in this study, SoftMax, Nave-Bayes (NB), decision trees (DT), random forests (RF), KNNs, and support vector machine (SVM) are used (35–37).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$PRE = \frac{TP}{TP + FP} \quad (5)$$

$$SEN = \frac{TP}{TP + FN} \quad (6)$$

$$SPE = \frac{TN}{TN + FP} \quad (7)$$

$$F1S = \frac{2TP}{2TP + FP + FN} \quad (8)$$

$$MCC = \frac{(TP*TN) - (FP*FN)}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}} \quad (9)$$

4. Results and discussions

The proposed study is implemented with MATLAB and Python on a workstation equipped with an Intel i7 2.9 GHz processor, 20 GB RAM, and 4 GB VRAM.

Initially, the proposed framework is implemented on the raw RCT images with the artifacts, and the classification performance

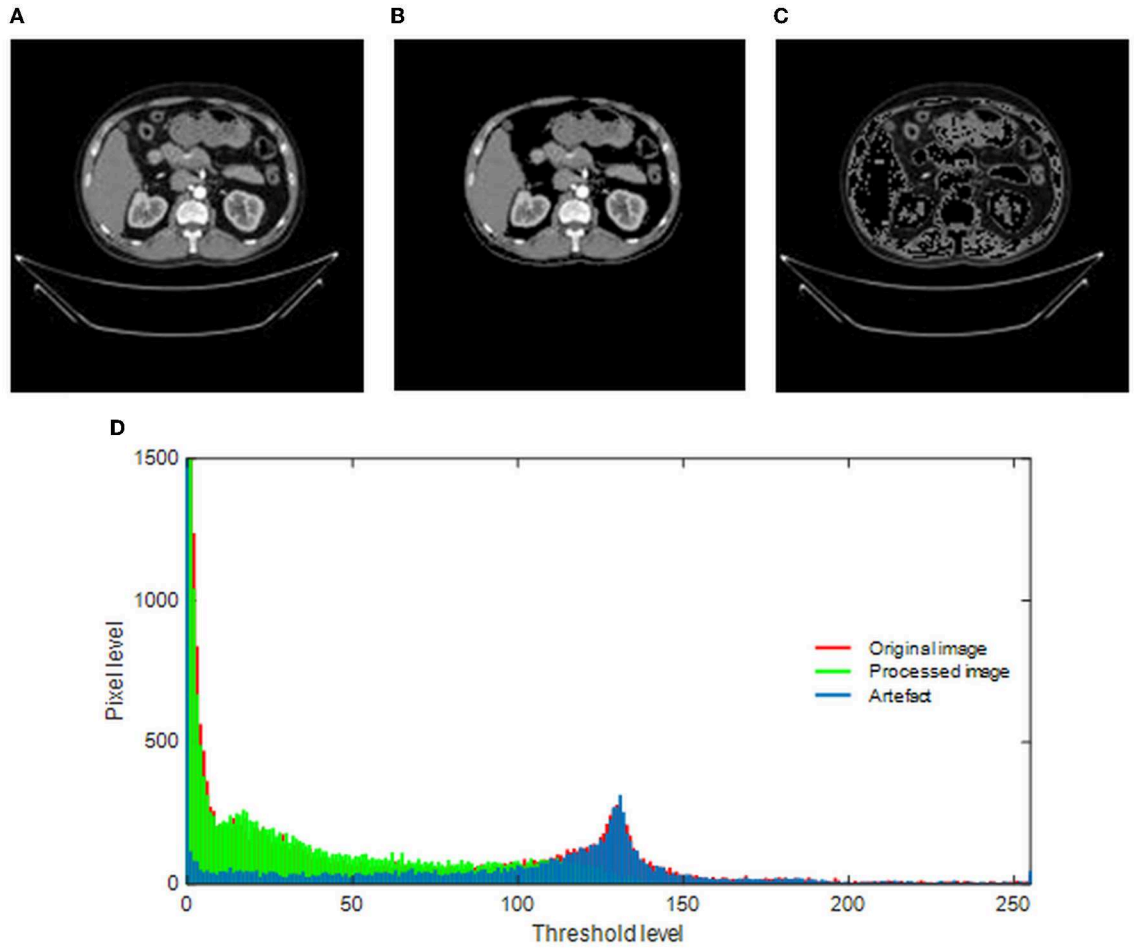


FIGURE 3
Implementation of threshold filter to eliminate the artifact. **(A)** Original image. **(B)** Processed image. **(C)** Artifact. **(D)** Gray-scale histogram.

TABLE 3 Classification results achieved for raw renal CT slice with a SoftMax classifier.

Scheme	TP	FN	TN	FP	ACC	PRE	SEN	SPE	F1S	MCC
VGG16	118	15	116	19	87.3134	86.1314	88.7218	85.9259	87.4074	74.6644
VGG19	118	17	121	12	**89.1791**	**90.7692**	**87.4074**	**90.9774**	**89.0566**	**78.4176**
ResNet50	120	13	116	19	88.0597	86.3309	90.2256	85.9259	88.2353	76.2024
ResNet101	117	20	117	14	87.3134	89.3130	85.4015	89.3130	87.3134	74.7144
DenseNet121	118	18	119	13	**88.4328**	**90.0763**	**86.7647**	**90.1515**	**88.3895**	**76.9269**
DenseNet201	118	20	116	14	87.3134	89.3939	85.5072	89.2308	87.4074	74.7130

The bold contents are the considered best metric.

is verified using the chosen binary classifiers. Then, the RCT classification performance of chosen pre-trained models is verified using the raw axial-plane images, and the outcomes are equated. The outcome of this experiment authorizes that the SoftMax-based binary classification with a 5-fold cross-validation provides a better detection performance with VGG19 and DenseNet121 methods compared with VGG16, ResNet50, ResNet101, and DenseNet201. Furthermore, along with the detection accuracy, the MCC achieved with these schemes is also better; this information is shown in Table 2.

A similar experimental task is repeated using the images whose artifacts are eliminated with a threshold filter. The results of this study confirm that this process offers a better ACC and MCC than other methods, as represented in Table 3. This table also approves that the VGG19 and DenseNet121 offer better performance. Table 5 presents the outcome for VGG16 with a SoftMax for various folds, and the best fold value is chosen as the outcome. The result of a chosen cross-validation approach is also presented in Figure 4. In this figure, the Glyph plot of Tables 3, 4 is separately developed and merged. These images are necessary to confirm the overall merit of this scheme, and this confirms that the artifact-removed RCT provides a better result than other methods. In addition, the result authorizes that this structure works fine on the chosen RCT images.

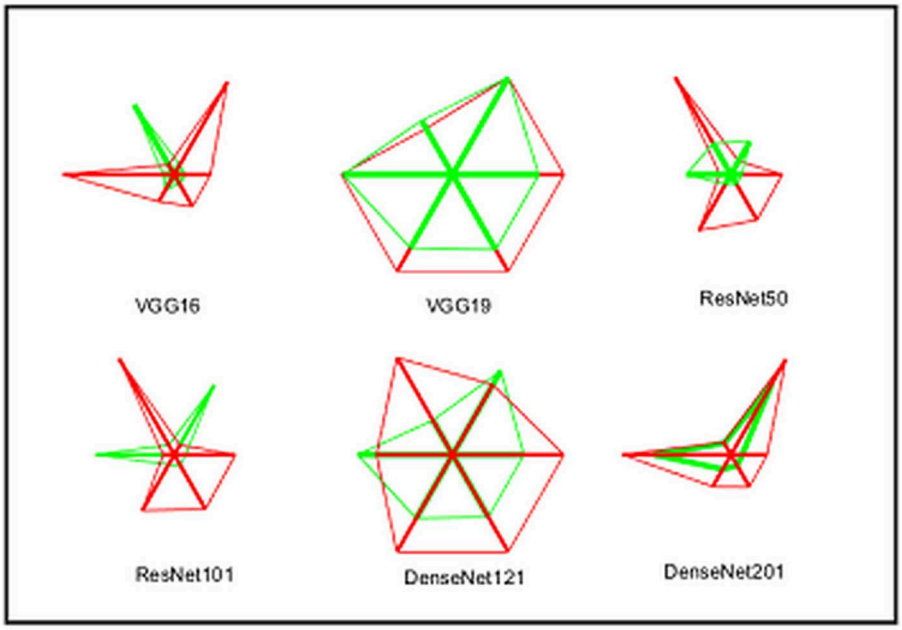


FIGURE 4
Integrated Glyph plot to demonstrate the overall performance of the considered methods.

TABLE 4 Classification results achieved for processed renal CT slice with a SoftMax classifier.

Scheme	TP	FN	TN	FP	ACC	PRE	SEN	SPE	F1S	MCC
VGG16	124	9	129	6	94.4030	95.3846	93.2331	95.5556	94.2966	88.8257
VGG19	127	7	128	6	**95.1493**	**95.4887**	**94.7761**	**95.5224**	**95.1311**	**90.3010**
ResNet50	125	8	127	8	94.0299	93.9850	93.9850	94.0741	93.9850	88.0590
ResNet101	128	5	126	9	94.7761	93.4307	96.2406	93.3333	94.8148	89.5939
DenseNet121	129	5	127	7	**95.5224**	**94.8529**	**96.2687**	**94.7761**	**95.5556**	**91.0549**
DenseNet201	126	9	127	6	94.4030	95.4545	93.3333	95.4887	94.3820	88.8295

The bold contents are the considered best metric.

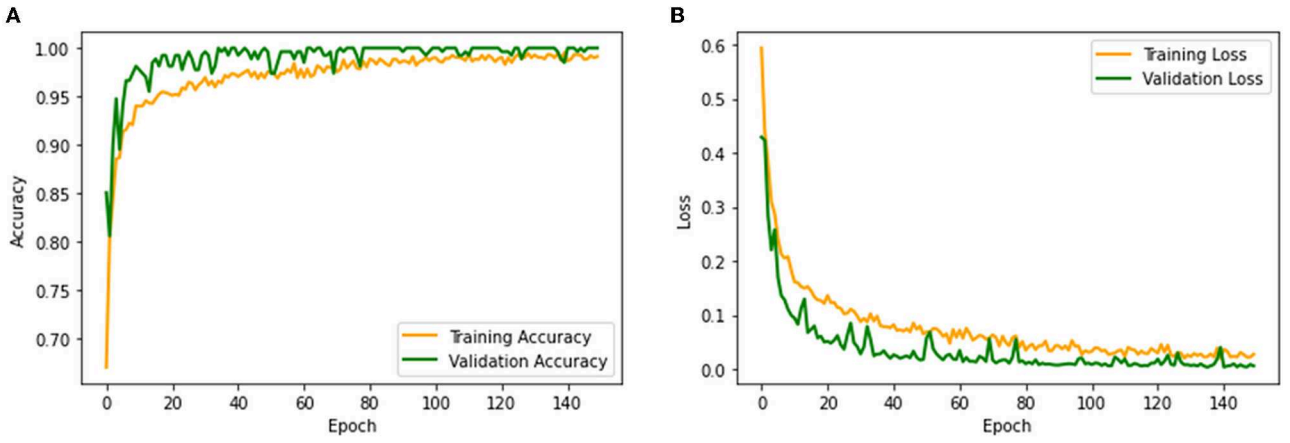


FIGURE 5
Convergence of training and validation process. **(A)** Accuracy. **(B)** Loss.

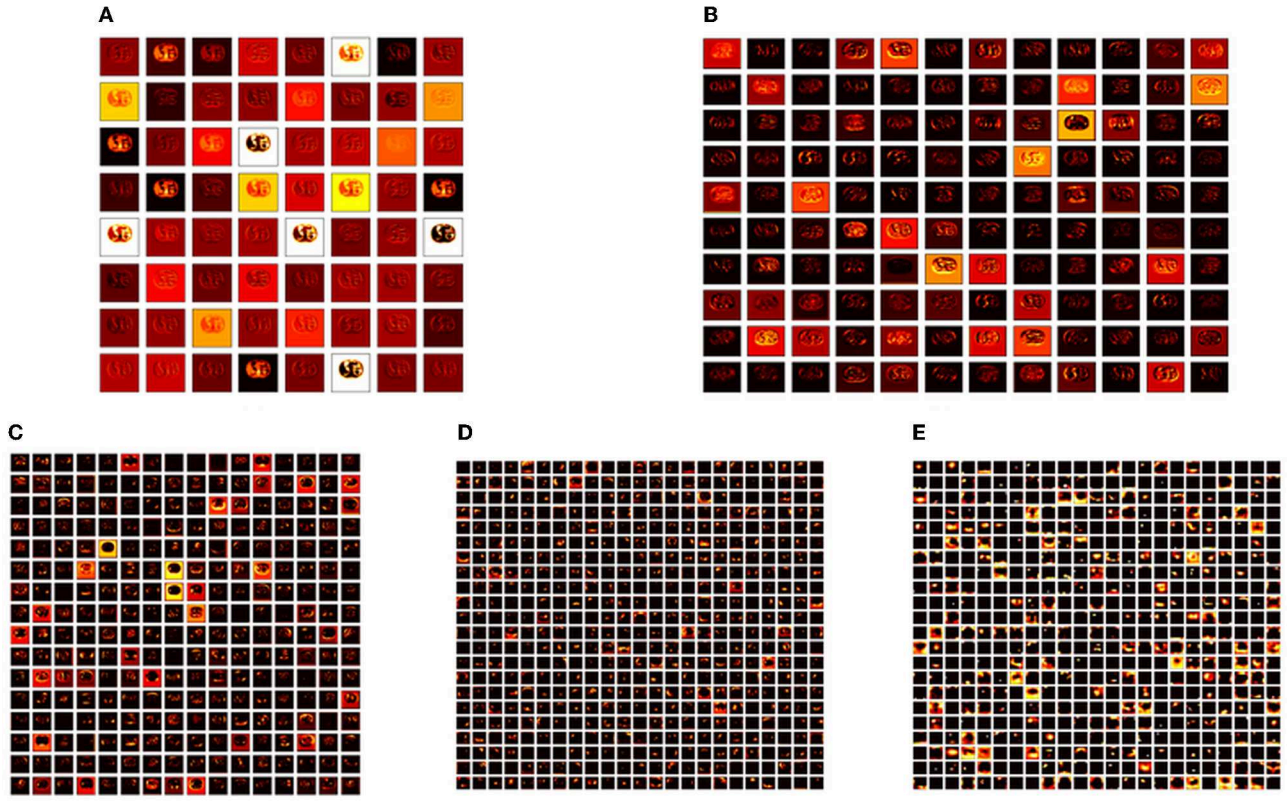


FIGURE 6
Intermediate layer outcomes collected from VGG19. **(A)** Conv1. **(B)** Conv2. **(C)** Conv3. **(D)** Conv4. **(E)** Conv5.

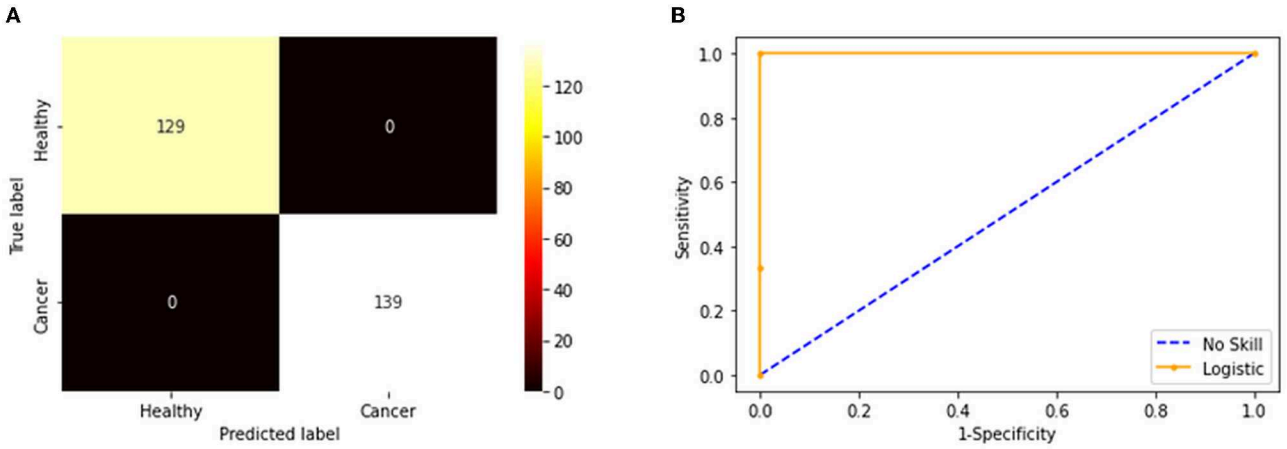


FIGURE 7
Confusion matrix and ROC curve achieved with fused features. **(A)** Confusion matrix. **(B)** the ROC curve.

The performance of the proposed system is then verified by considering the fused deep features of dimension 1 x 1 x 1,000. During this task, the VGG19 and DenseNet121 features are considered. Then, their features are sorted based on their value, and finally, a 50% dropout of these features is employed. To execute the classification task, the attained features are then serially fused to achieve a fused feature vector with dimensions of 1 x 1 x 1,000 pixels. The result of this experiment with fused features is presented in Figures 5–7. Figure 5 presents the convergence achieved with RCT image databases, and this figure confirms that the proposed method helps to achieve better detection accuracy (1,000%) than other methods. Figures 5A, B denote the experimental result achieved in this study.

The convolutional layer outcome was extracted with these results to verify the framework's performance with the chosen database. The results of Figure 6 show that this method will provide a better result

TABLE 5 Outcome of VGG16 with SoftMax for a 5-fold cross-validation.

Cross-validation	TP	FN	TN	FP	ACC	PRE	SEN	SPE	F1S	MCC
Fold 1	125	10	128	5	94.4030	96.1538	92.5926	96.2406	94.3396	88.8703
Fold 2	126	7	126	9	94.0299	93.3333	94.7368	93.3333	94.0299	88.0702
Fold 3	121	11	130	6	93.6567	95.2756	91.6667	95.5882	93.4363	87.3645
Fold 4	127	5	126	10	94.4030	92.7007	96.2121	92.6471	94.4238	88.8716
Fold 5	127	7	128	6	**95.1493**	**95.4887**	**94.7761**	**95.5224**	**95.1311**	**90.3010**

The bold contents are the considered best metric.

TABLE 6 Overall results achieved with the proposed framework for individual and fused features.

Image	Classifier	TP	FN	TN	FP	ACC	PRE	SEN	SPE	F1S	MCC
VGG19	SoftMax	127	7	128	6	95.1493	95.4887	94.7761	95.5224	95.1311	90.3010
	NB	128	6	130	4	**96.2687**	**96.9697**	**95.5224**	**97.0149**	**96.2406**	**92.5476**
	DT	127	4	130	7	95.8955	94.7761	96.9466	94.8905	95.8491	91.8141
	RF	128	7	128	5	95.5224	96.2406	94.8148	96.2406	95.5224	91.0554
	KNN	129	5	129	5	96.2687	96.2687	96.2687	96.2687	96.2687	92.5373
	SVM	130	5	127	6	95.8955	95.5882	96.2963	95.4887	95.9410	91.7927
DenseNet121	SoftMax	129	5	127	7	95.5224	94.8529	96.2687	94.7761	95.5556	91.0549
	NB	129	7	129	3	96.2687	97.7273	94.8529	97.7273	96.2687	92.5802
	DT	130	5	128	5	96.2687	96.2963	96.2963	96.2406	96.2963	92.5369
	RF	128	4	129	7	95.8955	94.8148	96.9697	94.8529	95.8801	91.8150
	KNN	129	6	130	3	**96.6418**	**97.7273**	**95.5556**	**97.7444**	**96.6292**	**93.3077**
	SVM	129	4	127	8	95.5224	94.1606	96.9925	94.0741	95.5556	91.0868
Fused deep features (VGG+DN)	SoftMax	133	0	134	1	99.6269	99.2537	100	99.2593	99.6255	99.2565
	NB	137	1	128	2	98.8806	98.5612	99.2754	98.4615	98.9170	97.7614
	DT	132	2	133	1	98.8806	99.2481	98.5075	99.2537	98.8764	97.7639
	RF	132	3	132	1	98.5075	99.2481	97.7778	99.2481	98.5075	97.0259
	KNN	139	0	129	0	**100**	**100**	**100**	**100**	**100**	**100**
	SVM	136	2	128	2	98.5075	98.5507	98.5507	98.4615	98.5507	97.0123

The bold contents are the considered best metric.

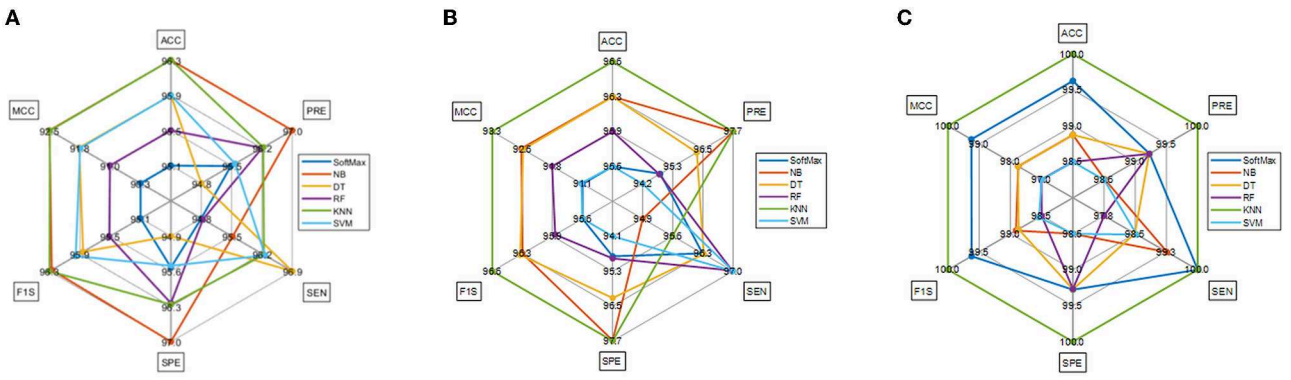


FIGURE 8
Spider plot achieved using the results of Table 6. **(A)** VGG19. **(B)** DenseNet121. **(C)** Fused deep features (VGG+DN).

and efficiency in completing the task. Figure 7 presents the outcome of the proposed technique, which shows the confusion matrix and the receiver operating characteristic (ROC), which depends mainly on the test images considered. The ROC value achieved is improved compared with the alternatives.

The result of this method authorizes that this system benefits in achieving a better result, and these measures for the experiment with conventional and fused features are shown in Table 5. The initial result for this table is achieved using a VGG19 and DenseNet121, which confirms the merit of the proposed technique. Finally, a spider plot is constructed to demonstrate the result in a graphical form, and the best result is highlighted.

The task of the proposed scheme is successfully employed using the fused features, and this scheme helps to accomplish an improved recognition accuracy (100%) compared with other methods found in the literature. The performance evaluation of Table 6 presented in Figure 8 confirms its overall merit on various classifiers. Figures 8A–C present the classification performance for different feature vectors. The main limitation of this research is the implementation of the threshold filter, which needs a manually verified Th. Nevertheless, the merit of the proposed scheme is verified using the clinical grade CT database, and the achieved experimental outcome verifies that the planned technique is better and helps to get better detection accuracy. The limitation of the proposed study is it needs an artifact removal process and it can be replaced by a chosen image enhancement scheme to achieve better disease detection accuracy.

5. Conclusion

The literature authorizes that cancer is a severe disease in human communities, and early diagnosis and treatment are necessary. When the cancer is accurately diagnosed, it can be controlled using a recommended clinical protocol. Due to its importance, a substantial amount of automatic cancer detection based on the bio-image-supported technique has been proposed and executed by researchers (38). The proposed study aims to develop a framework to effectively detect the KC in RCT images with the help of pre-trained deep-learning procedures. This study considered VGG19 and DenseNet121 schemes to classify the RCT into healthy/cancer classes with improved accuracy. As part of this study, individual DLFs and fused DLFs are employed to perform the binary classification task, and the results are compared to identify the most appropriate KC scheme. According to the results of this study, a binary classification with a KNN classifier was effective in achieving an accuracy of 100% for RCTs that had previously been preprocessed using a threshold filter. Based on the results of this research, the proposed framework appears to be effective, and it will be possible to test and validate its performance using clinically collected RCT slices in future.

Data availability statement

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

Author contributions

KS and C-YC conceptualized and supervised the research and carried out the project administration and validated the results. VR contributed to the development of the model, data processing, training procedures, and implementation of the model. VR, PV, and KS wrote the manuscript. VR, PV, KS, GA, and C-YC reviewed and edited the manuscript. C-YC carried out the funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This research was partially funded by the Intelligent Recognition Industry Service Research Center from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan and the Ministry of Science and Technology in Taiwan (Grant No. MOST 109-2221-E-224-048-MY2).

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.

References

1. Amin J, Sharif M, Fernandes SL, Wang SH, Saba T, Khan AR. Breast microscopic cancer segmentation and classification using unique 4-qubit-quantum model. *Microsc Res Tech.* (2022) 85:1926–36. doi: 10.1002/jemt.24054

2. Fernandes SL, Martis RJ, Lin H, Javadi B, Tanik UJ, Sharif M. Recent advances in deep learning, biometrics, health informatics, and data science. *Expert Systems.* (2022) 39:e13060. doi: 10.1111/exsy.13060

3. Wang SH, Fernandes SL, Zhu Z, Zhang YD, AVNC. attention-based VGG-style network for COVID-19 diagnosis by CBAM. *IEEE Sens J.* (2021) 22:17431–8. doi: 10.1109/JSEN.2021.3062442

4. Ueda T. Estimation of three-dimensional particle size and shape characteristics using a modified 2D−3D conversion method employing spherical harmonic-based principal component analysis. *Powder Technol.* (2022) 404:117461. doi: 10.1016/j.powtec.2022.117461

5. Krishnamoorthy S, Zhang Y, Kadry S, Yu W. Framework to segment and evaluate multiple sclerosis lesion in MRI slices using VGG-UNet. *Comput Intell Neurosci*. (2022) 2022:8096. doi: 10.1155/2022/4928096

6. Rajinikanth V, Fernandes SL, Bhushan B, Sunder NR. Segmentation and analysis of brain tumor using Tsallis entropy and regularised level set. *InProceedings of 2nd International Conference on Micro-Electronics, Electromagnetics, and Telecommunications*. (pp. 313-321). Springer, Singapore (2018). doi: 10.1007/978-981-10-4280-5_33

7. Arunmozhi S, Raja NS, Rajinikanth V, Aparna K, Vallinayagam V. Schizophrenia detection using brain MRI—A study with watershed algorithm. *In2020 International Conference on System, Computation, Automation and Networking (ICSCAN) 2020 Jul 3*. (pp. 1-4). IEEE (2020). doi: 10.1109/ICSCAN49426.2020.9262281

8. Rajinikanth V, Dey N, Raj AN, Hassanien AE, Santosh KC, Raja N. Harmony-search and otsu based system for coronavirus disease (COVID-19) detection using lung CT scan images. *arXiv preprint arXiv*. (2020).

9. Dey N, Zhang YD, Rajinikanth V, Pugalenthi R, Raja NS. Customized VGG19 architecture for pneumonia detection in chest X-rays. *Pattern Recognit Lett*. (2021) 143:67–74. doi: 10.1016/j.patrec.2020.12.010

10. Kadry S, Rajinikanth V. Computer assisted detection of low/high grade nodule from lung CT scan slices using handcrafted features. *Det Sys Lung Cancer Imag*. (2022) 1:3–1. doi: 10.1088/978-0-7503-3355-9ch3

11. Holzinger, Andreas. The next frontier: Ai we can really trust. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, Cham (2021). doi: 10.1007/978-3-030-93736-2_33

12. Alzu'bi D, Abdullah M, Hmeidi I, AlAzab R, Gharaibeh M, El-Heis M, et al. Kidney tumor detection and classification based on deep learning approaches: a new dataset in CT scans. *J Healthcare Engin*. (2022) 2022:1161. doi: 10.1155/2022/3861161

13. Xu L, Yang C, Zhang F, Cheng X, Wei Y, Fan S, et al. Deep learning using CT images to grade clear cell renal cell carcinoma: development and validation of a prediction model. *Cancers*. (2022) 14:2574. doi: 10.3390/cancers141 12574

14. Amiri S, Akbarabadi M, Abdolali F, Nikoofar A, Esfahani AJ, Cheraghi S. Radiomics analysis on CT images for prediction of radiation-induced kidney damage by machine learning models. *Comput Biol Med*. (2021) 133:104409. doi: 10.1016/j.compbiomed.2021.104409

15. Miskin N, Qin L, Matalon SA, Tirumani S.H, Alessandrino F, Silverman SG, et al. Stratification of cystic renal masses into benign and potentially malignant: applying machine learning to the Bosniak classification. *Abdominal Radiol*. (2021) 46:311–8. doi: 10.1007/s00261-020-02629-w

16. Shehata M, Alksas A, Abouelkheir RT, Elmahdy A, Shaffie A, Soliman A, et al. A comprehensive computer-assisted diagnosis system for early assessment of renal cancer tumors. *Sensors*. (2021) 21:4928. doi: 10.3390/s21144928

17. Heller N, Isensee F, Maier-Hein KH, Hou X, Xie C, Li F, et al. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: results of the KiTS19 challenge. *Med Image Anal*. (2021) 67:101821. doi: 10.1016/j.media.2020.101821

18. Bhandari A, Ibrahim M, Sharma C, Liong R, Gustafson S, Prior M. CT-based radiomics for differentiating renal tumours: a systematic review. *Abdom Radiol*. (2021) 46:2052–63. doi: 10.1007/s00261-020-02832-9

19. Islam MN, Hasan M, Hossain M, Alam M, Rabiul G, Uddin MZ, et al. Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography. *Sci Rep*. (2022) 12:1–4. doi: 10.1038/s41598-022-15634-4

20. Nikpanah M, Xu Z, Jin D, Farhadi F, Saboury B, Ball MW, et al. Deep-learning based artificial intelligence (AI) approach for differentiation of clear cell renal cell carcinoma from oncocytoma on multi-phasic MRI. *Clin Imag*. (2021) 77:291–8. doi: 10.1016/j.clinimag.2021.06.016

21. Abdelrahman A, Viriri S. Kidney tumor semantic segmentation using deep learning: a survey of state-of-the-Art. *J Imaging*. (2022) 8:55. doi: 10.3390/jimaging8030055

22. Wang J, Zhu H, Wang SH, Zhang YD. A review of deep learning on medical image analysis. *Mob Networks Appl*. (2021) 26:351–80. doi: 10.1007/s11036-020-01672-7

23. Khan MA, Rajinikanth V, Satapathy SC, Taniar D, Mohanty JR, Tariq U, et al. VGG19 network assisted joint segmentation and classification of lung nodules in CT images. *Diagnostics*. (2021) 11:2208. doi: 10.3390/diagnostics11122208

24. T Krishnan P, Balasubramanian P, Krishnan C. Segmentation of brain regions by integrating meta heuristic multilevel threshold with markov random field. *Curr Med Imag*. (2016) 12:4–12. doi: 10.2174/1573394711666150827203434

25. Liu S, Liu S, Zhang S, Li B, Hu W, Zhang YD. SSAU-net: a spectral-spatial attention-based u-net for hyperspectral image fusion. *IEEE Transact Geosci Remote Sens*. (2022). doi: 10.1109/TGRS.2022.3217168

26. Arco JE, Ortiz A, Ramírez J, Martínez-Murcia FJ, Zhang YD, Górriz JM. Uncertainty-driven ensembles of multi-scale deep architectures for image classification. *Inform Fusion*. (2023) 89:53–65. doi: 10.1016/j.inffus.2022.08.010

27. Mohan R, Kadry S, Rajinikanth V, Majumdar A, Thinnukool O. Automatic detection of tuberculosis using VGG19 with seagull-algorithm. *Life*. (2022) 12:1848. doi: 10.3390/life12111848

28. Nancy AA, Ravindran D, Raj Vincent PD, Srinivasan K, Gutierrez Reina D. IoT-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning. *Electronics*. (2022) 11:2292. doi: 10.3390/electronics11152292

29. Nandhini Abirami R, Durai Raj Vincent PM, Srinivasan K, Manic KS, Chang CY. Multimodal medical image fusion of positron emission tomography and magnetic resonance imaging using generative adversarial networks. *Behav Neurol*. (2022) 2022:8783. doi: 10.1155/2022/6878783

30. Mahendran N, PM DR A. deep learning framework with an embedded-based feature selection approach for the early detection of the Alzheimer's disease. *Comput Biol Med*. (2022) 141:105056. doi: 10.1016/j.compbiomed.2021.105056

31. Tiwari RS, Das TK, Srinivasan K, Chang CY. Conceptualising a channel-based overlapping CNN tower architecture for COVID-19 identification from CT-scan images. *Sci Rep*. (2022) 12:1–5. doi: 10.1038/s41598-022-21700-8

32. Bhardwaj P, Gupta P, Guhan T, Srinivasan K. Early diagnosis of retinal blood vessel damage via deep learning-powered collective intelligence models. *Comput Math Meth Med*. (2022) 5:2022. doi: 10.1155/2022/3571364

33. Gudigar A, Raghavendra U, Rao TN, Samanth J, Rajinikanth V, Satapathy SC, et al. FFCAEs: an efficient feature fusion framework using cascaded autoencoders for the identification of gliomas. *Int J Imag Sys Technol*. (2022) 3:2820. doi: 10.1002/ima.22820

34. Kadry S, Srivastava G, Rajinikanth V, Rho S, Kim Y. Tuberculosis detection in chest radiographs using spotted hyena algorithm optimized deep and handcrafted features. *Comput Intell and Neurosci*. (2022) 3:2022. doi: 10.1155/2022/9263379

35. Srinivasan K, Cherukuri AK, Vincent DR, Garg A, Chen BY. An efficient implementation of artificial neural networks with K-fold cross-validation for process optimization. *J Internet Technol*. (2019) 20:1213–25. doi: 10.3966/160792642019072004020

36. Amin J, Sharif M, Yasmin M, Fernandes SL. A distinctive approach in brain tumor detection and classification using MRI. *Pattern Recognit Lett*. (2020) 139:118–27. doi: 10.1016/j.patrec.2017.10.036

37. Fernandes SL, Gurupur VP, Sunder NR, Arunkumar N, Kadry S, A. novel non-intrusive decision support approach for heart rate measurement. *Pattern Recognit Lett*. (2020) 139:148–56. doi: 10.1016/j.patrec.2017.07.002

38. Srinivasan K, Ankur A, Sharma A. Super-resolution of magnetic resonance images using deep convolutional neural networks. *IEEE International Conference on Consumer Electronics—Taiwan (ICCE-TW)*. (2017), pp. 41-42, doi: 10.1109/ICCE-China.2017.7990985

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 02 February 2023
DOI 10.3389/fpubh.2023.1091850

OPEN ACCESS

EDITED BY
Steven Fernandes,
Creighton University, United States

REVIEWED BY
Reham Reda Mostafa,
Mansoura University, Egypt
Geno Peter,
University College of Technology Sarawak, Malaysia
Feras Alattar,
National University of Science and Technology, Oman

*CORRESPONDENCE
Chuan-Yu Chang
✉ chuanyu@yuntech.edu.tw

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 07 November 2022
ACCEPTED 16 January 2023
PUBLISHED 02 February 2023

CITATION
Vijay S, Guhan T, Srinivasan K, Vincent PMDR and Chang C-Y (2023) MRI brain tumor segmentation using residual Spatial Pyramid Pooling-powered 3D U-Net.
Front. Public Health 11:1091850.
doi: 10.3389/fpubh.2023.1091850

MRI brain tumor segmentation using residual Spatial Pyramid Pooling-powered 3D U-Net

Sanchit Vijay[1†], Thejineaswar Guhan[2†], Kathiravan Srinivasan[3], P. M. Durai Raj Vincent[2] and Chuan-Yu Chang[4,5*]

[1]School of Electronics Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India, [2]School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India, [3]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India, [4]Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Yunlin, Taiwan, [5]Service Systems Technology Center, Industrial Technology Research Institute, Hsinchu, Taiwan

Brain tumor diagnosis has been a lengthy process, and automation of a process such as brain tumor segmentation speeds up the timeline. U-Nets have been a commonly used solution for semantic segmentation, and it uses a downsampling-upsampling approach to segment tumors. U-Nets rely on residual connections to pass information during upsampling; however, an upsampling block only receives information from one downsampling block. This restricts the context and scope of an upsampling block. In this paper, we propose SPP-U-Net where the residual connections are replaced with a combination of Spatial Pyramid Pooling (SPP) and Attention blocks. Here, SPP provides information from various downsampling blocks, which will increase the scope of reconstruction while attention provides the necessary context by incorporating local characteristics with their corresponding global dependencies. Existing literature uses heavy approaches such as the usage of nested and dense skip connections and transformers. These approaches increase the training parameters within the model which therefore increase the training time and complexity of the model. The proposed approach on the other hand attains comparable results to existing literature without changing the number of trainable parameters over larger dimensions such as 160 × 192 × 192. All in all, the proposed model scores an average dice score of 0.883 and a Hausdorff distance of 7.84 on Brats 2021 cross validation.

KEYWORDS

brain tumor segmentation, 3D U-Net, Spatial Pyramid Pooling, image processing, healthcare

1. Introduction

Brain tumor segmentation using magnetic resonance images (MRI) is a vital step for treating tumors present in the brain and a specialist can use this to find the damage caused by a tumor in a region. The most frequent and severe malignant brain tumors are glioblastomas, often known as gliomas (GBM). Magnetic resonance imaging (MRI) with automated and exact segmentation of these malignancies is critical for early diagnosis as well as for administering and monitoring treatment progression. Assessment of tumor presence is the first step in brain tumor diagnosis and the assessment is done on the basis of segmentation of tumors present in MRI. This process is often done manually making it a time and human intensive task. Moreover, tumors exist in different forms and sizes making it a task requiring expertise. The process of assessment can be sped up by automating the segmentation of brain tumors (1).

The Brain Tumor Segmentation Challenge (BraTS) (2, 3) is a worldwide annual competition that has been concentrating on evaluation of state-of-the-art automated tumor sub-region segmentation algorithms since 2012. The American Society of Neuroradiology (ASNR), the Radiological Society of North America (RSNA), and MICCAI together hosted the BraTS

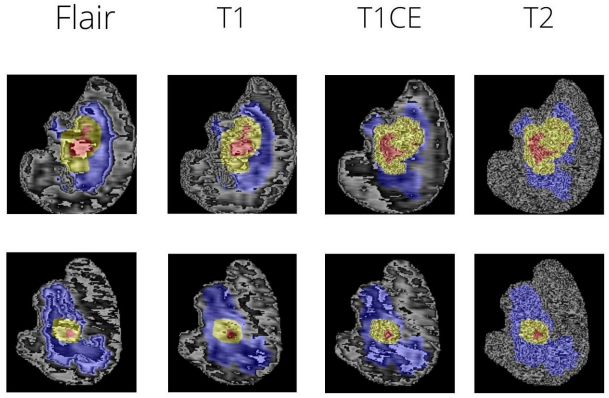


FIGURE 1
The different views of brain MRI slices with annotations.

2021 competition (1) honoring its 10th anniversary. With 1,251 meticulously annotated, multi-institutional, multi-parametric MR images (mpMRI) of patients with various degrees of gliomas, BraTS 2021 provides us with a sizable dataset. The segmentation of the histologically diverse brain tumor sub-regions and the classification of the tumor's O-methylguanine-DNA methyltransferase (MGMT) promoter methylation status are the two main goals of BraTS 2021. In this study, the first task will be the main focus.

The peritumoral edematous/invaded tissue (ED-label 2), the Gd-enhancing tumor (ET-label 4), and the necrotic tumor core are the tumor sub-regions for each patient (NCR-label 1). The peritumoral edematous and infiltrated tissue known as ED has an infiltrative non-enhancing tumor as well as peritumoral vasogenic edema and is linked with an abnormal hyperintense signal envelope on the T2 FLAIR volumes. ET stands for the tumor's enhancing segment and is identified by T1Gd MRI regions that exhibit some enhancement. On T1Gd MRI, the necrotic core of the tumor, or NCR, seems to be substantially less intense. Figure 1 depicts the various tumor sub-regions.

In many vision tasks like segmentation, particularly in the healthcare industry, deep learning-based segmentation systems have shown amazing success, outperforming other traditional methods in brain tumor analysis (4–8). With exceptional results, Fully Convolutional Networks (FCN) (9) achieve end-to-end semantic segmentation for the first time. The most popular architecture for medical picture segmentation is called U-Net (10), which combines a symmetric encoder-decoder topology with skip-connections to maximize information preservation. The performance for image segmentation is greatly improved by many U-Net variants, including U-Net++ (11), two-stage cascaded U-Net (12), and Res-U-Net (13). Although CNN-based techniques have great encoding capacities, because of the convolution kernels' constrained receptive fields, it is challenging to produce an apparent long-distance dependency. Learning global semantic information, which is essential for dense prediction issues like segmentation, is made more difficult by this constraint of convolution operation.

U-Nets consist of residual connections, and these connections are key for reconstruction. These connections pass local and global information to a particular decoder (10). However, information passed from one layer to another may be inadequate for reconstruction. Potentially passing information from a higher resolution may provide better clarity as inputs passed from one layer to another information is lost due to downsizing. Hence skip connections can further be employed to pass information from higher dimensional encoders.

Segmentation maps have been formed using a 3D U-Net which consists of three downsampling and upsampling blocks followed by a set of convolutional layers. The authors use a patching approach to train the model (14). Kaur et al. (15) proposes a 2D and 3D DGA-U-Net. In the 3D model, mainly the pooling layers are replaced with upsampling. The following is done to increase the resolution of the image within the contraction phase of the U-Net. Punn and Agarwal (16) utilized a multi-modal approach to segment brain tumors, where the multi-modalities of the dataset are fused across using deep inception encoding. Finally, a tumor extractor collects features from the fused images to the tumor segmenter. The extractor and segmentation have an U-Net-based architecture. Jiang et al. (12) used a cascaded U-Net in 2 stages. The approach is multi-modal in nature, where in all the class maps are concatenated and passed to the first U-Net. The output of the first U-Net along with the concatenated model input is passed to the second U-Net. Here, a triplet loss is used to train the model, where in the output of the first U-Net along with output of second U-Net and two output maps (Deconvolution and Interpolation approach). Isensee et al. (17) used the nn-U-Net (18) framework to propose a model which is then further enhanced by using post-processing, patching strategies and augmentations that are Brats specific. Qamar et al. (19) increased the contextual information by using a Hyperdense Inception (HI) 3D U-Net. The HI methodology builds the connections between factored convolutional layers to look more like dense connections. U-Nets have been versatile wherein transformer-based models are used within the model (20–22) and have provided significant improvement in results.

Wang et al. (23) proposed a SAR-U-Net which is based on the traditional U-Net with SE (Squeeze and Excitation) block to avoid focus on unnecessary regions within the dataset and Atrous Spatial Pyramid Pooling (ASPP) (24) to pass information on a multi scale basis. The model is trained on LITs dataset and has achieved significant results. Ahmad et al. (25) used a similar approach of using ASPP along with U-Net on Cardiac MRI dataset. The following two approaches are 2-dimensional in nature. Jiang et al. (26) used a 3D Atrous Inception U-Net where the Atrous pooling is used in the residual connections between the encoder and decoder on the Brats dataset. In this approach, the outputs of the succeeding encoder blocks are upsampled and concatenated across before sending to the decoder for reconstruction. Wang et al. (27) introduced the 3D CNN based Transformers for segmenting brain tumors.

Hence, we were able to identify some research gaps:

- As can be seen, existing literature uses heavy approaches such as the usage of nested and dense skip connections and transformers. Hence an approach which considers the parameters in mind is needed. Considering applications such as edge computing which heavily emphasize efficient and accurate predictions, the proposed mechanism fits such problem statement in hand.
- Moreover, the skip connections have always been an aspect of the experimentation. Additional information to the decoder

layers through mechanisms such as ASPP has given performance improvement. Hence utilizing a similar mechanism on multiple encodings in a 3-dimensional manner seemed to be an idea for the research.

1.1. Contributions

We propose a U-Net with SPP and attention. SPP takes information from three encoder layers and passes it to the decoder in the U-Net. The proposed addition provides the model with additional context and information for better reconstruction by providing scope from neighboring layers. The proposed mechanism does not have additional training parameters therefore the need for computational power remains the same. Therefore, the resultant model is lightweight in nature aiding for faster medical diagnosis and medical workflow in a production environment. To introduce reproducibility, the codebase utilized has been made public: https://github.com/sanchitvj/rsppUnet-BraTS-2021. We encourage the community to use and possibly improve the mechanism further in the form of open-source contributions.

2. Materials and methods

2.1. Data processing

The dataset used was Brats 2021. The MRI scans were firstly bought across to a common dimension of 160 × 192 × 192. This size was arrived upon based on experimentation and the comparison was done on the basis of Dice Score (further discussed in results). Figure 1 shows sample MRI slices from two MRI files. The scans are brought to a common dimension using padding and cropping. Padding is used whenever the image size is lower than the specified size and in cases where dimension of the original image being larger cropping takes place. Augmentations are key in this case as the number of data samples is low, hence a combination of augmentations are used at random. The following augmentations are used:

- Image flip
- Brightness adjust
- Rotation: Images can be rotated on the z-axis with the maximum angle of rotation being 30° and the minimum angle of rotation being −30°.
- Elastic transformation
- Intensity shift

Note that the choice of augmentations, within this set, used are random hence this makes the model robust to overfitting. The following is achieved by randomly choosing the augmentations on the basis of a threshold. K-Folds were used to divide the data into 5-folds, with Fold 1 being used to assess the model's performance and the other folds being used for training. Table 1 demonstrates the distribution of the data used. Fold 1 was chosen on the basis of metric stability. It was often noted that results achieved on Fold 1 had a relatively smooth progression. This dataset has a balance of noisy and normal data samples. In a way, training on these other noisy folds makes the model get a generalized understanding of the data.

TABLE 1 Data split for brats 2021.

Data split	
Split name	Number of samples
Train split	1,000
Validation split	250

2.2. Residual spatial pyramid pooling-powered 3D U-Net model

Spatial Pyramid Pooling (21) has been widely used in classification and object detection. The reason being, SPP provides an effective representation of varying sized images and it can be considered as an ensemble of pooling layers. In this way, the feature maps captured by convolutional layers can be deciphered in various ways, and pooling has often been the solution to aggregate the learning of convolutional layers. Hence, concentrating information using different dimensional pooling layers can provide representations that can further enhance the performance of the model.

Atrous Spatial Pyramid Pooling was proposed based on SPP and carries the concept of SPP by using parallel Atrous Convolutional layers. ASPP has been extensively used in semantic segmentation, and it serves the purpose of providing context at different levels or views. ASPP has been employed in various studies within brain segmentation. However, as per Tampu et al. (28), boosting context alone does not increase the performance of the model.

Attention is a process through which we humans put forth focus on doing certain tasks. While reading, we capture context by understanding neighboring words within a sentence. This mechanism is applied to the attention layer and its purpose is to capture context. The attention layer has been extensively employed in deep learning and has contributed to cutting-edge outcomes. Attention is obtained for the model by combining the output of two encoder layers. By feeding the output of two encoder layers into two different 3D convolutional layers, the following is accomplished. The output of the two layers is combined, and relu is then used to activate it. The activated output is passed through a 3D convolutional layer and is then normalized and activated. Fusing the output of these layers along with activation aids in maintaining context while not compromising on the dimensionality aspect.

Hence, SPP is used as a feature aggregator within the model, and to introduce context, attention layers are employed. The SPP layer, along with the attention layer, have been used to replace some residual connections within the U-Net. SPP is typically used at the end of the process, after the feature maps have been flattened so that fully connected neural networks can use the maps to predict class(es) or bbox(es). A 3D convolutional layer with a kernel size of 1 is utilized to modify SPP so that it functions as a residual connection. The output of the SPP is again converted to a 3D representation by this layer. Additionally, by sending input from many encoder levels to each pooling layer, information is gathered over a wide range. Figure 2 shows the architecture of the SPP Layer.

The U-Net used is based on the NvNet (29) and the following figure shows the architecture of the model. As shown in the architecture SPP is just used in two places, the reason for the same was to maintain the aspect of dimensionality. The SPP layer takes

input from various encoder layers therefore when pooling is applied the dimension of the output varies substantially.

In this paper, experimentation is done on 3 model architectures based on Figure 3:

- No SPP: The SPP blocks would be omitted therefore boosting the model only in terms of context.
- 1 SPP: The upper SPP block would be removed from the architecture keeping only the lower block (with 3rd encoder layer). Hence boosting the model with a combination of context and features.
- 2 SPP: Both the SPP blocks were used. This model carries more feature boost from the other two models used.

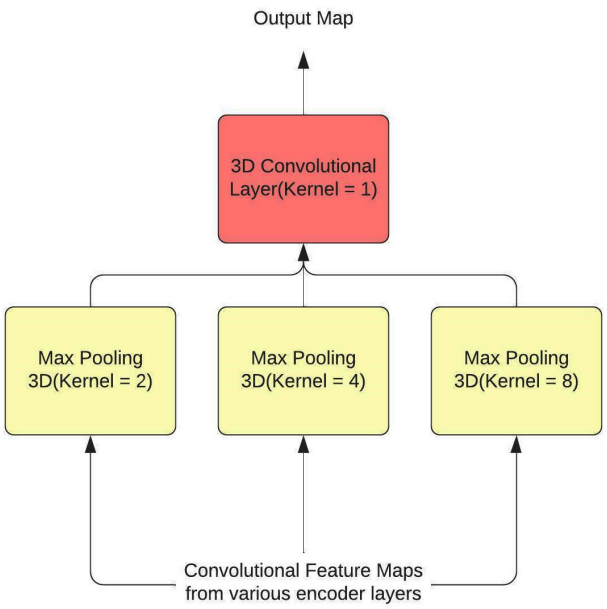


FIGURE 2
The architectural diagram for SPP.

2.2.1. Training procedure

Table 2 shows the hyperparameters used during training. In general, the increase in performance post 60 epochs was negligible hence the same was chosen. We experimented with different sizes (image size format: channel * length * width) such as 160 × 160 × 160, 128 × 160 × 160, and 160 × 192 × 192. The original dimensions of the slice were 155 × 240 × 240. Among these 160 × 192 × 192 showed the best convergence so we decided to go with it. The

TABLE 2 Hyperparameters used for training.

Hyperparameters used	
Hyperparameter	**Value**
Image size (channels * length * width)	160 × 192 × 192
Epochs	60
Learning rate	2.50E-04
Weight decay	1.00E-07
Scheduler	Cosine annealing LR
Criterion	Dice loss
Optimizer	Adam
Normalization	Group norm

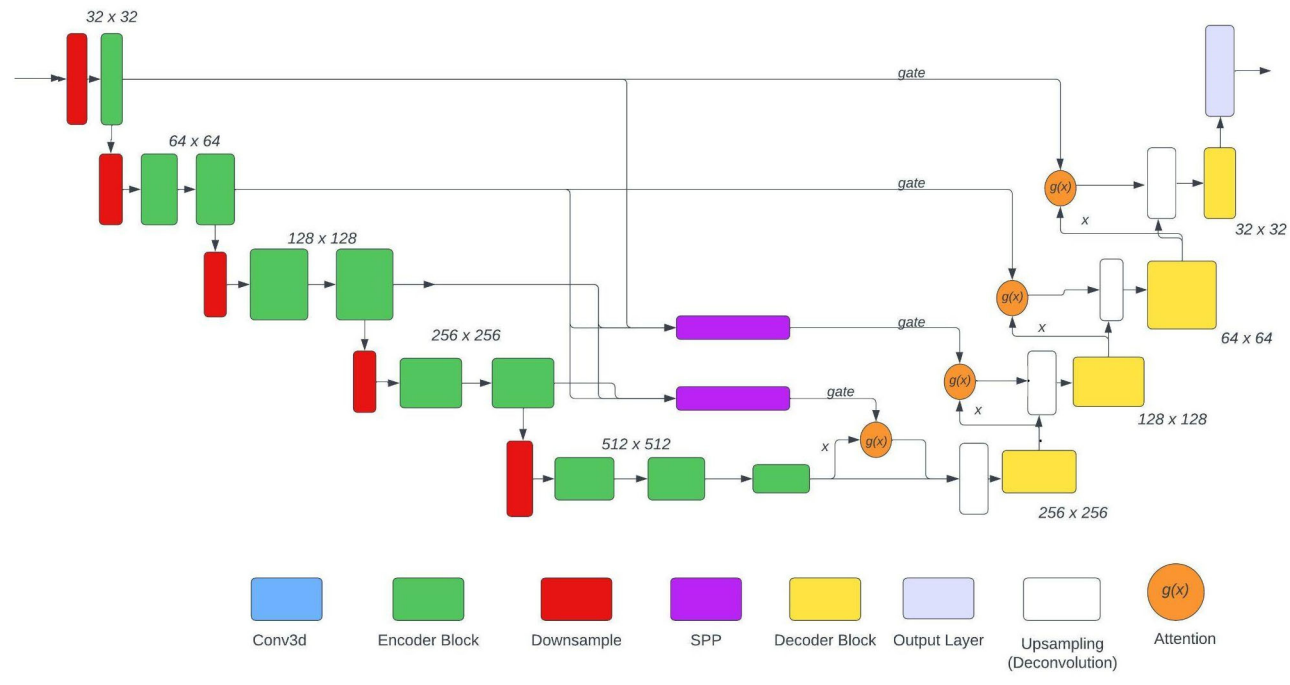


FIGURE 3
The architecture of U-Net used.

TABLE 3 Results obtained in standard setting.

Model name	Metrics							
	WT, whole tumor; TC, tumor core; ET, enhancing tumor (sub regions of tumor affected brain)							
	Hausdorff distance				Dice			
	WT	TC	ET	Average	WT	TC	ET	Average
No SPP	13.070	11.010	10.210	11.430	0.908	0.877	0.838	0.870
1 SPP	9.430	7.780	6.300	7.840	0.899	0.899	0.850	0.883
2 SPP	16.060	5.650	5.270	8.990	0.904	0.880	0.845	0.876
Hatamizadeh et al. (32)	4.739	15.309	16.326	12.120	0.927	0.876	0.853	0.890
Jia and Shu (22)	3.000	2.236	1.414	2.220	0.926	0.935	0.887	0.920
Qamar et al. (19)	–	–	–	–	0.875	0.837	0.795	0.840
Jiang et al. (12)	4.610	4.130	2.650	3.800	0.888	0.837	0.833	0.850

TABLE 4 Results obtained with image size 160 × 160 × 160.

Model name	Metrics							
	WT, whole tumor; TC, tumor core; ET, enhancing tumor (sub regions of tumor affected brain)							
	Hausdorff distance				Dice			
	WT	TC	ET	Average	WT	TC	ET	Average
No SPP	34.1	7.97	7.13	16.4	0.895	0.872	0.837	0.868
1 SPP	18.6	6.13	4.88	9.87	0.887	0.879	0.842	0.869
2 SPP	20.12	7.42	6.22	11.25	0.886	0.876	0.843	0.868

optimizer of choice was kept as Adam and the loss function of choice was Squared Soft Dice Loss as proposed by Milletari et al. (30) the working of the same shown in Equation (1). Here p_i denotes the truth label for the pixel and g_i denotes the model prediction where N denotes the number of voxels. The prime reason for choosing this function was to avoid the focus of the loss function from the background regions. In problems such as brain segmentation, the size of the regions consisting of tumors are very small relative to the background region and weighted losses have not been the most efficient solution for the same. This function ranges in the value of 0–1 with an objective to maximize the loss.

$$Dice\ Loss = \frac{2^* \sum_i^N p_i g_i}{p_i^2 g_i^2} \quad (1)$$

Based on our experimentation we found the issue of gradient explosion hence group normalization was employed. Batch Normalization did not work in our cases as high batch size could not be used for training. Batch size >1 did not provide the expected results and at times would also result in the GPU running out of memory and process killing. Hence, the choice for batch size was kept as 1. A Nvidia Tesla V100 GPU was utilized for the training of the models.

2.2.2. Training procedure

The evaluation process of the model has been done on the basis of cross-validation and the model was evaluated on two metrics:

- Dice Score (as showing in Equation 2): In short it is the F1-Score conveyed on behalf of image pixels: Wherein the ground truth is the annotated pixels. Dice score is an efficient metric as it penalizes false positives: If the predicted map has large false positives, it is used in the denominator rather than the numerator.

$$Dice\ Score = \frac{2^* Region\ of\ Overlap}{Region\ of\ Union} \quad (2)$$

- Hausdorff Distance: The Hausdorff distance (31) describes how closely each point in a model set resembles a point in an image set and vice versa. So, the degree of similarity between two items that are superimposed on one another can be gauged using this closeness.

It should be noted that Hausdorff Distance is unconcerned with the size of the image's background. By calculating the extensive distance between the extremes of the two outlines, the Hausdorff distance complements the Dice metric. A prediction may show nearly voxel-perfect overlap since it severely penalizes outliers, but the Hausdorff distance will only be meaningful if a certain voxel is far from the reference segmentation. This statistic is quite useful for determining the clinical importance of segmentation, despite being noisier than the Dice index.

3. Results and discussion

Two models from the Brats 2021 dataset, as well as one model from each of the Brats 2020 and Brats 2019 datasets, are compared to the suggested model.

As per Table 3 the following inferences can be made:

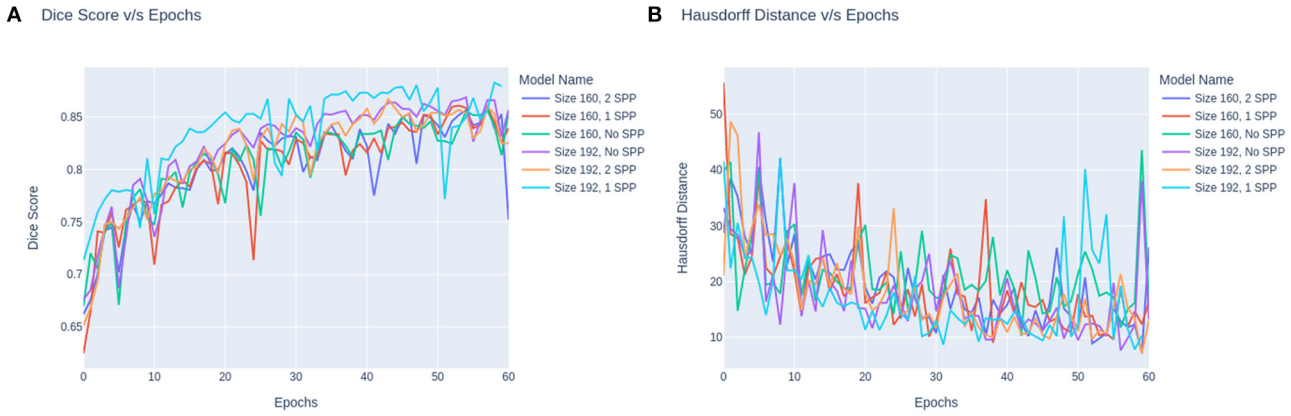


FIGURE 4
(A) The dice score vs. epochs. **(B)** Hausdorff scores vs. epochs.

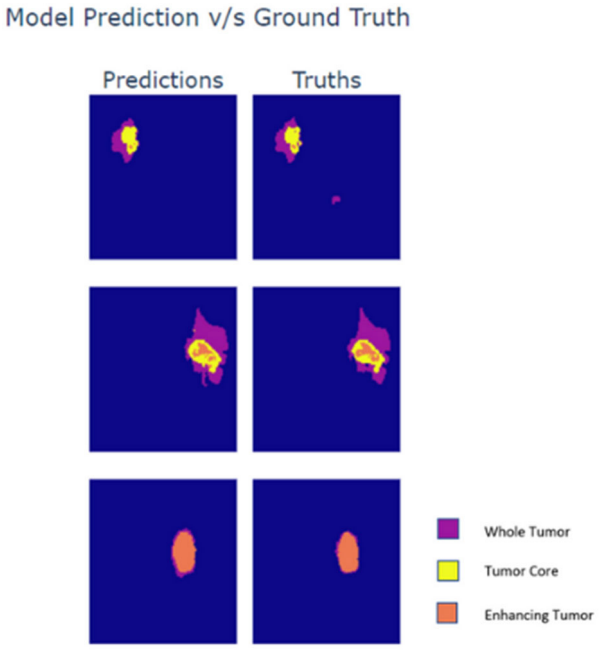


FIGURE 5
Prediction vs. ground truth segmentation mask comparison.

- The model with 1 SPP block performed the best amongst the architectures proposed. Therefore, it can be deduced that the boosting of context and features go hand in hand.
- In the case of Hausdorff Distance, all the three models have the lowest metric when the class is Enhancing Tumor.
- In the case of Dice Score, both No SPP and 2 SPP models achieve similar results in Whole Tumor and Tumor Core. Both the models outperform 1 SPP in Whole Tumor however lose out to 1 SPP in Tumor Core. All the models achieve the lowest Dice Score in the case of Enhancing Tumor.
- When comparing models trained on Brats 2019 and Brats 2020, the proposed work outperforms the model however this is a general trend.
- With respect to Brats 2021, the proposed work gives comparable results to Hatamizadeh et al. (32) however loses out to Jia and Shu (22) on a large margin. One thing to note, both of these models use transformers which naturally provide more context and features. Transformers are heavy on parameters, while the proposed approach requires no extra parameters.

The model was also trained on image size of 160 × 160 × 160. This was done to understand the impact of a smaller image size. Table 4 conveys the same.

The following inferences can be made from Table 4:

- 0.01 was the difference in average dice score between the models trained on different image sizes. However, a significant difference was observed in the Hausdorff distances.
- Again 1 SPP model performed the best but the margin of difference was next to none. Hence we can infer that a high image size is a key contributor to increase performance when SPP is utilized and the following inference proves the point of passing higher resolution features through residual connections.

The trend in the metrics can be seen as shown in Figure 4. Figure 4A represents the Dice Score vs. Epochs and Figure 4B portrays the Hausdorff Scores vs. Epochs. Regardless of the training image size, the models carry a similar trend where the dice scores plateau at 60 epochs. Secondly, the performance of models trained on 160 × 160 × 160 are lower than the models trained on 160 × 192 × 192. Moreover, it can also be observed that the convergence of the loss is delayed for models with SPP. Although SPP does not bring any extra trainable parameters it still keeps the model from converging. Based on the trend, the model can be fine-tuned at extremely small magnitudes of learning rates to increase the performance of the model. In the case of Figure 4B it can be observed that models without SPP tend to provides metric stability once the model reaches the last few epochs.

Lastly, the output maps from the model are analyzed in Figure 5. For the analysis three slices from different MRI scans are taken wherein each slice exhibits varying presence of class. The first row within the plot contains a sparse volume of enhancing tumor, while the second row contains a moderate volume of enhancing tumor. The last row majorly contains enhancing tumors. Based on the predictions it can be observed that the model is able to predict all types of cases with great accuracy. The reason for the visualization is to showcase the model's ability to predict enhancing tumors accurately as its presence in data is limited. Moreover, the model is also able to detect abnormal whole tumor shapes with ease which conveys that the model is fit for real world diagnosis. In the first scan the model is able to predict sparse presence of enhancing tumors which is very crucial.

4. Conclusion

We propose U-Net with SPP and Attention Residual Connections in this work. The proposed model attachment is a lightweight mechanism which boosts information and context in the model by passing high and low resolution information to the decoders in the Unet. The proposed mechanism is applied to the NvNet model in varying frequencies which then produces different variants: Model with attention, Model with attention and 1 SPP, and Model with attention and 2 SPP. The model with 1 SPP and attention performs the best and provides comparable results to heavy models with transformer residual attachments. The average Dice Score and Haussdorf distance for the model with 1 SPP and attention are 0.883 and 7.99, respectively. The proposed mechanism is an approach to boost information and context hence giving considerable performance boosts. This approach plays well in applications such as edge computing which requires a balance of computational efficiency and performance. Such an approach could be utilized in mobile healthcare stations which need immediate diagnosis with less computation power. However, the impact of performance improvement at times falls a bit short compared to heavy approaches and it boils down to the extra trainable parameters brought by the components which eventually capture more patterns. In the current work, the mechanism is only adapted to one particular model and in the future, we aim to make the mechanism adaptable to various other 3D-Unet architectures.

Data availability statement

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

Author contributions

KS and C-YC conceptualized and supervised the research, carried out the project administration, and validated the results. SV and TG contributed to the development of the model, data processing, training procedures, and the implementation of the model. SV, TG, and KS wrote the manuscript. SV, TG, KS, PV, and C-YC reviewed and edited the manuscript. C-YC carried out the funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This research was partially funded by Intelligent Recognition Industry Service Research Center from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan and Ministry of Science and Technology in Taiwan (Grant No. MOST 109-2221-E-224-048-MY2).

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.

References

1. Baid U, Ghodasara S, Mohan S, Bilello M, Calabrese E, Colak E, et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv [Preprint]*. (2021).

2. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging*. (2014) 34:1993–2024. doi: 10.1109/TMI.2014.2377694

3. Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv [Preprint]*. (2018).

4. Srinivasan K, Selvakumar R, Rajagopal S, Velev DG, Vuksanovic B. Realizing the effective detection of tumor in magnetic resonance imaging using cluster-sparse assisted super-resolution. *Open Biomed Eng J*. (2021) 15:170–9. doi: 10.2174/1874120702115010170

5. Das TK, Roy PK, Uddin M, Srinivasan K, Chang CY, Syed-Abdul S. Early tumor diagnosis in brain MR images via deep convolutional neural network model. *Comp Mater Cont*. (2021) 68:2413–29. doi: 10.32604/cmc.2021.016698

6. Kanniappan S, Samiayya D, Vincent PMDR, Srinivasan K, Jayakody DN, Reina DG, et al. An efficient hybrid fuzzy-clustering driven 3D-modeling of magnetic resonance imagery for enhanced brain tumor diagnosis. *Electronics*. (2020) 9:475. doi: 10.3390/electronics9030475

7. Srinivasan K, Sharma V, Jayakody DNK, Vincent DR. D-ConvNet: Deep learning model for enhancement of brain MR images. *Basic Clin Pharmacol Toxicol*. (2018) 124:3–4.

8. Arabahmadi M, Farahbakhsh R, Rezazadeh J. Deep learning for smart healthcare—A survey on brain tumor detection from medical imaging. *Sensors*. (2022) 22:1960. doi: 10.3390/s22051960

9. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA) (2015). p. 3431–40.

10. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer (2015). p. 234–41.

11. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham: Springer (2018). p. 3–11.

12. Jiang Z, Ding C, Liu M, Tao D. Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task. In: *International MICCAI Brainlesion Workshop*. Cham: Springer (2019). p. 231–41.

13. Zhang Z, Liu Q, Wang Y. Road extraction by deep residual u-net. *IEEE Geosci Remote Sens Lett.* (2018) 15:749–53. doi: 10.1109/LGRS.2018.2802944

14. Baid U, Talbar S, Rane S, Gupta S, Thakur MH, Moiyadi A, et al. Deep learning radiomics algorithm for gliomas (drag) model: a novel approach using 3D U-Net based deep convolutional neural network for predicting survival in gliomas. In: *International MICCAI Brainlesion Workshop*. Cham: Springer (2018). p. 369–79.

15. Kaur A, Kaur L, Singh A. GA-UNet: UNet-based framework for segmentation of 2D and 3D medical images applicable on heterogeneous datasets. *Neural Comp Appl.* (2021) 33:14991–5025. doi: 10.1007/s00521-021-06134-z

16. Punn NS, Agarwal S. Multi-modality encoded fusion with 3D inception U-net and decoder model for brain tumor segmentation. *Multimed Tools Appl.* (2021) 80:30305–20. doi: 10.1007/s11042-020-09271-0

17. Isensee F, Jäger PF, Full PM, Vollmuth P, Maier-Hein KH. nnU-Net for brain tumor segmentation. In: *International MICCAI Brainlesion Workshop*. Cham: Springer (2020). p. 118–32.

18. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods.* (2021) 18:203–11. doi: 10.1038/s41592-020-01008-z

19. Qamar S, Ahmad P, Shen L. Hi-net: hyperdense inception 3d unet for brain tumor segmentation. In: *International MICCAI Brainlesion Workshop*. Cham: Springer (2020). p. 50–7.

20. Liu Q. iSegFormer: interactive image segmentation with transformers. *arXiv [Preprint].* (2021).

21. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell.* (2015) 37:1904–16. doi: 10.1109/TPAMI.2015.2389824

22. Jia Q, Shu H. Bitr-unet: a cnn-transformer combined network for mri brain tumor segmentation. In: *International MICCAI Brainlesion Workshop*. Cham: Springer (2022). p. 3–14.

23. Wang J, Lv P, Wang H, Shi C. SAR-U-Net: squeeze-and-excitation block and atrous spatial pyramid pooling based residual U-Net for automatic liver segmentation in computed tomography. *Comput Methods Programs Biomed.* (2021) 208:106268. doi: 10.1016/j.cmpb.2021.106268

24. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans Pattern Anal Mach Intell.* (2017) 40:834–48. doi: 10.1109/TPAMI.2017.2699184

25. Ahmad I, Qayyum A, Gupta BB, Alassafi MO, AlGhamdi RA. Ensemble of 2D residual neural networks integrated with atrous spatial pyramid pooling module for myocardium segmentation of left ventricle cardiac MRI. *Mathematics.* (2022) 10:627. doi: 10.3390/math10040627

26. Jiang Y, Ye M, Huang D, Lu X. AIU-Net: an efficient deep convolutional neural network for brain tumor segmentation. *Math Prob Eng.* (2021) 2021. doi: 10.1155/2021/7915706

27. Wang W, Chen C, Ding M, Yu H, Zha S, Li J. Transbts: Multimodal brain tumor segmentation using transformer. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer (2021). p. 109–19.

28. Tampu IE, Haj-Hosseini N, Eklund A. Does anatomical contextual information improve 3D U-Net-based brain tumor segmentation? *Diagnostics.* (2021) 11:1159. doi: 10.3390/diagnostics11071159

29. Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization. In: *International MICCAI Brainlesion Workshop*. Cham: Springer (2018). p. 311–20.

30. Milletari F, Navab N, Ahmadi SA. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. Stanford, CA: IEEE (2016). p. 565–71.

31. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell.* (1993) 15:850–63. doi: 10.1109/34.232073

32. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: *International MICCAI Brainlesion Workshop*. Cham: Springer (2022). p. 272–84.

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 03 March 2023
DOI 10.3389/fpubh.2023.1124998

OPEN ACCESS

EDITED BY
Steven Fernandes,
Creighton University, United States

REVIEWED BY
Akriti Nigam,
Birla Institute of Technology, Mesra, India
Fnu Bilal,
University of Texas Southwestern Medical Center, United States
Nidhi Arora,
University of Delhi, India
Neetu Kushwaha,
The Alan Turing Institute, United Kingdom

*CORRESPONDENCE
Arti Tiwari
✉ atiwari1@as.iitr.ac.in
Kamanasish Bhattacharjee
✉ kb020sayan@gmail.com
Millie Pant
✉ pant.milli@as.iitr.ac.in

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 15 December 2022
ACCEPTED 10 February 2023
PUBLISHED 03 March 2023

CITATION
Tiwari A, Bhattacharjee K, Pant M, Srivastava S and Snasel V (2023) An AI-enabled research support tool for the classification system of COVID-19. *Front. Public Health* 11:1124998. doi: 10.3389/fpubh.2023.1124998

An AI-enabled research support tool for the classification system of COVID-19

Arti Tiwari[1]*, Kamanasish Bhattacharjee[2]*, Millie Pant[1,3]*, Shilpa Srivastava[4] and Vaclav Snasel[5]

[1]Department of Applied Mathematics and Scientific Computing, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India, [2]Machine Intelligence in Medicine and Imaging (MI-2) Lab, Mayo Clinic, Phoenix, AZ, United States, [3]Mehta Family School for Data Science and Artificial Intelligence, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India, [4]CHRIST (Deemed to be University) Delhi NCR, Ghaziabad, India, [5]Department of Computer Science, VŠB-Technical University of Ostrava, Ostrava, Czechia

The outbreak of COVID-19, a little more than 2 years ago, drastically affected all segments of society throughout the world. While at one end, the microbiologists, virologists, and medical practitioners were trying to find the cure for the infection; the Governments were laying emphasis on precautionary measures like lockdowns to lower the spread of the virus. This pandemic is perhaps also the first one of its kind in history that has research articles in all possible areas as like: medicine, sociology, psychology, supply chain management, mathematical modeling, etc. A lot of work is still continuing in this area, which is very important also for better preparedness if such a situation arises in future. The objective of the present study is to build a research support tool that will help the researchers swiftly identify the relevant literature on a specific field or topic regarding COVID-19 through a hierarchical classification system. The three main tasks done during this study are data preparation, data annotation and text data classification through bi-directional long short-term memory (bi-LSTM).

KEYWORDS

COVID-19, long short-term memory, classification, bi-directional LSTM, Artificial Intelligence

1. Introduction

Early in the year 2020, the outbreak of COVID-19 created havoc around the world, leading to mental trauma, shattered economies and, above all, the loss of human life. While the researchers and scientists were trying to understand more about the virus and a possible antidote/vaccine for it, the challenge for the Government was to keep its people safe by enforcing preventive measures like lockdowns. The uncertainty of the situation affected almost all sections of society. Despite all this grimness, the scientific and research community was doing its bit through experiments and observations and publishing research articles and reports on its basis. The COVID pandemic, perhaps, also is the first case of its kind that provoked research in all possible dimensions. Although the situation is not alarming anymore, with people getting vaccinated and economies getting back on pace, the research on COVID-19 is still continuing, and a noticeable quantity of research articles are being published.

The internet now contains a plethora of literature dedicated to the various aspects of COVID-19 ranging from studies related to lab experiments to clinical studies to vaccines and drug development to diagnostic techniques and many more. There are several studies dedicated to economics and mathematical models, forecasting methods to estimate the spread of the virus, supply chain models and several others.

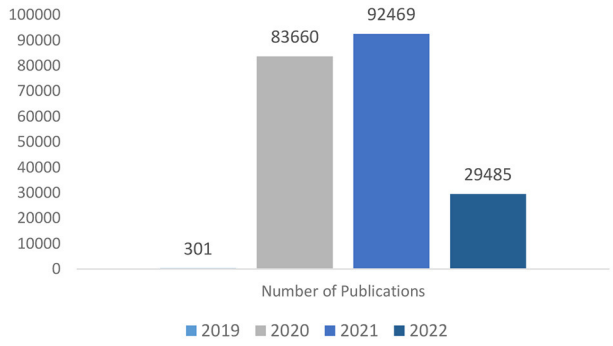


FIGURE 1
Year-wise number of publications listed in CORD-19 dataset.

1.1. Bibliometric analysis

A selected bibliometric analysis was performed on the CORD-19 dataset for articles related to COVID-19 which were later used for model training and database development. The results are obtained to show the trend of publications for COVID-19 articles and the "*terms*" used in the paper to label the classes.

Figure 1 shows that in 2019, at the onset of COVID-19, the publications were 301 in number, which raised to 83,660 in 2020 and further raised to 92,469 in 2021 and although in 2022, the number of articles became 29,485, the trends are good enough to indicate that the research is still continuing in this area with new research papers being published from time to time.

Figure 2 shows a network visualization created using Vosviewer (https://www.vosviewer.com/). The network map includes the terms/items (object of interest) represented by a circle driven by the title and abstract of the selected articles and the links between the terms based on their pair-wise occurrence. The higher the occurrence of an item, the bigger the circle. In this map total of 612 terms are selected and grouped into four non-overlapping clusters. Cluster-one (red) consists of 223 terms, cluster-two (green) contains 186 items, cluster-three (blue) incorporate 149 items, and cluster-four (yellow) contains 54 terms.

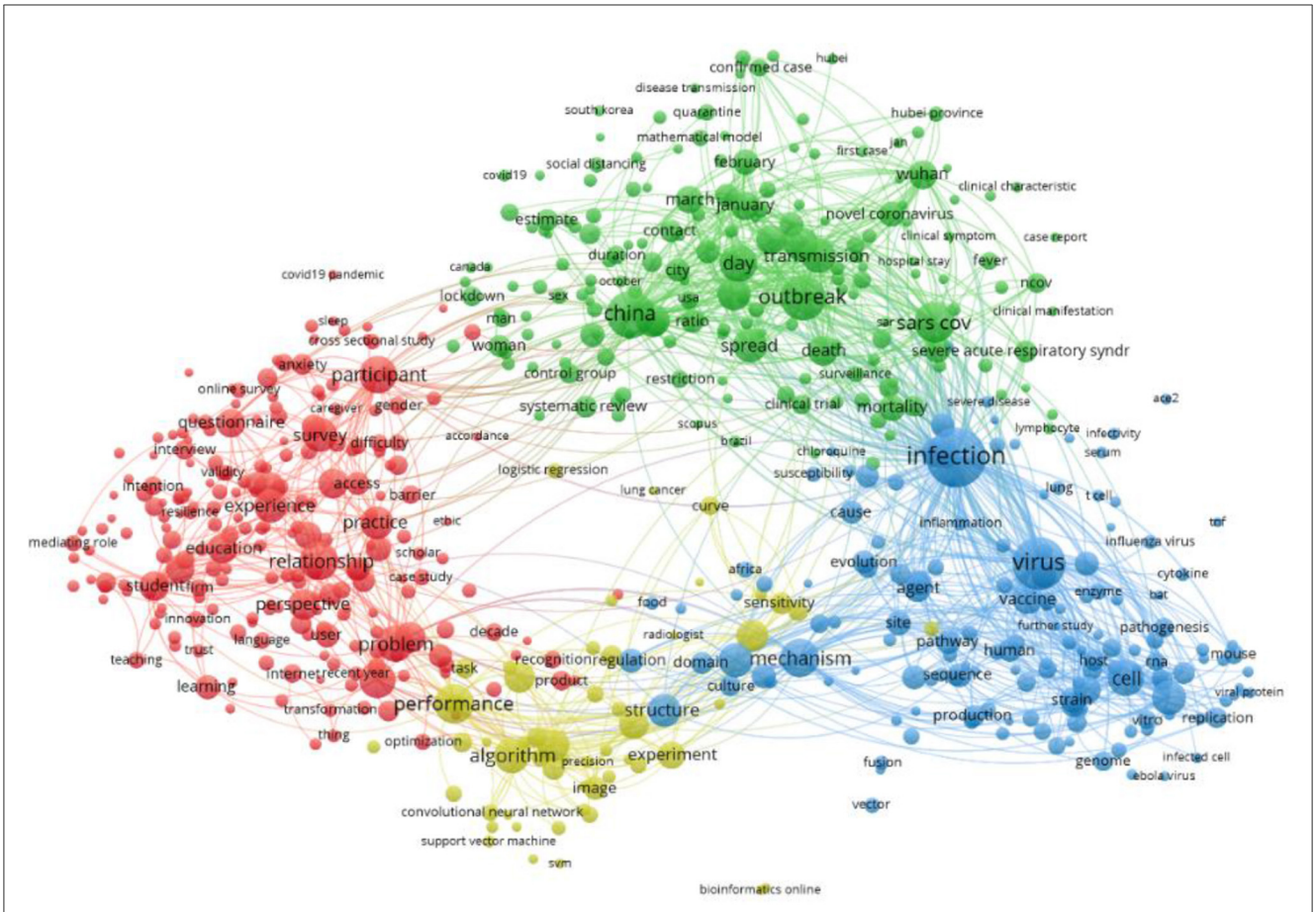


FIGURE 2
Network visualization map based on Abstract and Title.

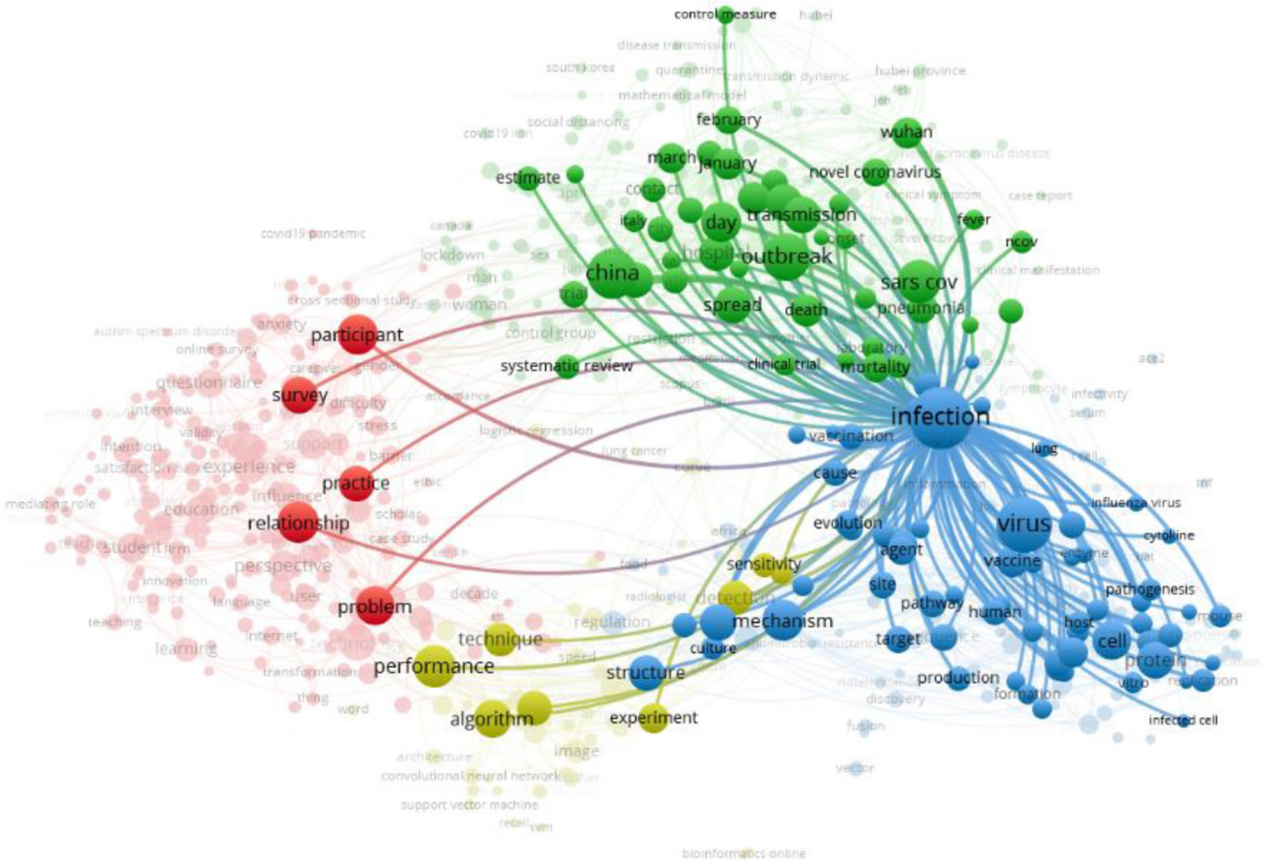


FIGURE 3
Network visualization map of the term "infection".

In Figure 2, the term "infection" is depicted with the biggest circle, as this term shares the highest co-occurrence with the other terms.

Figure 3 describes the network visualization map of the term/item "infection," which possesses the highest occurrence value and link strength value as 433 and 5,014, respectively. The link strength value shows the number of articles where two terms occurred together.

On the basis of the publication years of the selected articles, an overlay visualization map is created in Figure 4. This visualization of this map is identical to the network map, however, its interpretation is based on the score of the average publication year.

This map shows the focus shifts on the area of research publication throughout the year 2019 to the year 2021. Since early to mid-2019, the published articles were subjected toward the infection, virus, and vaccine. From mid-2019 to mid-2020 the published articles were tend toward China, the outbreak, SARS-CoV, and its spread. After mid-2020, the articles are focused on problem-solving, algorithms, perspective, experiments and performance.

1.2. Need for a dedicated research support tool

The graphs given above clearly indicate, how the research is growing in the area of COVID-19. These graphs also show that there are several categories (fields) of research and every category can be further divided into sub-categories (subfields). For a new researcher, digging into this plethora of information can be quite overwhelming. It becomes difficult for a researcher to identify the correct literature relevant to one's area of interest. This difficulty may be eased to some extent if there is a dedicated platform which can easily guide them to their area of interest. In the literature, very few dedicated research support tools are available as per the authors understanding. The closest works to this study can be found in Simon et al. (1). Here the authors have presented a text mining based tool called BioReader for the classification of Biomedical research. In (2), R-classify is a web tool developed by Aggarwal et al. to help users in finding out the relevant literature in the area of Computer Science. Doty et al. (3) developed a python-based graphical user interface to conduct the classification and visualization of electron microscopy data.

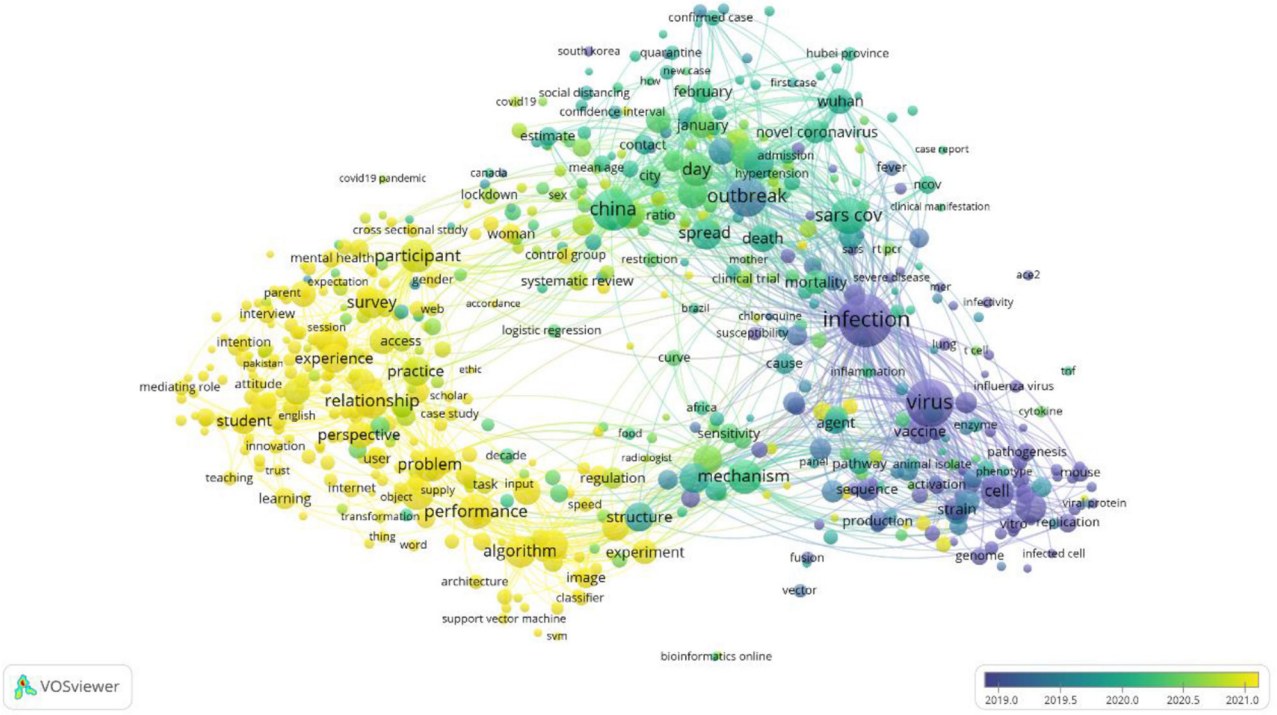


FIGURE 4
Overley visualization of the articles based on publication year.

In the present article, an Artificial Intelligence (AI) enabled automatic classification tool called Research Support Tool (RST) is developed for COVID-19-related literature. Since the problem is of text (literature) classification, a Bi-LSTM neural network is used. The Bi-LSTM model is trained on the abstract and title of the selected articles. The articles are taken from the CORD-19 dataset and are divided into seven categories (class labels) based on their subjects. The RST is developed using IONIC and Angular framework. Remaining of the article consists of three more sections. In Section 2, the methodology followed in the present study is described. In Section 3, the user interface is presented, and the workflow is defined. Finally, Section 4 provides the concluding remarks and also provides some future directions in which the work can be extended.

2. Materials and methodology

The work done in this study can be divided into four major steps, which start from data collection to its preparation to its labeling and finally to its classification. The steps are defined below in Figure 5.

2.1. Step 1: Data collection

The first step in this study is the collection of data for which the COVID-19 Open Research Dataset or CORD-19 (4) was selected. It is curated by Allen Institute for AI (AI2) and is available on Kaggle (5) as well as on AI2's Semantic Scholar website (6). This database is periodically updated. At the time of the preparation of this article, it contained resources for almost 4,00,000 scholarly articles, including over 2,45,000 full-text articles on COVID-19, SARS-CoV-2, and variants of coronaviruses.

2.2. Step 2: Data preparation

Once the data source has been identified, the next step is to prepare the data for further usage. To make the study more relevant in terms of the COVID scenario, only the studies subjected to COVID-19 or SARS-CoV-2 were considered. This was done by using the keywords like "COVID-19," "Wuhan," "Hubei," "SARS-CoV-2," "2019 novel coronavirus," "2019-nCoV," "coronavirus disease 2019," "corona pandemic," "coronavirus outbreak," and their combinations and filtering out the studies not meeting up with our criteria. Initially, 4,532 articles were selected based on different subjects, after filtering out the inconsistent, incomplete data, a total of 3,011 articles are taken for the model training and database development.

2.3. Step 3: Data labeling

The third step, and also one of the key tasks of this study, was to label the articles, which can be classified later as per the machine learning algorithms. The literature was segregated into seven major

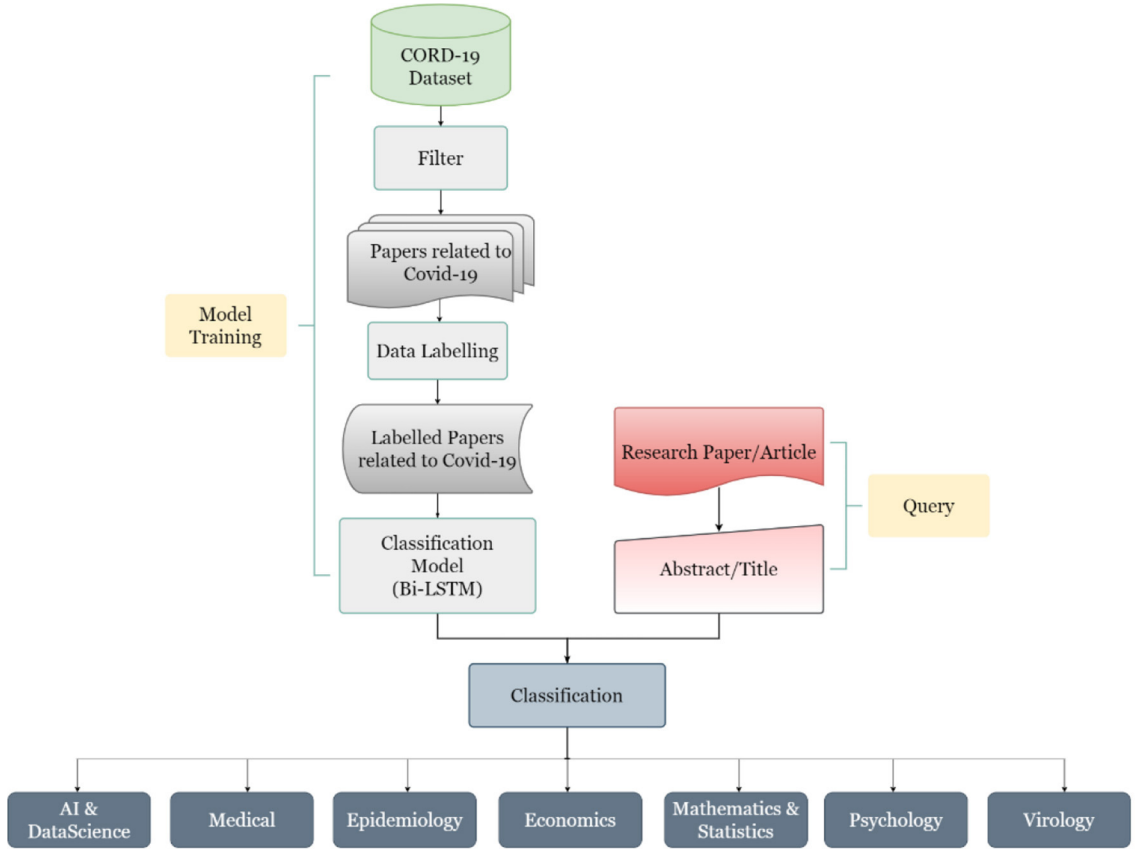


FIGURE 5
The workflow diagram of the AI-enabled research support tool process.

classes per the experts' discussion. These seven classes are Artificial Intelligence (AI) and Data Science, Economics, Epidemiology, Mathematics and Statistics, Medical, Psychology and Virology. A brief description of the classes is shown in Table 1, and the subclasses of the selected articles are shown in Figure 6.

In the dataset created for this work, each data contains the title, abstract, and class label of the literature. The data distribution among the selected seven categories is shown in Figure 7.

2.4. Step 3: Classification

The AIRST developed in the present study is based on the classification of text, for which the Bi-directional long short-term memory (Bi-LSTM) neural network (43) is implemented. Vanilla neural networks are not found to be suitable for texts as these are unable to process the sequences.

Recurrent neural networks (RNN), have a loop-like architecture which allows the information to persist. RNNs have been successfully applied to various areas including speech recognition, speech synthesis, language translation, image captioning and many more (44–46). However, in the case of sequential data, it sometimes becomes susceptible to vanishing gradient due to long-term dependency. The problem of vanishing gradient can be resolved with the help of LSTMs (47), a type of RNN which are capable of learning long-term dependencies. The LSTM models are made up of cell states and various gates. While the cell state in LSTM acts like a memory of the network and transfers relevant information down the sequence chain model; gates are the neural networks that decide the information to be retained and the information to be forgotten during training. An LSTM model consists of three gates viz. forget gate, input gate, and output gate. These gates are described in brief as follows.

2.4.1. Forget gate

The first step of the LSTM cell is to retain the relevant information and to discard the information that is not of significance. This is done with the help of the sigmoid layer known as the *"forget gate layer."* The activation value for the forget gate can be given as:

$$f_t = \sigma(w_f[h_{t-1}, x_t] + b_f) \quad (1)$$

where x_t is input vector at timestamp t $h_{t-1}$is a hidden state or output of the previous timestamp, w, b represent the weight and deviation matrix, respectively.

The sigmoid function normalizes all the activation values between 0 and 1. The value 0 implies all forgotten, and the value 1 implies nothing forgotten.

TABLE 1 A description of class labels categorization.

Class	Description	References
Artificial Intelligence (AI) and Data Science	This class is divided into five subclasses of AI and Data Science—machine learning, deeplearning, social media infodemic, thematics analysis, and big data analysis for selecting the related articles	(7–15)
	The articles that deal with AI and Data Science being used for automatic screening of COVID-19 using computer tomography scans and X-ray images of the lungs of patients, prediction and forecasting of virus spread, mortality risk etc. It is further subdivided into Machine Learning and Deep Learning, Data Mining, Data Analysis methods for social media infodemic, misinformation spreading, patient report analysis, sentiment analysis, infoveillance, and information on datasets which are relevant to deal with COVID-19 are classified under this category	
Economics	This class has four subcategories—industrial organization, economic system, stock market, public economy and government spending. The articles that belong to these categories discuss the consequences of COVID-19 on the economy of a country, the economy of a specific product, the economy of a segment of the market, and stock markets are classified under this category	(16–20)
Epidemiology	This class considers three subcategories—transmission modeling, disease surveillance, and occupational epidemiology. The articles categorized in this class deal with outbreak control measures, the effect of COVID-19 on various occupations and environments, risk assessment, transmission monitoring, transmission pattern recognition, analysis and forecasting are classified under this category	(21–25)
Mathematics and Statistics	This class considers three subcategories—data-based analysis, mathematical modeling, and forecasting. Articles that explain how mathematical modeling and statistical analysis are utilized to predict the transmission and spread of COVID-19 and also to identify the mitigation strategies are classified under this category	(26–30)
Medical	Diagnosis, therapeutics, pharmaceuticals, pediatrics, oncology, neurology, and anesthesiology are the subclasses of class medical. The papers dealing with COVID-19 diagnosis, therapeutics, immunology, pharmacology, anaesthesiology, oncology, neurology, pediatrics, hematology etc. medical related issues are classified under this category	(31–34)
Psychology	Two subcategories—health psychology, and neuropsychology are considered for selecting the articles that belong to this class. The papers that discuss the impact of the COVID-19 epidemic on the mental health and psyche of human beings and their behavior are classified under this category	(35–39)
Virology	There are four different subclasses—viruses, viral disease, viral protein, and viral life cycle are considered for this class. The papers with research work on the virus structure, genome, molecular characterization, and mutation are classified under this category	(40–42)

2.4.2. Input gate

The second step in an LSTM model is to identify the information that will be stored in the state of a cell. The input gate layer quantifies the crucial information carried by the input. This step is further divided into two parts. First, an "input gate layer" (sigmoid layer) decides the values to be added to the cell state C_t and then, a tan h layer derives a vector of new candidate value N_t, that has to be added to the state. This is followed by the combination of the aforementioned steps to update the state. The input gate activation value is as follows:

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \tag{2}$$

where, x_t is input vector at timestamp t, $h_{t-1}$is a hidden state or output of the previous timestamp, w, b represent the weight and deviation matrix, respectively.

N_t is defined as:

$$N_t = \tanh(w_c[h_{t-1}, x_t] + b_C) \tag{3}$$

Cell state is updated as:

$$C_t = f_t^* C_{t-1} + i_t^* N_t \tag{4}$$

Where, C_{t-1} is the previous cell state.

2.4.3. Output gate

The objective of the output gate is to decide the output which in turn will be n the basis of the state of the cell. Here, a sigmoid layer identifies the part of the cell state that will be the output. This information is further processed by passing the cell state through the activation function tan h and multiplying it with the output of the sigmoid gate. Finally, the output h_t is obtained as:

$$O_t = \sigma(w_o[h_{t-1}, x_t] + b_o)h_t = O_t^* \tanh(C_t) \tag{5}$$

2.4.4. Bi-directional long short-term memory

Bi-directional long short-term memory (Bi-LSTM) is an extended and improved version of LSTM; it is an integration of

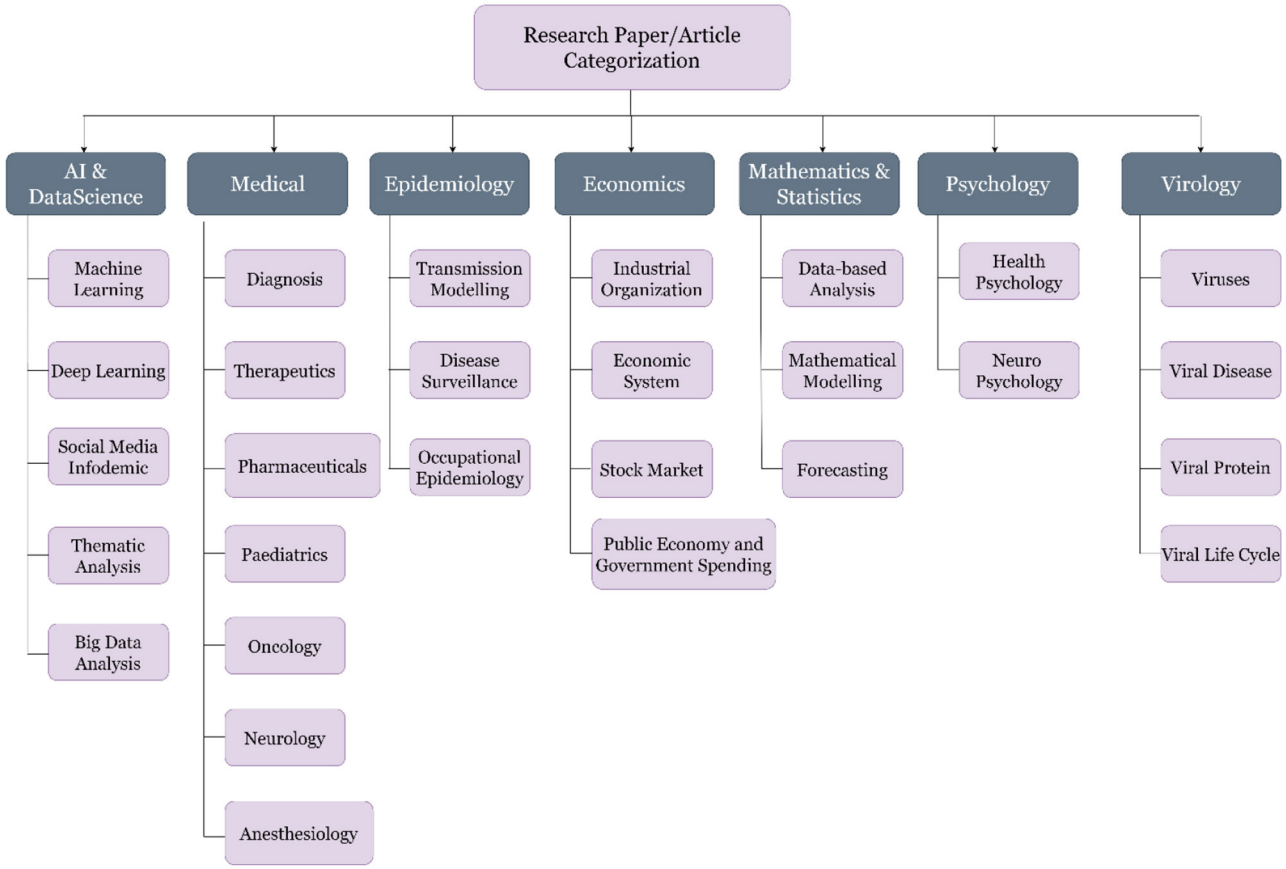


FIGURE 6
Categorization and sub-categorization of the articles selected for the study.

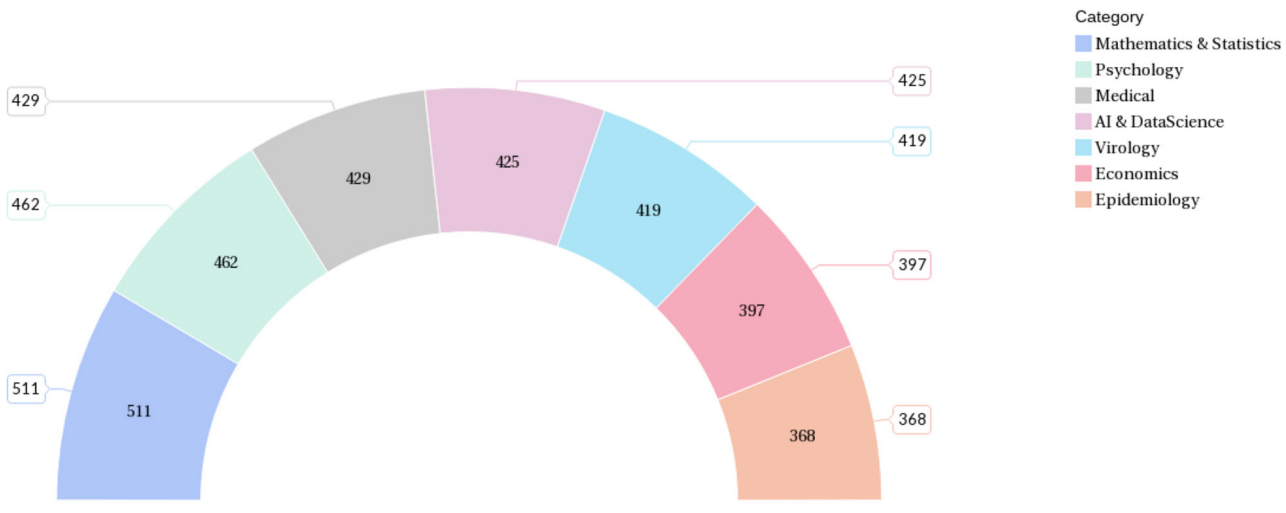


FIGURE 7
Number of data samples (articles) assigned to the class categories.

two independent RNN models. Unlike unidirectional LSTM, in Bi-LSTM, the information flows in both directions: backward as well as in the forward direction. This is illustrated in Figure 8.

Bi-LSTM exploits the information about the sequence in both directions at every timestamp by connecting two hidden layers to the same output. The management of the past and future

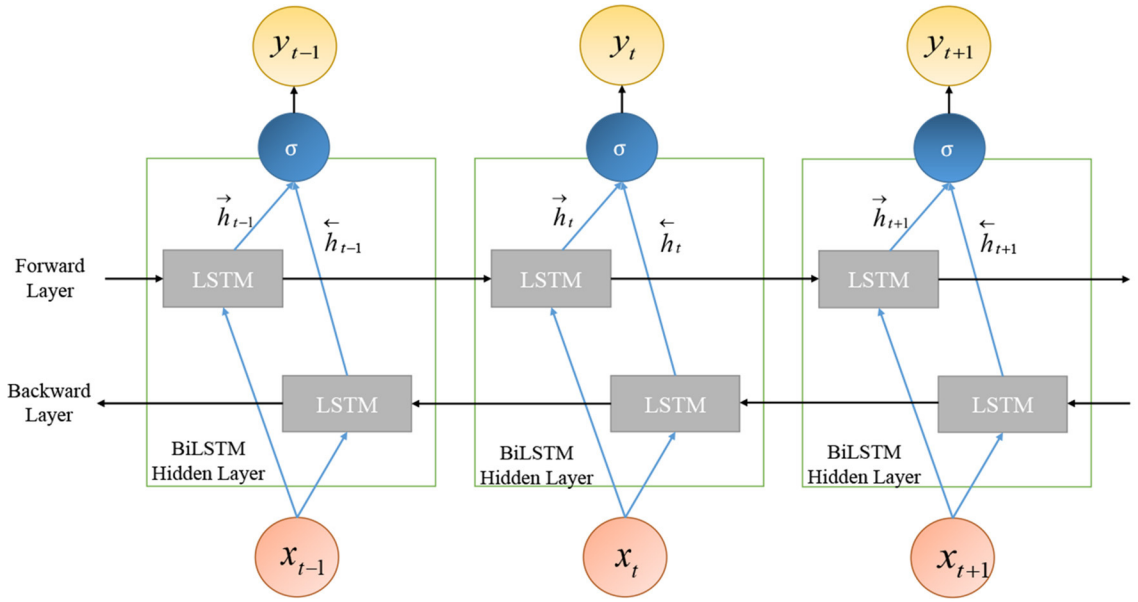


FIGURE 8
Bi-directional long short-term memory (Bi-LSTM).

TABLE 2 Parameters of Bi-LSTM model architecture.

Parameters	Size
Embedding layer vocab size	10,000
Embedding dimension	64
Maximum length of a unique word	200
Bi-LSTIM size	32
Batch size	64

information, for a sequence, leads to better predictions for Bi-LSTM. The output of the hidden layer of Bi-LSTM is made up of the activation output of forward as well as backward hidden layers:

$$\vec{h}_t = \sigma(W_{xh}^{\rightarrow} x_t + W_{hh}^{\rightarrow\rightarrow} \vec{h}_{t-1} + b_h^{\rightarrow}) \quad (6)$$

$$\overleftarrow{h}_t = \sigma(W_{xh}^{\leftarrow} x_t + W_{hh}^{\leftarrow\leftarrow} \overleftarrow{h}_{t-1} + b_h^{\leftarrow}) \quad (7)$$

$$H_t = W_{xh}^{\rightarrow} \vec{h} + W_{hy}^{\leftarrow} \overleftarrow{h} + b_y \quad (8)$$

where, H_t represents the hidden layer, and its output includes the forward layer output $\vec{h}_t$ and backward layer output $\overleftarrow{h}_t$.

The Bi-LSTM model was trained on a total of 3,011 samples of seven different categories of research articles related to COVID-19 that are collected from the CORD-19 dataset. The parameters of the Bi-LSTM model architecture are mentioned in Table 2.

The final layer of the model is the Dense output layer with seven neurons representing the total number of class labels and Softmax activation function. To avoid overfitting while training the model, each layer is followed by the Dropout layer with an alpha value as 0.35.

2.5. User interface and workflow

The workflow of the research support tool has two components—the objective of the user interface development and the cloud environment-based application development tools.

2.5.1. Objective

A research support tool has been designed to meet the following three primary objectives:

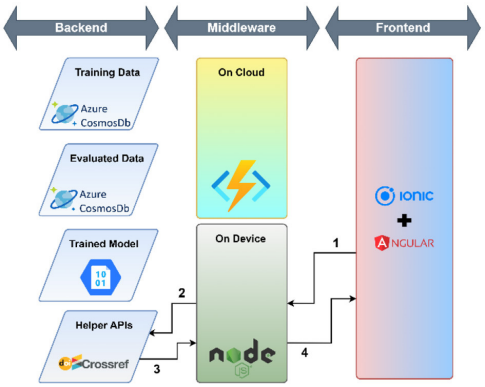
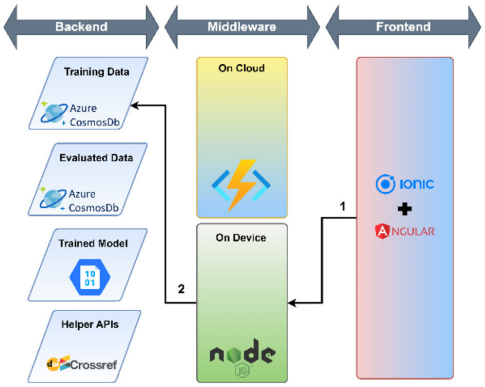
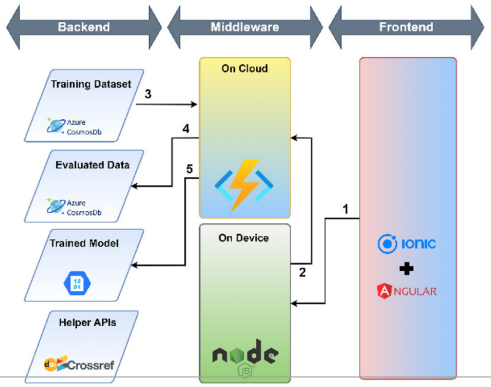
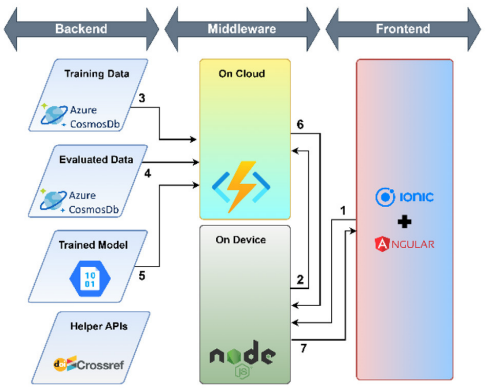
1. Enable users to view COVID-19-related research papers and articles under different categories. The users are also enabled to filter and search for research papers based on the title of the research papers.
2. Enable users to categorize an article not available in the dataset. The user can do that by providing DOI and proceeding after checking the extracted title and abstract.
3. Enable users to contribute to the labeled dataset by providing the title and abstract of the research paper and assigning a category manually.

2.5.2. Application development tool

A cloud environment-based application was developed that used a micro-service architecture to meet the mentioned requirements. The following technology stack was selected to develop the tool:

1. Azure Cloud platform—Azure Cosmos DB (NoSQL) and Azure Cloud Functions were used for storing and retrieving data, executing the Python script to categorize research papers based on the trained model.
2. Ionic + Angular—Ionic and Angular frameworks were used to develop the user interface because of easily available

TABLE 3 A detailed workflow of the developed user interface.

Use of helper APIs 	The application utilizes DOI to ensure the uniqueness of documents in the training and evaluation dataset. CrossRef APIs are used to make the user experience smoother for the end-user. Since the CrossRef API does not guarantee that the abstract will be available in all DOIs, or the veracity of the abstract, the end-user can make adjustments to the abstract to ensure it is correct. The following steps are executed— 1. The user enters the DOI in the textbox provided. Once the user clicks on the check button, the NodeJS service picks the DOI 2. The DOI is passed the CrossRef Works API 3. The CrossRef database returns the details of the work in a semi-structured JSON 4. The Title and Abstract from the response are extracted and displayed to the end user
Load data for Training 	Once the user has entered the DOI, the title and abstract have been verified, the user can select the category. The following steps are executed— 1. The user ensures the title and the abstract are correct and fall under the selected category from the drop-down. When the user clicks on the "Submit for Training" button the details are passed to the NodeJS Service 2. The NodeJS service ensures that the data provided is in the correct structure. If the same DOI is present in the training dataset, the service will overwrite the record. Otherwise, the service will create a new record in the training dataset
Model Training 	The following steps are executed— 1. The user clicks on the "Retrain Model" button on the "Submit New" page which triggers the NodeJS service 2. The NodeJS service authenticates and triggers an Azure Cloud Function to retrain the model 3. The Azure Cloud Function fetches all the records from the training dataset 4. The Azure Cloud Function deletes all the records from the evaluation dataset 5. The Azure Cloud Function converts the training dataset in the required format and trains the model and stores it in the Azure Blob Storage service
Evaluation against the model 	The following steps are executed 1. Once the user ensures that the DOI, Title and Abstract are correct and clicks on the "Get Category" button, the details are passed to the NodeJS Service 2. The NodeJS service authenticates with the Azure Cloud function and passes the DOI, Title and Abstract as arguments to the Azure Cloud Function 3. The Azure Cloud Function checks whether the DOI is present in the training dataset. If the DOI is present in the Training Dataset, the cloud function will return the category present in the training dataset 4. The Azure Cloud Function checks whether the DOI is present in the evaluation dataset. If the DOI is present in the Evaluation Dataset, the cloud function will return the category present in the Evaluation Dataset 5. Steps 3 and 4 are done to reduce unnecessary computation against the model since it is a computationally expensive process. If the DOI is not present in either the training or evaluation dataset, the Azure Cloud Function will retrieve the model stored in the Azure Blob Storage and evaluate the category against the provided Title and Abstract 6. The Azure Cloud Function returns the evaluated Category to the NodeJS service and stores the DOI, Title, and Abstract along with the category in the Evaluation Dataset 7. The NodeJS service displays the evaluated Category against the given DOI, Title and Abstract

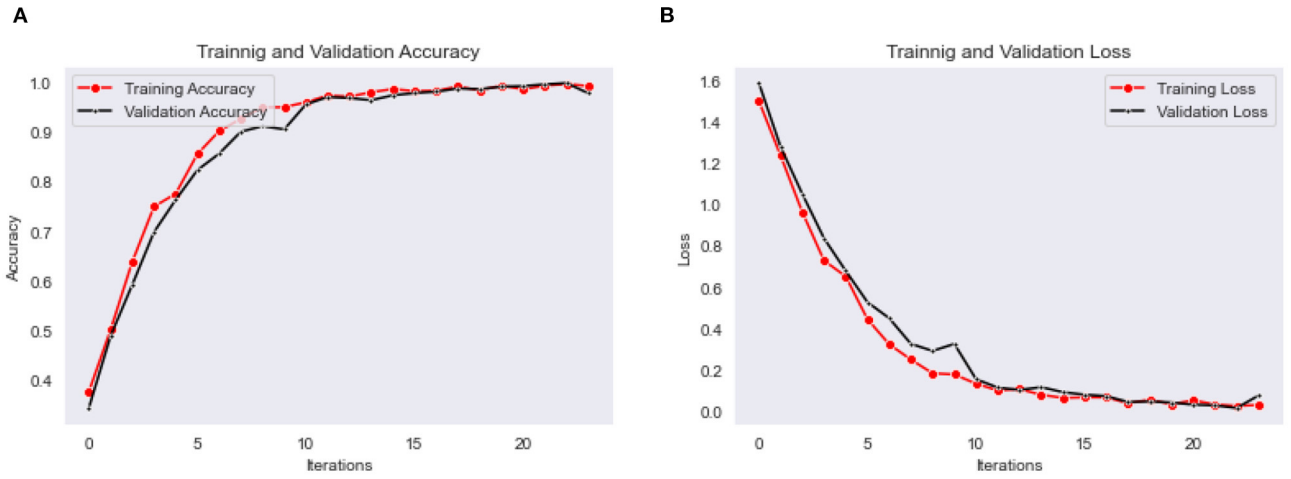


FIGURE 9
(A) Training and validation accuracy, **(B)** training and validation loss plots of Bi-LSTM classification model.

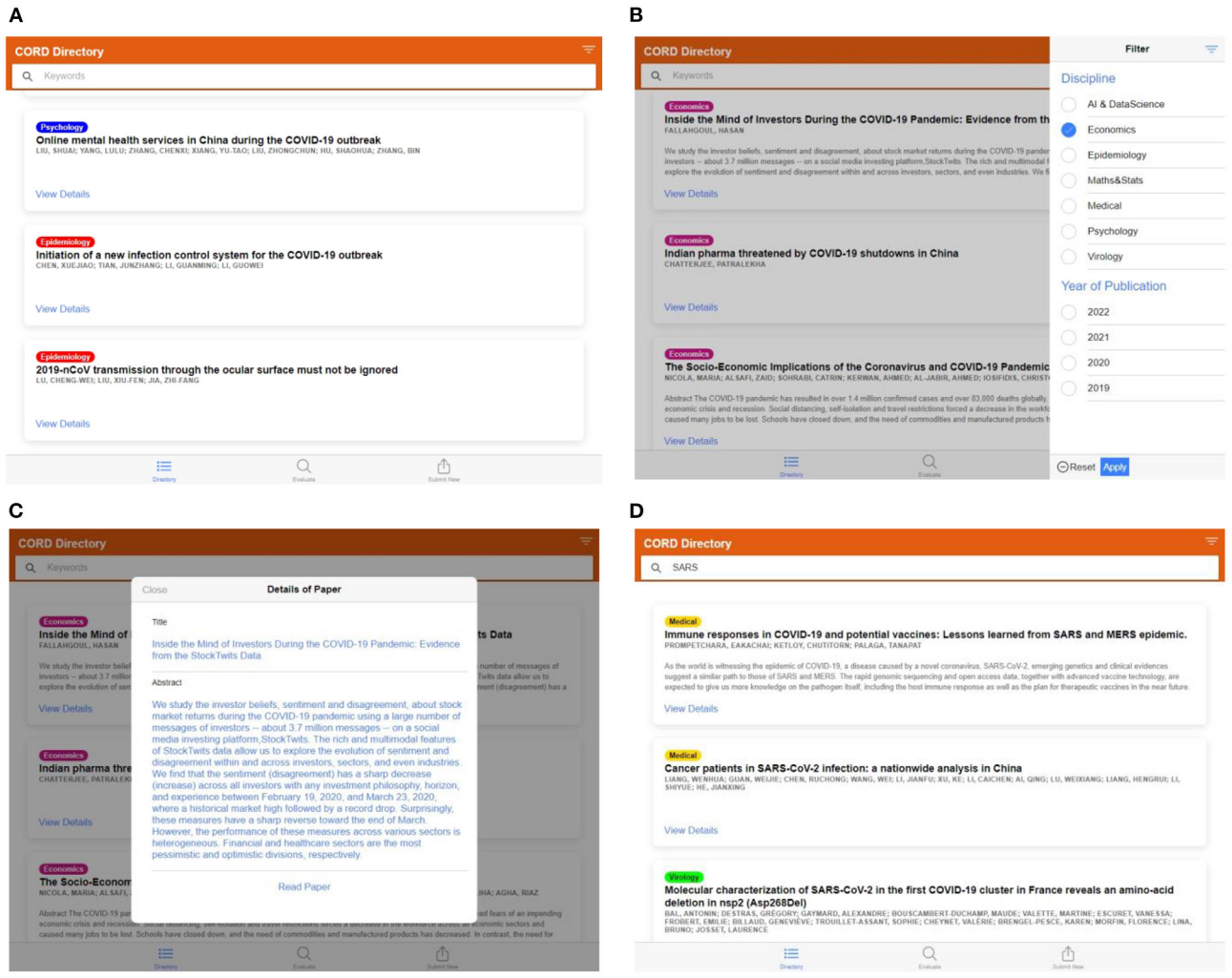


FIGURE 10
(A) Screen capture of the "Directory" tab—list of labeled articles in the database, **(B)** filter panel on the left side to select the articles of the particular category, **(C)** details of the listed articles, read paper tab will redirect to the original article page through DOI, **(D)** keyword tab can be used to search the article from the labeled dataset.

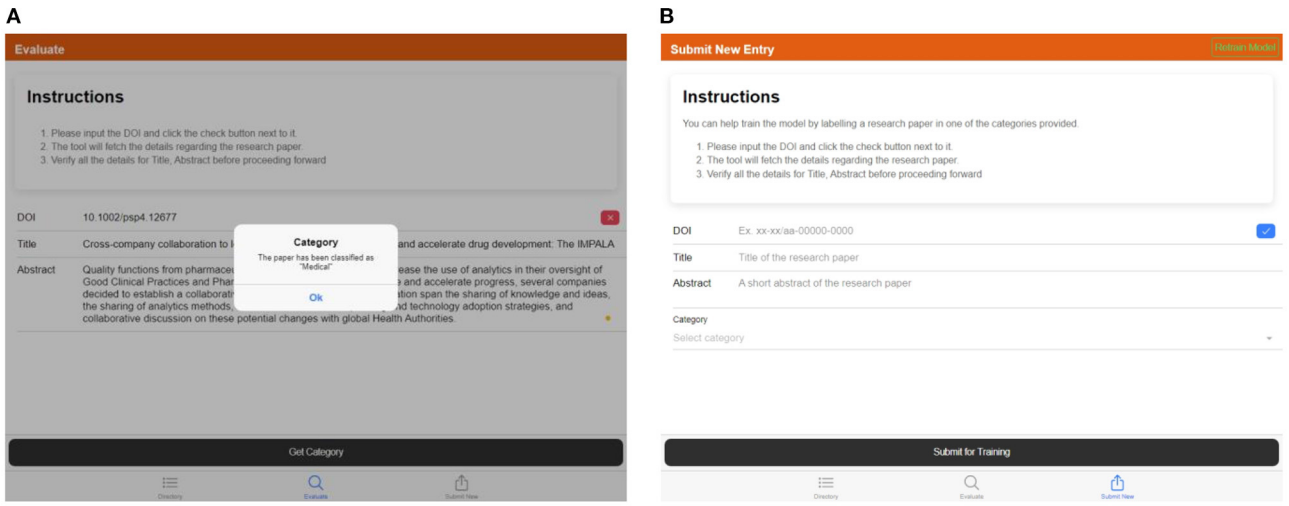


FIGURE 11
(A) Screen capture of the "Evaluate" tab—popup showing the category of the provided DOI, **(B)** screen capture of the "Submit New Entry" tab—details of the article can be provided by manually feeding the DOI, title, abstract, and the relevant category.

components and ability to deploy on multiple platforms such as Desktop, Mobile (Android and iOS using Cordova or Capacitor), Progressive Web Apps (PWA) and Cloud Hosted Web.

3. NodeJ—NodeJS middleware was used to access micro-services and respond to user interactions.

The workflow of the developed user interface consists of three parts: (1) use of helper APIs, (2) load data for training and training the model, (3) evaluation: evaluation again consists of two parts—the use of helper APIs and Evaluation against the model. The complete process of user interface workflow is shown in Table 3.

3. Results and analysis

The Bi-LSTM classification model is trained for the 25 epochs, and obtained maximum validation accuracy as 0.97, with a minimum validation loss as 0.015. The accuracy and loss for every epoch of training and validation are shown in graphs plots in Figures 9A, B, respectively.

The performance of the research support tool is presented through the screen captures of the developed user interface. Users can see the following view upon landing. The view is divided into three segments to meet the three objectives mentioned above. These segments can be accessed using the three tabs at the bottom of the interface.

1. The "Directory" tab is used to view, search and filter the research papers already categorized by the model. These include records from the training dataset and any records generated when a customer is evaluating a research paper using the model, shown in Figures 10A–D.
2. The "Evaluate" tab is used to provide the details of a research paper and categorize it using the trained model, shown in Figure 11A.
3. The "Submit New Entry" tab is used to manually label any research paper and add it to the training dataset. This will allow us to grow the training dataset and re-train the model periodically, as shown in Figure 11B.

The user interface requests DOI to enable CrossRef API to get details regarding the research paper, such as the title and abstract.

4. Conclusion and future scope

This work primarily intends to communicate the idea of developing a Research Support Tool for researchers around the world. The conclusive statements can be drawn from this study as shown below:

- The researchers can leverage this tool to delve deeper into COVID-19 research and make the relevant literature identification smoother.
- A multi-platform graphical user interface is developed to fulfill the primary objectives of extracting the COVID-19 related articles effortlessly and classifying them based on the particular research area.
- The classification system uses the Bi-LSTM model, which enhances efficiency by feeding the input in both backward and forward directions. The results regarding the system's performance have been presented.
- The research support tool can further be extended for different research areas, and the classification model can also be trained on different datasets for other application areas.
- This article considers the abstract and title while training the model. In future, the conclusion and the related work part of the articles can also be included for increasing the better exploration.

Data availability statement

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

Author contributions

KB: interpretation and acquisition of data. AT: development of model architecture and user interface. MP, SS, and VS: conception of ideas and formulation and development of designing concepts. All authors contributed to the article and approved the submitted version.

Funding

This publication was realized with support of the Operational Program Integrated Infrastructure in frame of the project: Intelligent technologies for protection of health-care personnel in the front line and operation of medical facilities during spreading of disease COVID-19, code ITMS2014+: 313011ATQ5 and co-financed by the Europe Regional Development Found.

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.

References

1. Simon C, Davidsen K, Hansen C, Seymour E, Barnkob MB, Olsen LR, et al. BioReader: a text mining tool for performing classification of biomedical literature. *BMC Bioinformatics.* (2019) 19:165–70. doi: 10.1186/s12859-019-2607-x

2. Aggarwal T, Salatino A, Osborne F, Motta E. R-classify: extracting research papers' relevant concepts from a controlled vocabulary. *Softw Impacts.* (2022) 14:100444. doi: 10.1016/j.simpa.2022.100444

3. Doty C, Gallagher S, Cui W, Chen W, Bhushan S, Oostrom M, et al. Design of a graphical user interface for few-shot machine learning classification of electron microscopy data. *Comput Mater Sci.* (2022) 203:111121. doi: 10.1016/j.commatsci.2021.111121

4. Wang LL, Lo K, Chandrasekhar Y, Reas R, Yang J, Eide D, et al. Cord-19: the COVID-19 open research dataset. *arXiv.* (2020). doi: 10.48550/arXiv.2004.10706

5. COVID-19 Open Research Dataset Challenge (CORD-19). *Kaggle.* Available online at: https://www.kaggle.com/datasets/allen-institute-for-ai/CORD-19-research-challenge (accessed December 15, 2022).

6. CORD-19: COVID-19 Open Research Dataset. *Allen Institute for AI.* Available online at: https://allenai.org/data/cord-19 (accessed December 15, 2022).

7. Yan L, Zhang H, Goncalves J, Xiao Y, Wang M, Guo Y, et al. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. *MedRxiv.* (2020). doi: 10.1101/2020.02.27.20028027

8. Randhawa GS, Soltysiak MPM, El Roz H, de Souza CPE, Hill KA, Kari L, et al. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. *PLoS ONE.* (2020) 15:e0232391. doi: 10.1371/journal.pone.0232391

9. Song Y, Zheng S, Li L, Zhang X, Zhang X, Huang Z, et al. Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. *IEEE/ACM Trans Comput Biol Bioinform.* (2021) 18:2775–80. doi: 10.1109/TCBB.2021.3065361

10. Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J, et al. A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). *Eur Radiol.* (2021) 31:6096. doi: 10.1007/s00330-021-07715-1

11. Yu J. Open access institutional and news media tweet dataset for COVID-19 social science research. *ArXiv* [preprint] *arXiv*:2004.01791 (2020).

12. Medford RJ, Saleh SN, Sumarsono A, Perl TM, Lehmann CU. *An Infodemic Leveraging High-Volume Twitter Data to Understand Public Sentiment for the COVID-19 Outbreak*. Oxford: Oxford University Press. (2020). doi: 10.1101/2020.04.03.20052936

13. Callaghan S. COVID-19 is a data science issue. *Patterns.* (2020) 1:100022. doi: 10.1016/j.patter.2020.100022

14. Thelwall M, Thelwall S. A thematic analysis of highly retweeted early COVID-19 tweets: consensus, information, dissent, and lockdown life. *ASLIB J Inform Manage.* (2020) 72:945–62. doi: 10.1108/AJIM-05-2020-0134

15. Baïna K. Leveraging data preparation, HBase NoSQL storage, and HiveQL querying for COVID-19 big data analytics projects version 1. *arXiv:2004.00253.* (2020). doi: 10.48550/arXiv.2004.00253

16. Tanne JH. COVID-19: trump proposes tax cuts and improved health insurance, but millions are not covered. *BMJ.* (2020) 993:10–1. doi: 10.1136/bmj.m993

17. Kristoufek L. Grandpa, grandpa, tell me the one about bitcoin being a safe haven : new evidence from the COVID-19 pandemic. *Front Phys.* (2020) 8:296. doi: 10.3389/fphy.2020.00296

18. Chatterjee P. Indian pharma threatened by COVID-19 shutdowns in China. *Lancet.* (2019) 395:675. doi: 10.1016/S0140-6736(20)30459-1

19. Stephany F, Neuhäuser L, Stoehr N, Darius P, Teutloff O, Braesemann F, et al. The CoRisk-Index: a data-mining approach to identify industry-specific risk perceptions related to COVID-19. *Human Soc Sci Commun.* (2022) 9:1–5. doi: 10.1057/s41599-022-01039-1

20. Inoue H, Todo Y. The propagation of the economic impact through supply chains: the case of a mega-city lockdown against the spread of COVID-19. *SSRN Electron J.* (2020) 1–11. doi: 10.2139/ssrn.3564898

21. Zhang R, Liu H, Li F, Zhang B, Liu Q, Li X, et al. Transmission and epidemiological characteristics of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infected pneumonia (COVID-19): preliminary evidence obtained in comparison with 2003-SARS. *MedRxiv.* (2020). doi: 10.1101/2020.01.30.20019836

22. Cho HW. Effectiveness for the response to COVID-19: the MERS outbreak containment procedures. *Osong Public Health Res Perspect.* (2020) 11:1–2. doi: 10.24171/j.phrp.2020.11.1.01

23. Bobdey S, Ray S. Going viral-COVID-19 impact assessment: a perspective beyond clinical practice. *J Mar Med Soc.* (2020) 22:9. doi: 10.4103/jmms.jmms_12_20

24. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Travel Med.* (2020) 1:1–4. doi: 10.1093/jtm/taaa021

25. Wu Y, Chen C, Chan Y. The outbreak of COVID-19: an overview. *J Chin Med Assoc.* (2020) 83:217–20. doi: 10.1097/JCMA.0000000000000270

26. Li R, Lu W, Yang X, Feng P, Muqimova O, Chen X, et al. Prediction of the epidemic of COVID-19 based on quarantined surveillance in China. *medrxiv. [Preprint].* (2020). doi: 10.1101/2020.02.27.20027169

27. Odendaal WG. A method to model outbreaks of new infectious diseases with pandemic potential such as COVID-19. *medRxiv.* (2020). doi: 10.1101/2020.03.11.20034512

28. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data based analysis, modelling and forecasting of the COVID-19 outbreak. *PLoS ONE*. (2020) 15:e0230405. doi: 10.1371/journal.pone.0230405

29. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect Dis.* (2020) 20:553–8. doi: 10.1016/S1473-3099(20)30144-4

30. Liu Z, Magal P, Seydi O, Webb G. Understanding unreported cases in the COVID-19 epidemic outbreak in Wuhan, China, and the importance of major public health interventions. *Biology*. (2020) 9:50. doi: 10.3390/biology9030050

31. Shi F, Yu Q, Huang W, Tan C. 2019 novel coronavirus (COVID-19) pneumonia with hemoptysis as the initial symptom: CT and clinical features. *Korean J Radiol.* (2020) 21:537. doi: 10.3348/kjr.2020.0181

32. Sun Q, Xu X, Xie J, Li J, Huang X. Evolution of computed tomography manifestations in five patients who recovered from coronavirus disease 2019 (COVID-19) pneumonia. *Kor J Radiol.* (2020) 21:614. doi: 10.3348/kjr.2020.0157

33. Prompetchara E, Ketloy C, Palaga T. Immune responses in COVID-19 and potential vaccines: lessons learned from SARS and MERS epidemic. *Asian Pac J Allergy Immunol.* (2020) 38:1–9. doi: 10.12932/AP-200220-0772

34. Pua U, Wong D. What is needed to make interventional radiology ready for COVID-19? Lessons from SARS-CoV epidemic. *Korean J Radiol.* (2020) 21:629. doi: 10.3348/kjr.2020.0163

35. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China : a web-based cross-sectional survey. *Psychiatry Res.* (2020) 288:112954. doi: 10.1016/j.psychres.2020.112954

36. Xiao C. A novel approach of consultation on 2019 novel coronavirus (COVID-19)-related psychological and mental problems: structured letter therapy. *Psychiatry Investig.* (2020) 17:175–6. doi: 10.30773/pi.2020.0047

37. Liu S, Yang L, Zhang C, Xiang YT, Liu Z, Hu S, et al. Online mental health services in China during the COVID-19 outbreak. *Lancet Psychiatry.* (2020) 7:e17–8. doi: 10.1016/S2215-0366(20)30077-8

38. Liu CY, Yang YZ, Zhang XM, Xu X, Dou QL, Zhang WW, et al. The prevalence and influencing factors in anxiety in medical workers fighting COVID-19 in China: a cross-sectional survey. *Epidemiol Infect.* (2020) 148. doi: 10.1017/S0950268820001107

39. Li Z, Ge J, Yang M, Feng J, Qiao M, Jiang R, et al. Vicarious traumatization in the general public, members, and non-members of medical teams aiding in COVID-19 control. *Brain Behav Immun.* (2020) 88:916–9. doi: 10.1016/j.bbi.2020.03.007

40. Herst CV, Burkholz S, Sidney J, Sette A, Harris PE, Massey S, et al. An effective CTL peptide vaccine for Ebola Zaire based on survivors' CD8+ targeting of a particular nucleocapsid protein epitope with potential implications for COVID-19 vaccine design. *Vaccine.* (2020) 38:4464–75. doi: 10.1016/j.vaccine.2020.04.034

41. Li Q, Ding X, Xia G, Geng Z, Chen F, Wang L, et al. A simple laboratory parameter facilitates early identification of COVID-19 patients. *MedRxiv*. (2020). doi: 10.1101/2020.02.13.20022830

42. Goh GKM, Dunker AK, Foster JA, Uversky VN. Rigidity of the outer shell predicted by a protein intrinsic disorder model sheds light on the COVID-19 (Wuhan-2019-nCoV) infectivity. *Biomolecules*. (2020) 10:331. doi: 10.3390/biom10020331

43. Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM networks. In: *Proceedings. 2005 IEEE International Joint Conference on Neural Networks*. IEEE (2005).

44. Yadav SP, Zaidi S, Mishra A, Yadav V. Survey on machine learning in speech emotion recognition and vision systems using a recurrent neural network (RNN). *Arch Comput Methods Eng.* (2022) 29:1753–70. doi: 10.1007/s11831-021-09647-x

45. Ackerson JM, Dave R, Seliya N. Applications of recurrent neural network for biometric authentication and anomaly detection. *Information.* (2021) 12:272. doi: 10.3390/info12070272

46. Dixon M, London J. Financial forecasting with α-rnns: a time series modeling approach. *Front Appl Math Stat.* (2021) 6:551138. doi: 10.3389/fams.2020.551138

47. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* (1997) 9:1735–80. doi: 10.1162/neco.1997.9.8.1735

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 03 March 2023
DOI 10.3389/fpubh.2023.1125011

OPEN ACCESS

EDITED BY
Steven Fernandes,
Creighton University, United States

REVIEWED BY
Sidra Abbas,
COMSATS University Islamabad, Sahiwal Campus, Pakistan
Jatinderkumar R. Saini,
Symbiosis Institute of Computer Studies and Research (SICSR), India

*CORRESPONDENCE
J. Andrew
✉ andrew.j@manipal.edu
J. Karthikeyan
✉ karthikeyan.jk@vit.ac.in

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 15 December 2022
ACCEPTED 06 February 2023
PUBLISHED 03 March 2023

CITATION
Andrew J, Eunice RJ and Karthikeyan J (2023) An anonymization-based privacy-preserving data collection protocol for digital health data. *Front. Public Health* 11:1125011. doi: 10.3389/fpubh.2023.1125011

An anonymization-based privacy-preserving data collection protocol for digital health data

J. Andrew[1]*, R. Jennifer Eunice[2] and J. Karthikeyan[3]*

[1]Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India, [2]Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India, [3]School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India

Digital health data collection is vital for healthcare and medical research. But it contains sensitive information about patients, which makes it challenging. To collect health data without privacy breaches, it must be secured between the data owner and the collector. Existing data collection research studies have too stringent assumptions such as using a third-party anonymizer or a private channel amid the data owner and the collector. These studies are more susceptible to privacy attacks due to third-party involvement, which makes them less applicable for privacy-preserving healthcare data collection. This article proposes a novel privacy-preserving data collection protocol that anonymizes healthcare data without using a third-party anonymizer or a private channel for data transmission. A clustering-based *k*-anonymity model was adopted to efficiently prevent identity disclosure attacks, and the communication between the data owner and the collector is restricted to some elected representatives of each equivalent group of data owners. We also identified a privacy attack, known as "leader collusion", in which the elected representatives may collaborate to violate an individual's privacy. We propose solutions for such collisions and sensitive attribute protection. A greedy heuristic method is devised to efficiently handle the data owners who join or depart the anonymization process dynamically. Furthermore, we present the potential privacy attacks on the proposed protocol and theoretical analysis. Extensive experiments are conducted in real-world datasets, and the results suggest that our solution outperforms the state-of-the-art techniques in terms of privacy protection and computational complexity.

KEYWORDS

anonymization, data privacy, healthcare data, *k*-anonymity, privacy-preserving, data collection

1. Introduction

Healthcare industries have seen a significant transition since the advancements in communication technologies (1). E-health services (2) have become popular for their wide range of advantages such as accuracy, timeliness, easy access, and efficiency (3, 4). Electronic health records (EHRs) (5) are the major step toward the transformation of traditional healthcare services into paperless medical practice that can reduce the risk of medical errors (6–8). Digitized patients' health record benefits both patients and healthcare providers in sharing, monitoring, tracking, and analyzing the healthcare of patients (9). As EHRs follow a standard health record format, it is possible to make them available worldwide (10). EHRs reduce administrative overhead, costs, and medical errors through efficient communication of health information (11). Healthcare organizations often collect EHRs for medical and

research purposes (12). EHRs generally contain information concerning individual health records, medical history, medications, physical conditions, etc. (13). Since there is a huge amount of personal information contained in EHRs, it is crucial to consider privacy issues more carefully (14–16).

Collecting personal health records without breaching the privacy of involved individuals is essential for its success (17–20). In the data collection problem, the data collector is usually an untrusted third-party service provider who collects data from a set of individual data owners (21, 22). Assume that a medical researcher requests data from a number of patients who hold the healthcare demographics. The schema of demography consists of user ID, age, sex, weight, and diagnosis that every patient provides to the data collector. The health record schema is a combination of personal identifiers (e.g., user ID), quasi-identifiers (QI) (e.g., age, sex, weight), and a sensitive attribute (e.g., diagnosis). A sample healthcare records collection table is shown in Table 1.

In the aforementioned example, although there are no direct identifiers such as name and social security number (SSN) in the EHR, privacy breaches can still arise. An untrusted data collector can ascertain the identity of the patient through the explicit identifier *userID* and sensitive attribute *diagnosis* of each individual. Although QI cannot be used to directly identify a person, by connecting them to the data in a published database, it may be possible to do so. The QI can act as an identifier in the absence of a direct identifier. Hence, identity disclosure is one of the major privacy issues in EHR. In the data collection problem, identity disclosure (23) can arise both at internal and external levels. Internal identity disclosure (24) generally happens within the organization either through the data owners or the data collectors. External identity disclosure (25) takes place when the data is transmitted between the owner and the collector.

Unsurprisingly, privacy-preserving healthcare data collection has become a recent research focus where a good number of literature exists (26–32). Cryptography or anonymization-based approaches are widely used to prevent the identity disclosure of EHR (33, 34). Symmetric key and asymmetric key cryptography, multiparty computation, and homomorphic encryption are some of the cryptographic approaches used for privacy-preserving data collection (35); although it guarantees privacy to a certain extent, significant challenges such as heavy computation and key propagation make it a difficult choice. The anonymization approach (36), in general, removes the identifiers and generalizes the QIs excluding the sensitive attribute. Traditional anonymization techniques, such as *k*-anonymity (37), *l*-diversity (38), *t*-closeness (39), clustering-based *k*-anonymity (40), (α, k)-anonymity (41), *p*-sensitive *k*-anonymity (42), and others, anonymize the personal records by grouping similar QI attributes to make them indistinguishable from other sets of records in the same table.

Most of the literature for privacy-preserving data collection has not considered distributed data owners, and it is assumed that personal data are already collected in a common place to be anonymized (43). Hence, in centralized solutions for privacy-preserving data collection, it has become essential to employ a third-party anonymizer (44). However, it is highly undesirable for a patient to share his/her original EHR with a third party. There is also a huge risk of a privacy breach when a data owner (patient) directly shares their personal information with the data collector. The existing privacy models drudged to control the disclosure by deploying an anonymization layer or private unidentified channel between the data collector and the data owner. Nonetheless, such assumptions are not practical as the layer or channel is not persistent. Cryptographic approaches also encrypt the healthcare records to prevent identity disclosure at the data collector's end; furthermore, the data are anonymized, resulting in poor data utility.

In this research, we propose a data collection protocol for EHRs that is effective and protects privacy in order to address the aforementioned problems. In the proposed protocol, multiple data owners anonymize their health records in a distributed and collaborative fashion before submitting the data to the data collector. This protocol's main goal is to forbid explicit exchanges between data owners and data collectors. The data owners submit their anonymized QIs through a set of representatives elected for their equivalent group. Representatives are data owners of the equivalent group with common quasi attributes. Every equivalent group should satisfy the clustering-based *k*-anonymity property (i.e., at least *k*-1 records share the same quasi attributes); therefore, the anonymized records with common QIs are submitted to the data collector through group representatives. This approach of the proposed protocol is efficient in tackling internal and external identity disclosure. Table 1 shows the original EHR of *n* patients, Table 2 shows the anonymized version of the original records by the proposed protocol. As shown in Table 2, there are two equivalent groups that share common QIs of size $k = 3$. Such equivalent groups, along with sensitive values (e.g., diagnosis), are collected by the data collector, which reduces the risk of identity disclosure. Furthermore, dynamic data owners

TABLE 1 Electronic health records.

User ID	Age	Sex	Weight	Diagnosis
1,2,3	30–40	F	55	Gastritis
		F	50	Flu
		F	60	Dyspepsia
4,5,6	55–65	M	65	Pneumonia
		M	75	Flu
		M	68	Cancer

TABLE 2 3-anonymized health records.

User ID	Age	Sex	Weight	Diagnosis
1	35	F	55	Gastritis
2	40	F	50	Flu
3	45	F	60	Dyspepsia
4	55	M	65	Pneumonia
5	60	M	75	Flu
6	65	M	68	Cancer

who join or leave an equivalent group are handled by a greedy heuristic method.

The major contributions of the proposed protocol are as follows:

(1) **Privacy-preserving healthcare data collection protocol:** A novel *k*-anonymity-based data collection protocol specifically for healthcare data collection is proposed.
(2) **Leader election:** A leader election algorithm is proposed to elect representatives of equivalent groups of anonymized records that share similar generalized quasi attributes.
(3) **Greedy heuristic method:** Data owners who dynamically join or leave the group is efficiently managed without affecting the data utility and privacy.
(4) **Leader collision mitigation and sensitive attribute protection:** We propose solutions for privacy breach through leader collision and methods to enhance the protection of sensitive attributes.

The remainder of this article is structured as follows. The recent state-of-the-art literature is discussed in the Section 2. In the Section 3, an adversarial model of the proposed protocol is presented, along with a data model and other definitions. In the Section 4, the proposed protocol is formally defined, along with the proposed algorithms. In the Section 5, data utility and possible privacy attacks on the proposed protocol are discussed. In the Section 6, experiments conducted are presented. Finally, the Section 7 concludes the article.

2. Literature survey

In the last decade, a huge number of research studies were conducted in privacy-preserving data publication and data collection. This section presents a detailed study of various state-of-the-art literature available in the field of preserving the privacy of personal data. In privacy-preserving data collection and publication, disclosure or reidentification of data owners has been a significant issue. The state-of-the-art literature consists of cryptographic and anonymization-based approaches for privacy preservation. The collection of personal data is accomplished through devices and sensors. The device periodically collects and transmits the data to the data collector upon request. The data transmission is generally conducted in a closed or open network. Hence, it is essential to ensure the secure transmission of data. Hussien et al. (45) used a symmetric key cryptographic technique to propose a secure and energy-efficient method to collect data in wireless sensor networks.

Most privacy-preserving schemes require a secure transmission channel or a third-party authentication system. However, they are impractical due to various challenges. In (46), Beg et al. have proposed a reversible data transform (RDT) algorithm for privacy-preserving data collection in the mobile recommendation system (MRS). The proposed RDT algorithm is used to protect sensitive attributes. To avoid the third-party role in the data collection process, the data transfer is done through elected representatives. However, the leader election process is straightforward, and leader collision is possible that can breach privacy. However, the same authors in (47) addressed the RDT prior data sharing and its parameter protection challenges by proposing a chaotic RDT for PPDP MRS. The authors also claim that the proposed approach can replace homomorphic encryption techniques and preserve the privacy of the MRS. The leader collusion problem is addressed by Sajjad et al. (48) through a random leader election mechanism that elects the leaders randomly and maintains a leader table for maintaining the records. However, this scheme is inefficient, which simply uses a random function to select the leaders, and leader collusion is still possible when the number of available groups is minimal. Data anonymization is vital in protecting big data and IoT data. Ni et al. (49) evaluated the performance of data anonymization schemes in an IoT environment for big data. The authors addressed the reidentification risks and evaluated the schemes based on privacy preserving-level and data utility metrics. Traditional anonymization schemes like *k*-anonymity, *l*-diversity, obfuscation, permutation, and differential privacy techniques (50) are evaluated through information loss, data utility, and conditional entropy. A similar study was presented by Sun et al. (51) for trajectory data publishing. Canbay et al. (52) proposed a Mondrian-based utility aware anonymization approach called u-Mondrian. This approach is aimed to address the upper-bound problem in the Mondrian anonymization approach that leads to poor data utility.

Healthcare data contain sensitive information that must be protected concurrently; it is very vital for healthcare research. Hence, it is essential for protecting the privacy of healthcare data with appropriate data utility. In (53), we proposed a clustering-based anonymization approach for privacy-preserving data collection in a healthcare IoT environment. The proposed approach utilizes a client–server model to anonymize the healthcare data before it reaches the data collector. The model is evaluated with information loss and other data utility metrics. A similar approach was proposed by Abbasi and Mohammadi (54) to protect the privacy of healthcare data in cloud-based systems. They proposed an optimal *k*-anonymity technique called the *k*-means++ method and used the normal distribution function to improve the anonymization data utility. We performed another study called an attribute-focused approach (55) to protect the privacy of healthcare data during data publishing. In this study, the healthcare attributes are categorized as numerical and sensitive attributes. A fixed-length interval approach is used to protect the numerical attributes and an improved *l*-diversity approach is used to protect the sensitive attributes. Avraam et al. (56) proposed a deterministic approach for protecting the privacy of sensitive attributes. This approach identifies the categorical and continuous attributes from the dataset and applies different mechanisms to prevent a privacy breach. The stratification technique is used for categorical and continuous attributes that are redistributed based on *k*-nearest-neighbor algorithms. The proposed approach is claimed to be efficient in preventing the data from reidentification. Kanwal et al. proposed multiple anonymization-based approaches to preserve the privacy of health records. In (57), they proposed a privacy scheme called horizontal sliced permuted permutation to protect multiple records of data owners. They considered the protection of multiple sensitive attributes by proposing 1: M MSA-(p, l)-diversity approach (58). Furthermore, the authors proposed

an anonymization technique with an access control mechanism for hybrid healthcare cloud services. In all the studies, they evaluated data privacy for various privacy attacks such as identity disclosure attacks, membership disclosure, and sensitive attribute disclosures. Jayapradha and Prakash (59) presented a privacy-preserving model called f-slip that uses a frequency-slicing approach to protect sensitive attributes. Sensitive attributes are correlated to maintain the linking relationship during the anonymization process. Khan et al. (60) used phonetic encoding and generalization approaches for record linkage problems. The authors used phonetic encoding for anonymizing textual data, and for categorical and numerical attributes, the k-anonymization-based approach is utilized. Raju and Naresh (61) proposed a distributed algorithm to merge the datasets from different sources to maintain their privacy. To preserve the privacy of the sensitive attributes, they proposed a bucketization-based approach called $(l,m,d)^*$- anonymity. The proposed approach anonymizes the data and transforms the data into a sensitive attribute and quasi-attribute table.

Based on the in-depth literature study of the recently published literature, most of the privacy-preserving models are still using the k-anonymization-based approach. However, they either use a private secure channel or a third-party anonymizer for privacy-preserving data collection. This may lead to a possible privacy breach. Hence, a k-anonymity-based privacy-preserving protocol for data collection without a third-party anonymizer is on demand.

3. Preliminaries

Various terminologies used in this study are introduced in this section. The components of the proposed protocol such as the data model, adversary model, and system architecture are defined.

3.1. Data model

We assume that EHRs are generated periodically on the users' devices. Out of the different attributes of personal healthcare data, only the major attributes such as personal identifiers, QIs, and sensitive attributes are considered in this article. Personal identifiers are explicit attributes that unambiguously distinguish a particular individual (e.g., social security number, name, IP address, and phone number). Identifiers are generally removed in the process of data collection and publication to avoid identity and attribute disclosure.

QIs are common attributes that can be shared by more than one data owners (e.g., age, sex, and zip code). Although they cannot directly identify an individual, the combination of QIs with publicly available datasets may breach privacy. In general, generalization and suppression approaches are used to protect QIs. Sensitive attributes (S) are details about a person that should not be shared (e.g., diagnosis). Identification of an individual's sensitive information, along with the identity, is a serious privacy breach. Hence, sensitive information is needed and protected with top priority.

3.1.1. Definition 1: (Personal health data)

In personal health records table T, let H be a unique record in the table and H^{qi} be one of the QIs, and H^{si} be the single sensitive attribute (S) of the particular record. The health data schema is then defined as follows:

$$\left(H_1^{qi}, H_2^{qi}, H_3^{qi}, \ldots, H_m^{qi}, H^{si}\right)$$

where m is the number of QIs for the record. In this article, a single sensitive attribute problem is considered.

3.1.2. Definition 2: (Anonymization)

The term anonymization means protecting the identity. Hence, it involves a process of transforming the original health records to an equivalent less significant record. The original health record table T is mapped with an anonymization function f to generate an anonymized table T^*. Every record of t in T is mapped to a record in T^*. The anonymized QI attribute QI^* for every t in T^* is then defined as $t_i\,[QI] \prec t_i^*[QI]$.

3.1.3. Definition 3: (k-anonymity)

A personal health dataset T satisfies k-anonymity when a record t of T^* is imperceptible from at least k-1 other records. It is given by $k \leq N(t\,(QI))$ for every record $t \in T$, $N(t\,(QI))$ – number of records shares the same QI.

3.1.4. Definition 4: (Clustering-based k-anonymity)

A personal health dataset T satisfies the clustering-based k-anonymity (25) property if a set of clusters formed from n records where each cluster consists of k records where $k \leq n$.

3.1.5. Definition 5: (Equivalence class)

To create an equivalent class, at least k data owners' anonymized records with related quasi characteristics must be used. Let G^E represent the collection of data owners k who are grouped by the same anonymized quasi attributes QI^*. G^E is an equivalent group if and only if $G^E = \{d|d\,[QI] = qi\}$ and $k \leq G^E$, where d represents an arbitrary data owner with quasi attribute $d\,[QI\,]$.

3.2. Adversary model

In privacy-preserving healthcare data collection context, there could be a single data collector and multiple data owners.

Personal health data are generated by data owners (Definition 1). We assume that there are n data owners in the network and can communicate with other data owners and the collector. The client devices (e.g., medical sensors) at the data owner's end perform communication. The data owners collaborate with other clients not only to protect their health data but also patients in the network.

The data collector collects anonymized health records from the patients. In our protocol, the data collector is assumed to be a single semi-honest collector in the network. A semi-honest entity in a

network generally follows the protocols but sometimes breaches the protocol to acquire more information. An attempt may be made to learn more about a person by a semi-honest data collector. This leads to identity disclosure.

A group of data owners who share the same quasi attributes forms an equivalent group (Definition 5) satisfying the *k*-anonymity and clustering-based *k*-anonymity model (i.e., at least *k* data owners in an equivalent group). Table 2 shows the example of an anonymity model that contains two groups with the value of $k = 3$. The records in the equivalent group share similar quasi attributes. The data owners interact with the data collector through the equivalent groups. Thus, it protects the data from external identity disclosure. Since the data owners share common quasi attributes in an equivalent group, internal identity disclosure is also protected.

An adversarial model is necessary to identify possible privacy attacks in the system. In a privacy-preserving data collection model, an adversary could be a data collector and data owner. The data collector is considered to be a malicious component in the network. Therefore, giving the data collector access to the original records is not appropriate. The clustering-based *k*-anonymity model ensures anonymized data is submitted to the data collector. The data owner can also be an adversary. An adversarial data owner generates fake quasi attributes and gets added to a specific equivalent group. During the random election of group representatives, if the adversarial data owners are elected as the first and second leaders of the group, then the sensitive attributes are disclosed. Such an attack is called a leader collision attack (LCA).

3.3. Overview of the protocol

Initialization, leader election, and data collecting phases make up the proposed data collection process. In the initialization step, the data owners (patients) create QI attributes and provide them to the data collector (without sensitive attributes). The data collector applies the provided clustering-based k-anonymity model to anonymize the health records. This results in the original QI being equivalent to at least *k*-1 generalized quasi characteristics (GQI). The appropriate data owners are then given the GQI and the list of data owners. The data owners then create comparable groupings that comply with the privacy policy.

In the leader election phase, members of an equivalent group are assigned with unique numbers; then based on a random number generation function, two leaders are elected for each equivalent class. The first leader obtains each member's hidden sensitive attributes from the phase of data collecting that uses sensitive values that are not real. The GQI and list of sensitive data are then given to the data collector. Without actually possessing sensitive information, the second leader gathers counterfeit sensitive information. In order to obtain the anonymized dataset, the data collector then executes intersection operations on the first and second leader datasets. The proposed privacy-preserving data collection protocol's architecture is depicted in Figure 1.

The proposed approach additionally takes into consideration of dynamic data owners who join or depart the equivalent class during the anonymization process. Dynamic join or leave follows the privacy requirement and ensures the required number of members for each group.

4. Privacy-preserving healthcare data collection protocol

Initialization, leader election, and data collection are the three phases of the protocol. The anonymization network is organized during the initiation phase, and the QI properties of the data owners are generalized. Representatives from related groups were chosen to serve as the leader during the election process. The data collector is finally given access to the anonymized records

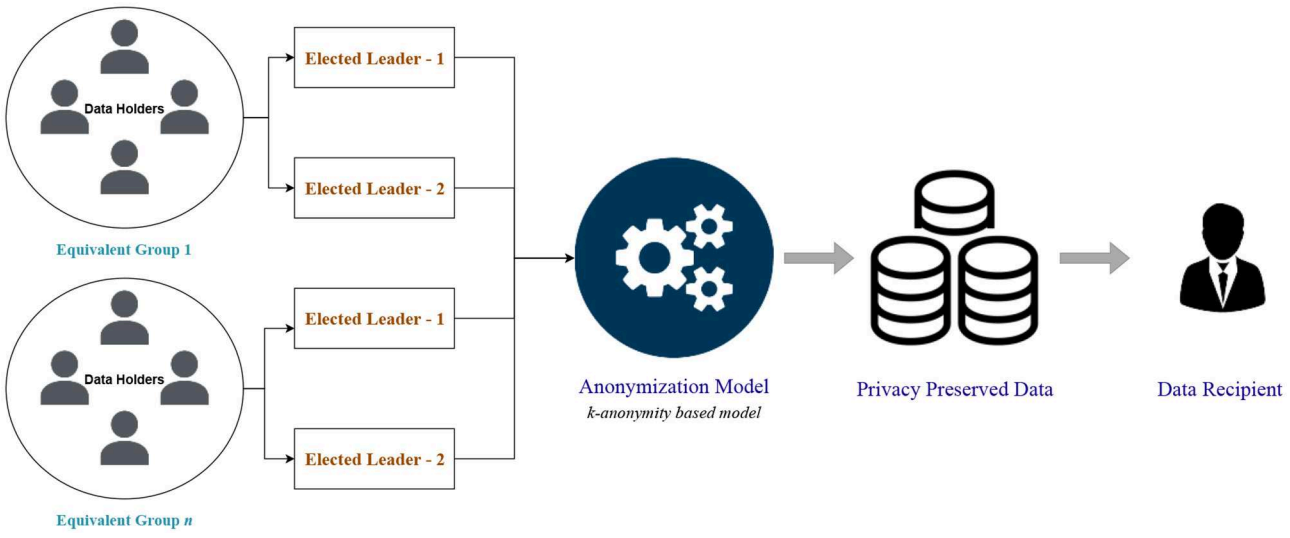


FIGURE 1
Privacy-preserving data collection protocol architecture.

with quasi characteristics and sensitive attributes during the data collecting phase. We also outline techniques for managing data owners who join or leave the network on a dynamic basis.

4.1. Initialization phase

The anonymization network is set up by the initialization phase. Data owners and data collectors are required to initialize their attributes for the network. There are two algorithms proposed for data owner initialization and data collector initialization. Data owners initially transmit their QI attributes to the data collector over the specified network. It should be highlighted that the data owners do not send their sensitive qualities. Over time, the data collector gets QI attributes from n data owners. Then the data collector anonymizes the QI attributes based on any given privacy model (37–40) to generate generalized quasi attributes (GQI). For example, Table 1 shows the original health records of n ($n = 6$) data owners that are sent to the data collector without the sensitive attribute (e.g., diagnosis). Table 2 shows the anonymized version of Table 1 with the value of $k = 3$.

The generated GQIs are distributed to the relevant data owners together with a list of data owners who have common GQIs. The list is then used by the data owners to connect with other data owners who have the same GQI. Every data owner then verifies their GQI with other data owners to form an equivalent group. Equivalent groups should satisfy the privacy policy of at least k data owner records present in every group. For example, Table 2 shows two equivalent groups that share the same GQI. The detailed steps of initialization for the data owner and data collector are shown in Algorithms 1, 2. Table 3 describes the symbols used in the algorithms.

Algorithm 1 runs at the data collector end to receive the quasi attributes from the data owners and to generate GQI based on any given anonymization techniques. It then disseminates the GQIs to the data owners. Algorithm 2 runs at the data owner's end to send the QIs to the data collector and to form equivalent groups based on the received GQI.

```
Input: QI - Data owners quasi attributes, k - privacy parameter
Output: GQI - Data collector's generalized quasi attributes
for each QI received from data owner D_i do
        insert QI into QIT
end for
G_ID= Group ID
while QIT ≠ NULL do
        anonymize QI to GQI w.r.t k
        insert GQI into GQIT
        G_ID = G_ID + 1
end while
return GQIT
return D list of data owners
```

Algorithm 1. Data collector—initialization.

4.2. Leader election phase

On the data owners' side, equivalent classes are formed as per the privacy requirement k. In the leader election phase, two leaders are elected to represent the group and interact with the data collector. Algorithm 3 shows the detailed steps for leader election. First, the equivalent class members are counted. Then the *random()* function is used to generate two random numbers between 1 and the maximum number of members in the group. First, the randomly generated *userID* is considered as the first and second leader. Then we identified the energy and delay-less efficient leaders by utilizing the firefly-based algorithm proposed by Sarkar and Senthil Murugan (62). Firefly-based algorithm calculates the Euclidean distance between the elected leader and the nodes in the network then based on the distance metrics a firefly with cyclic randomization is performed to select the best leaders from among the groups. After every leader election, the leader table is updated. This algorithm ensures a single data owner is selected as the first and second leader. The elected leaders then transfer data to the data collector in the data collection phase.

4.3. Data collection phase

The major task of the data collection phase is to collect the anonymized personal health records from the data owners. During the data collection initialization stage, QI attributes of data owners are generalized by the data collector then equivalent groups are formed on the data owners' side. To avoid explicit interaction of data owners with the data collector, group leaders are elected in the leader election phase. The leaders of each group are responsible for communicating QIs and sensitive identifiers. There are two leaders elected, the first leader (L^1) is responsible to send the generalized QIs and multivalued sensitive attributes (MSA). The members equivalent group sends

```
Input: GQIT from data collector, D list of data owners
Output: G^E - set of equivalent groups
for all d ∈ D_i do
        generate QI
        send QI to the data collector
end for
receive GQIT, D from data collector
for all gqi ∈ GQIT do
        if gqi_i == d(GQI) then
                insert GQI into G_i^E
                continue
        else
                break
        end if
end for
Get consent to add d in equivalent group G_i^E
return G^E
```

Algorithm 2. Data owner—initialization.

TABLE 3 Symbols.

Symbols	Description	Symbols	Description
QI	Quasi identifier	R_{GE}	Number of records in G^E
QIT	Quasi identifier table	G	Number of groups in anonymized dataset
GQI	Generalized quasi identifier	L^1	First leader
GQIT	Generalized quasi identifier table	L^2	Second leader
G_{ID}	Group ID	LT	Leader information table
D	Data owner	U^{ID}	Group member user ID
G^E	Equivalent group	CS_j	Counterfeit sensitive information of L^1
ST_R	Sensitive information of L^2	ST_j	Number sensitive information in L^1
AT	Anonymized table	S_j	Sensitive attribute in final table AT

Input: G^E- set of equivalent groups
Output: LT - Leader Table with their respective group id G_{ID}
1: R_{GE} - Number of records in an equivalent group G^E
2: R_1, R_2 = Values ranging from 1 to R_{GE} for every $G_i^E \in G^E$
3: L^1 = ***rand*** (R_1, R_2)
4: L^2 = ***rand*** (R_1, R_2)
5: Calculate Euclidean distance between leader and the group members
6: Identify the leaders by firefly cyclic randomization (62)
7: **if** ($L^1 \neq L^2$) **then**
8: insert L^1,L^2 into LT
9: insert respective G_{ID} into LT
10: **end if**
11: **return** LT

Algorithm 3. Leader election.

TABLE 4 Anonymized data collection (first leader).

User ID	Age	Sex	Weight	Diagnosis
1,2,3	30–40	F	50–60	Gastritis, heart disease, pneumonia
		F		Flu, cancer, osteoarthritis
		F		Dyspepsia, gastritis, flu
4,5,6	55–65	M	65–75	Pneumonia, cancer, arrhythmia
		M		Flu, bronchitis, pneumonia
		M		Cancer, heart disease, gastritis

TABLE 5 Anonymized data collection (second leader).

User ID	Diagnosis
1,2,3	Heart disease, pneumonia
	Cancer, osteoarthritis
	Gastritis, flu
4,5,6	Cancer, arrhythmia
	Bronchitis, pneumonia
	Heart disease, gastritis

their anonymized records along with the multivalued sensitive attribute to the first leader. The MSA is a combination of an original sensitive attribute and n-1 counterfeit-sensitive attributes (where n is the size of the equivalent group's records). Hence, the first leader cannot discern the sensitive attributes of others in the group. Table 4 shows the example of the first leader anonymized dataset. The members of an equivalent class send their counterfeit sensitive attributes (CSA) (without the original sensitive attribute) to the second leader (L^2). Table 5 shows the example of the second leader dataset that only contains CSA along with the userID.

The data collector receives the datasets for the first and second leaders from each equivalent group during the data collecting phase. Elimination of counterfeit information from the first leader dataset is another important process for data collectors. It is hard for the data collector to identify the first and second leader datasets of each equivalent class as it performs subtraction and aggregation to eliminate the CSA from the dataset. The detailed steps of the data collection phase are given in Algorithm 4.

4.4. Dynamic data collection phase

The data collection protocol is designed in a way that it can consider data owners who join or depart the network dynamically. Dynamic data owners have to be efficiently managed to avoid any privacy breach to the network. The challenges with dynamic data owners are when a dynamic data owner joins the network, he/she should be placed in the appropriate equivalent group with minimal information loss and when a dynamic data owner leaves the network it should not affect the required privacy policy and without any privacy breach. During dynamic join or leave, the entire equivalent group needs to be reorganized, which incurs huge computational costs. Hence, the greedy heuristic method is proposed to efficiently handle dynamic data owners.

```
Input: L^1- Dataset, L^2 - Dataset, GQIT
Output: AT - Anonymized Table
1: g = number of groups
2: U^ID = user id of group g_i
3: CS_j = counter sensitive attribute of L_i^1 at
   column j
4: ST_R =sensitive information of L_i^2 of a
   particular U^ID
5: ST_j = number of QIs in L^2
6: S_j = sensitive attribute after removing
   counterfeit information
7: for i = 1 to g do
8:          for j = 1 to ST_j do
9:                   if CS_j = = ST_j then
10:                            S_j = CS_j − ST_j
11:                            insert S_j to AT
12:                   end if
13:          end for
14: end for
15: return AT
```

Algorithm 4. Data collection.

4.4.1. Dynamic join

When dynamic data owners try to join the network, they transmit the data collector their QI attributes. The data collector considers the QI attribute as a dynamic join request and finds appropriate GQI from the existing GQIT to minimize the information loss. The new data owner is then added to the particular equivalent group who the share same GQI. The representatives (the first leader and the second leader) and group members are then notified about the new member in the group along with the modified GQI. Thereafter, the new data owner is considered for anonymization and GQI communication in the network.

4.4.2. Dynamic leave

Data owners may leave the network due to unforeseen situations like power failure, system failure, and network failure. In such situations, a data owner leaves the network dynamically. It should be handled efficiently without breaching privacy. Each equivalent class consists of k or more data owners based on the privacy requirement. When a data owner departs the network, the corresponding equivalent class will be updated as per the number of remaining data owners to maintain the k-value for privacy. After the dynamic leave if the number of data owners is less than k, then the members of the equivalent group should be released to form a new group; otherwise, privacy would be breached. If a dynamic leave does not affect the minimum k-value of the group, then no specific handling is required as it is still within the privacy policy. But if the data owner who left is the first or second leader, then the leader election process should be carried out to elect new leaders.

Dismantling an existing equivalent group to form new groups during a dynamic leave is a heavy computational process. In the proposed protocol, such situations are handled by enforcing a threshold time limit. Dynamic leave of a data owner may be temporary or permanent. In temporary leave, the data owner rejoins the network within a particular time period, whereas, in permanent leave, the data owner will not join the network for further process. Hence, the threshold time is enforced to wait for any temporary leave data owner to rejoin. This reduces the computation cost as there is no further process required. If a data owner cannot rejoin within the time limit, then the members of the group will be released and a new group is formed based on the available data owners by satisfying the k-value and new leaders are elected. Thus, the dynamic leave of a data owner is efficiently handled in the protocol without a privacy breach.

5. Experiments

We evaluate our protocol in terms of computational complexity with respect to CSA elimination. In our privacy-preserving data collection protocol, we evaluate the computational complexity of the data collection phase only. The initialization and leader election phase has a complexity similar to traditional centralized anonymization techniques. Hence, the performance of the proposed protocol can be evaluated through CSA elimination of the data collection phase.

5.1. Experimental settings

The algorithms are implemented in Python programming and executed on Quad-Core Intel i7 at 2.2 GHz with 16 GB of RAM running Mac OS 10.15.3. We experimented our protocol on real-world public available datasets: the adult (63) and the informs (64).

5.2. Experimental analysis

The efficiency of the protocol in real-world datasets is analyzed in this section. First, the analysis is done with the adult dataset. There are 32,561 records with 14 attributes available in the adult dataset. The attributes "salary" and "occupation" are considered sensitive attributes. The sensitive attributes are merged as a single attribute "occupation-salary"; thus we increased the number of sensitive attributes to 30. It should be noted that our protocol does not consider multiple sensitive attributes. The computational complexity of the protocol is evaluated with the number of sensitive attributes (s) vs. time taken (in ms) by the protocol to eliminate the CSA. Figure 2 shows the computational complexity of the adult dataset with s as the x-axis and computational complexity (ms) as the y-axis. It is observed from the graph that the computational complexity increases with the number of sensitive attributes the protocol has to deal with is increased. Since the model deals with fewer sensitive attributes, the overhead seems to be stable with a slight increase in the s value.

The informs demographic dataset consists of 102,581 records and has 18 attributes. We consider "income" as the sensitive attribute and the domain size is 23,784. Figure 3 illustrates the computational complexity of the informs dataset. It is observed that the counterfeit elimination with larger domains incurs more overhead to the protocol. In the graph, the computational

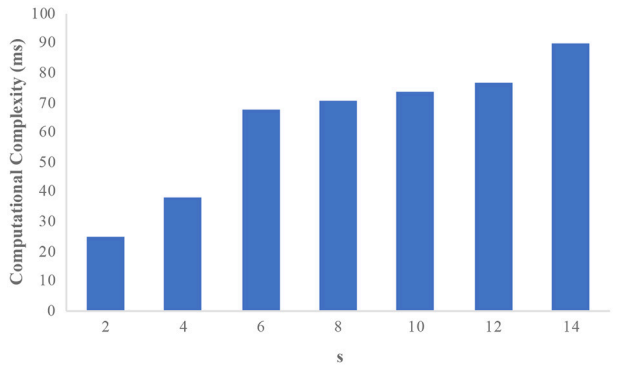


FIGURE 2
Computational complexity on the adult dataset.

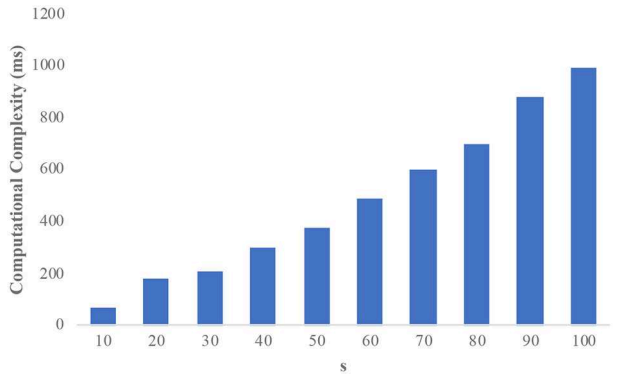


FIGURE 3
Computational complexity on the informs dataset.

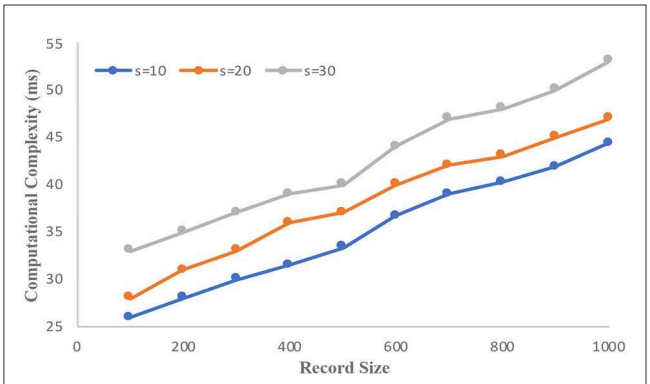


FIGURE 4
Computational complexity vs. record size.

complexity constantly increases with the size of the sensitive attributes (s) in the network. Figure 4 illustrates how the informs dataset's computing complexity varies depending on the number of sensitive features. It is understood from the graphs that computational overhead increases with the size of the dataset and the domain size. The rise is caused by the volume of fake sensitive qualities that must be addressed.

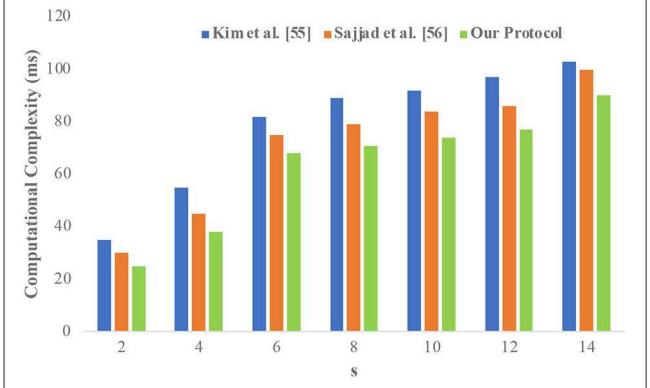


FIGURE 5
Performance evaluation of the proposed system.

5.3. Performance evaluation

The performance evaluation of the proposed study is compared with similar studies conducted by Kim and Chung (65) and Sajjad et al. (48). Figure 5 compares the performance of the proposed protocol with the state-of-the-art literature (experiments on the adult dataset). It is observed that the proposed protocol has considerably minimized the computational complexity. It is due to the slight changes in the CSA elimination where the distinct rows are compared instead of the whole dataset.

6. Discussion

This section outlines potential attacks on the suggested protocol as well as the measures the protocol uses to defend against them. We also discuss other important issues in the protocol such as leader collision mitigation and determination of CSA count. Furthermore, we discuss the complexity analysis and data utility of the protocol.

6.1. Internal and external identity disclosure attacks

When a legitimate member in the anonymization network tries to determine a person's identity, internal identity disclosure occurs. In our protocol, we consider a data collector as an adversary who seeks to gain more information about an individual. The adversary may target an individual to discern the sensitive attribute and to try to distinguish through the combination of quasi attributes. We employ a clustering-based *k*-anonymity (40) privacy model to anonymize personal health records that prevent identity disclosure. Clustering-based *k*-anonymity model generalizes the quasi attributes and forms clusters that contain at least *k* records each. As a result, the probability of identity disclosure is limited to *1/k* or less. Although the adversarial data collector has access to the generalized quasi attributes and sensitive attributes, the clustering-based *k*-anonymity policy makes internal identity disclosure nearly impossible.

External identity disclosure can happen when the data is transmitted using the given network. A practical data transmission environment is considered in the protocol, so it is necessary to add headers to the microdata. Our proposed protocol avoids direct connection between the data owner and collector in order to protect the external identity exposure, and instead relies on representatives (such as group leaders) to deliver the data to the data collector. Since all data owners in an equivalent group share the same generalized quasi attributes and the sensitive attributes are covered by a list of CSA, the group leaders are unable to determine who the data owners are. In addition to the original sensitive property, every record in the first leader dataset also contains at least *k-1* CSA. This ensures that the representative's identity disclosure does not exceed *1/k*.

6.2. Leader collision mitigation

Leader collision is a privacy attack where the elected representatives are adversarial data owners and attempt to discern sensitive information. In the leader election phase, each equivalent group elects two leaders. The first leader gathers the sensitive attribute along with the CSA. The second leader collects the CSA without real sensitive attributes and QIs. In an equivalent group if a single data owner is elected as the first and second leader, then the sensitive attributes can be discerned through the elimination of second leader sensitive attributes from the first leader dataset. In the proposed protocol, we verify the elected leaders' userIDs to make sure they are of a single data owner. Algorithm 3 shows the steps to elect different data owners as representatives.

Another type of LCA is identified by Sajjad et al. (48). Adversarial data owners may join the network by generating fake quasi attributes. They intend to be grouped under a particular equivalent group and try their chance to be elected as the group leaders. If both first and second leaders are elected from the adversarial data owners, they can collaborate and discern the sensitive attribute. This type of attack is called LCA. In our proposed protocol, we utilized firefly with a cyclic randomization algorithm (62) to elect the leaders. First, the number of data owners and their userIDs (index values) are collected, and based on the minimum and maximum index values, the random function generates two different userIDs. The generated userID is then considered the first and second leader for that specific data collection phase. The leader information is then stored in the leader information table for further verification.

6.3. Determination of counterfeit sensitive attribute count

Counterfeit sensitive attributes play an important role in protecting the sensitive attributes of the equivalent group. Similar privacy preserving data collection studies (48, 65) proposed the method of adding CSA to the anonymization network. However, the number of CSA to be added to the original sensitive attribute is not specified. It is important to determine the number of CSA required to protect the sensitive attribute in the anonymization network. In our protocol, we determine the count of CSA based on the privacy parameter k. It is proved from the k-anonymity-based privacy model that the identity can be disclosed only at the probability of $1/k$. So, we consider the privacy parameter k as the count of CSA along with the actual sensitive attribute. Hence, the sensitive attribute of each data owner is protected and the probability to disclose the sensitive attribute is not $>1/k$. In our protocol, the privacy parameter value k is shared with every data owner as the CSA count. Each data owner generates k-1 counterfeit attributes to be added with the real sensitive attributes. To improve the quality of CSA, semantic diversity (66) among the sensitive attributes can be pitched in.

6.4. Complexity analysis

The complexity of the proposed protocol can be analyzed for the three phases of the data collection protocol: initialization, leader election, and data collection phase. The data owner's initialization phase comprises *QI* generation, submission, and *GQI* validation tasks. Let Ct_{gen}, Ct_{sub}, *and* Ct_{val} be the complexity of the three tasks. *QI* generation is the basic operation of the data owner, the cost Ct_{gen} is in $O(1)$ where the *QI* is generated at a constant time. The complexity of Ct_{sub} is in $O(1)$ where each data owner can submit the *QI* at a given time. Ct_{val} is in $O(k)$ where k is the number of records in each equivalent group. In the data collector's initialization phase, the major tasks are *QI* generalization and *QI* distribution. Let Ct_{anon} and Ct_{dist} be the cost of the two tasks. Ct_{anon} is the cost of the anonymization technique that is adopted in the protocol. In traditional k-anonymity models, the cost of anonymization is NP-hard with complexity $O(n^2)$. In our proposed protocol, we adopted a clustering-based k-anonymity model so the cost Ct_{anon} is in $O\left(\frac{n^2}{k}\right)$. The distribution cost Ct_{dist} is in $O(n)$ where n is the number of data owners in the network. The total cost of the data collector at the initialization stage is in $O\left(\frac{n^2}{k}\right) + O(n)$. Leader election is another trivial task, the cost of Ct_{elec} is in $O(u)$, where u denotes the users in the network. In an equivalent group, Ct_{elec} is in $O(k)$, where k is the records in the equivalent class.

In the data collection phase, the elimination of CSA from the first leader dataset using the dataset of the second leader is a major task. The CSA values obtained from the second leader dataset are required to be compared with anonymized records of the first leader dataset. Let s be the sensitive attributes in an equivalent group then the number of sensitive attributes in a group is $k \times s$. The list of CSA in the dataset is $k \times s - 1$. If g is the number of equivalent classes, then the cost of CSA elimination is $O(g \cdot k^2 \cdot s^2)$. In our protocol, counterfeit elimination is carried out by comparing the CSA only with distinct sensitive attributes. Hence, the cost of CSA elimination is restricted to $O(kds)$ where d denotes the sensitive attribute domain size.

6.5. Data utility

In the process of anonymization, the original dataset tends to suffer from poor data utility. The data utility is generally measured through various information loss metrics. Likewise, a

dataset with minimum or no information loss may leak privacy. Hence, it is important to maintain the trade-off between privacy and data utility. In our protocol, the anonymization process is carried out only during the initialization phase. The QI attributes are anonymized by the data collector through a utilized clustering-based *k*-anonymity model (53) that forms clusters as the equivalent groups with *k* or more records in each group. Thus, data utility is inherited from the adopted privacy model. Furthermore, our protocol can adopt any *k*-anonymity based privacy model. The information loss and data utility are based on the chosen privacy model. Hence, in this study, we did not present the results of the information loss as our protocol is independent of the privacy model.

6.6. Healthcare data security analysis

Beyond privacy protection, it is also essential to secure healthcare data from unauthorized access and disclosure (67). The potential security threats to a healthcare system are covered in this section.

Due to the requirements of the legal, ethical, and medical domains, healthcare data must be protected from unauthorized access and disclosure (68). To protect health information, three data security techniques are widely in use; they are cryptographic security, blockchain based security, and network security. Cryptography is the most commonly used technique to protect data from unauthorized access, tampering, and an interception. Data encryption plays a major role in protecting data. Qiu et al. (69) proposed a selective encryption algorithm to secure healthcare data sharing with fragmentation and dispersion techniques. This algorithm ensures data safety even when the cloud servers and keys are compromised. Blockchain based security techniques are popular because of their unhackable distributed ledger and smart contracts. Zhuang et al. (70) proposed a blockchain model to protect patient records from unauthorized access and disclosure. The blockchain properties such as immutability, smart contract, and distributed ledgers ensure data consistency, quick access, and patient authorization. The network is another essential part of the healthcare domain that needs proper security to avoid eavesdropping, intrusion, and tampering attacks. Most healthcare systems employ IoT, wireless networks, and body area networks. So appropriate network security is required to protect the data transferred between the data owner and the collector (71–73).

7. Conclusion and future work

In this article, we presented a privacy-preserving healthcare data collection protocol. The state-of-the-art privacy-preserving data collection models, coerce strict assumptions such as secure private channels or third-party anonymization between the data owners and the collector. The proposed protocol eliminates such assumptions and offers anonymous data collection through the elected representatives among the data owners. The protocol is efficient in tackling internal and external identity disclosure through an adopted clustering-based *k*-anonymity model. We proposed solutions for possible collisions among the elected representatives within the equivalent group. We also proposed a new efficient method to add CSA to protect the real sensitive attributes. Furthermore, dynamic data owners are efficiently handled in the protocol by a greedy heuristic method. Through extensive experimental analysis, we proved that our protocol incurs considerably minimum computational complexity compared with state-of-the-art techniques. This makes our protocol more suitable for collecting huge amounts of healthcare datasets without privacy breach. Our protocol is built to accommodate any *k*-anonymity-based privacy models; hence, the data utility can be optimized as per the requirement.

We intend to conduct several future studies to address the limitations of this study. First, we would like to focus on minimizing the other privacy risks such as attribute disclosure, membership disclosure, and similarity attacks. Currently, our study is focused mainly on protecting personal data from identity disclosure. Considering other privacy attacks would make our protocol more robust for healthcare data collection. Second, we would like to employ anonymization techniques other than *k*-anonymity such as bucketization and anatomy to enhance the data utility of the protocol.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://archive.ics.uci.edu/ml/datasets/adult.

Author contributions

JA and JK conceived the idea and worked on the technical details. JA, RE, and JK devised the work, the main conceptual ideas, the proof outline, and worked on the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

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

Publisher's note

References

1. Varshney U. Pervasive healthcare and wireless health monitoring. *Mob Networks Appl.* (2007) 12:113–27. doi: 10.1007/s11036-007-0017-1

2. Andreassen HK, Bujnowska-Fedak MM, Chronaki CE, Dumitru RC, Pudule I, Santana S, et al. European citizens' use of E-health services: a study of seven countries. *BMC Public Health.* (2007) 7:1–7. doi: 10.1186/1471-2458-7-53

3. Benharref A, Serhani MA. Novel cloud and SOA-based framework for E-health monitoring using wireless biosensors. *IEEE J Biomed Heal Informat.* (2014) 18:46–55. doi: 10.1109/JBHI.2013.2262659

4. Aldosari B. Patients' safety in the era of EMR/EHR automation. *Informatics Med Unlocked.* (2017) 9:230–3. doi: 10.1016/j.imu.2017.10.001

5. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform.* (2008) 77:291–304. doi: 10.1016/j.ijmedinf.2007.09.001

6. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet.* (2012) 13:395–405. doi: 10.1038/nrg3208

7. Saleh S, Abdouni L, Dimassi H, Nabulsi D, Harb R, Jammoul Z, et al. Prevalence of non-communicable diseases and associated medication use among Syrian refugees in Lebanon: an analysis of country-wide data from the Sijilli electronic health records database. *Confl Health.* (2021) 15:77. doi: 10.21203/rs.3.rs-58159/v1

8. Saleh S, Alameddine M, Farah A, Arnaout NE, Dimassi H, Muntaner C, et al. eHealth as a facilitator of equitable access to primary healthcare: the case of caring for non-communicable diseases in rural and refugee settings in Lebanon. *Int Public Health J.* (2018) 63:577–88. doi: 10.1007/s00038-018-1092-8

9. Jha AK, DesRoches CM, Campbell EG, Donelan K, Rao SR, Ferris TG, et al. Use of Electronic Health Records in U.S. Hospitals. *N. Engl. J. Med.* (2009) 360:1628–38. doi: 10.1056/NEJMsa0900592

10. Blumenthal D, Tavenner M. The 'meaningful use' regulation for electronic health records. *N Engl J Med.* (2010) 363:501–4. doi: 10.1056/NEJMp1006114

11. Garde S, Knaup P, Hovenga EJS, Heard S. Towards semantic interoperability for electronic health records: domain knowledge governance for openEHR archetypes. *Methods Inf Med.* (2007) 46:332–43. doi: 10.1160/ME5001

12. Lingren T, Sadhasivam S, Zhang X, Marsolo K. Electronic medical records as a replacement for prospective research data collection in postoperative pain and opioid response studies. *Int J Med Inform.* (2018) 111:45–50. doi: 10.1016/j.ijmedinf.2017.12.014

13. Haas S, Wohlgemuth S, Echizen I, Sonehara N, Müller G. Aspects of privacy for electronic health records. *Int J Med Inform.* (2011) 80:e26–31. doi: 10.1016/j.ijmedinf.2010.10.001

14. Demuynck L, De Decker B. Privacy-preserving electronic health records. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 3677* (Salzburg: Springer Berlin Heidelberg). (2005). p. 150–9.

15. Rizwan M, Shabbir A, Javed AR, Srivastava G, Gadekallu TR, Shabir M, et al. Risk monitoring strategy for confidentiality of healthcare information. *Comput Electr Eng.* (2022) 100:107833. doi: 10.1016/j.compeleceng.2022.107833

16. El Zarif O, Haraty RA. Toward information preservation in healthcare systems. *Innov Heal Informat A Smart Healthc Prim.* (2020) 163–85. doi: 10.1016/B978-0-12-819043-2.00007-1

17. Xue M, Papadimitriou P, Raïssi C, Kalnis P, Pung HK. Distributed privacy preserving data collection. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6587* (Hong Kong: Springer Berlin Heidelberg). (2011). p. 93–107.

18. Andrew J, Karthikeyan J, Jebastin J. Privacy preserving big data publication on cloud using mondrian anonymization techniques and deep neural networks. In: *2019 5th International Conference on Advanced Computing and Communication Systems.* (2019). p. 722–7.

19. Dhasarathan C, Hasan MK, Islam S, Abdullah S, Mokhtar UA, Javed AR, et al. COVID-19 health data analysis and personal data preserving: a homomorphic privacy enforcement approach. *Comput Commun.* (2023) 199:87–97. doi: 10.1016/j.comcom.2022.12.004

20. Haraty RA, Boukhari B, Kaddoura S. An effective hash-based assessment and recovery algorithm for healthcare systems. *Arab J Sci Eng.* (2022) 47:1523–36. doi: 10.1007/s13369-021-06009-4

21. Liu YN, Wang YP, Wang XF, Xia Z, Xu JF. Privacy-preserving raw data collection without a trusted authority for IoT. *Comput Networks.* (2019) 148:340–8. doi: 10.1016/j.comnet.2018.11.028

22. Sei Y, and Okumura AJH, Ohsuga A. Privacy-preserving collaborative data collection and analysis with many missing values. *IEEE Trans. Dependable Secur. Comput.* (2022). doi: 10.1109/TDSC.2022.3174887

23. Krasnova H, Günther O, Spiekermann S, Koroleva K. Privacy concerns and identity in online social networks. *Identity Inf Soc.* (2009) 2:39–63. doi: 10.1007/s12394-009-0019-1

24. Fung BCM, Wang K, Yu PS. Anonymizing classification data for privacy preservation. *IEEE Trans Knowl Data Eng.* (2007) 19:711–25. doi: 10.1109/TKDE.2007.1015

25. Byun JW, Kamra A, Bertino E, Li N. Efficient k-anonymization using clustering techniques. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 4443* (Bangkok: Springer Berlin Heidelberg). (2007). p. 188–200.

26. Zakerzadeh H, Osborn SLAANST. Fast anonymizing algorithm for numerical streaming data. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6514.* (2011). p. 36–50.

27. Prakash M, Singaravel G. An approach for prevention of privacy breach and information leakage in sensitive data mining. *Comput Electr Eng.* (2015) 45:134–40. doi: 10.1016/j.compeleceng.2015.01.016

28. Li HT, Ma JF, Fu SA. privacy-preserving data collection model for digital community. *Sci China Inf Sci.* (2015) 58:1–16. doi: 10.1007/s11432-014-5197-2

29. Yang Z, Zhong S, Wright RN. Anonymity-preserving data collection. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* (2005). p. 334–43.

30. Erlingsson Ú, Pihur V, Korolova A. Rappor: randomized aggregatable privacy-preserving ordinal response. In: *Proceedings of the ACM Conference on Computer and Communications Security.* (2014). p. 1054–67.

31. Kim JW, Jang B, Yoo H. Privacy-preserving aggregation of personal health data streams. *PLoS ONE.* (2018) 13:e0207639. doi: 10.1371/journal.pone.0207639

32. Fung BCM, Wang K, Chen R, Yu PS. Privacy-preserving data publishing: a survey of recent developments. *ACM Comput. Surv.* (2010) 42. doi: 10.1145/1749603.1749605

33. Wang M, Xiao D, Liang J, Hu G. Distributed privacy-preserving nested compressed sensing for multiclass data collection with identity authentication. *Signal Process.* (2023) 204:108823. doi: 10.1016/j.sigpro.2022.108823

34. Zhang W, Jiao H, Yan Z, Wang X, Khan MK. Security analysis and improvement of a public auditing scheme for secure data storage in fog-to-cloud computing. *Comput Secur.* (2023) 125:103019. doi: 10.1016/j.cose.2022.103019

35. Kaaniche N, Laurent M. Data security and privacy preservation in cloud storage environments based on cryptographic mechanisms. *Comp Commun.* (2017) 111:120–41. doi: 10.1016/j.comcom.2017.07.006

36. Puri V, Sachdeva S, Kaur P. Privacy preserving publication of relational and transaction data: survey on the anonymization of patient data? *Comp Sci Rev.* (2019) 32:45–61. doi: 10.1016/j.cosrev.2019.02.001

37. Sweeney L. k-anonymity: a model for protecting privacy. *Int Uncert J Fuzziness Knowl Based Syst.* (2002) 10:557–70. doi: 10.1142/S0218488502001648

38. Machanavajjhala A, Gehrke J, Kifer D, Venkitasubramaniam M. ℓ-Diversity: privacy beyond k-anonymity. *Proc Int Conf Data Eng.* (2006) 2006:24. doi: 10.1109/ICDE.2006.1

39. Ninghui L, Tiancheng L, Venkatasubramanian S. t-closeness: privacy beyond k-anonymity and ℓ-diversity. in *Proceedings - International Conference on Data Engineering.* (2007). p. 106–15.

40. Lin JL, Wei MC. An efficient clustering method for k-anonymization. *ACM Int Conf Proc Ser.* (2008) 331:46–50. doi: 10.1145/1379287.1379297

41. Li H, Guo F, Zhang W, Wang J, Xing J. (a,k)-anonymous scheme for privacy-preserving data collection in IoT-based healthcare services systems. *J Med Syst.* (2018) 42:1–9. doi: 10.1007/s10916-018-0896-7

42. Truta TM, Vinay B. Privacy protection: P-sensitive k-anonymity property. In: *ICDEW 2006 - Proceedings of the 22nd International Conference on Data Engineering Workshops* (2006).

43. Zhong S, Yang Z, Wright RN. Privacy-enhancing k-anonymization of customer data. In: *Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems.* (2005). p. 139–47.

44. Zhong S, Yang Z, Chen T. k-anonymous data collection. *Inf Sci.* (2009) 179:2948–63. doi: 10.1016/j.ins.2009.05.004

45. Hussien AA, Hamza N, Hefny HA. Attacks on anonymization-based privacy-preserving: a survey for data mining and data publishing. *J Inf Secur.* (2013) 04:101–12. doi: 10.4236/jis.2013.42012

46. Beg S, Anjum A, Ahmad M, Hussain S, Ahmad G, Khan S, et al. A privacy-preserving protocol for continuous and dynamic data collection in IoT enabled mobile app recommendation system (MARS). *J Netw Comput Appl.* (2021) 174:102874. doi: 10.1016/j.jnca.2020.102874

47. Beg S, Anjum A, Ahmed M, Malik SUR, Malik H, Sharma N, et al. Dynamic parameters-based reversible data transform (RDT) algorithm in recommendation system. *IEEE Access.* (2021) 9:110011–25. doi: 10.1109/ACCESS.2021.3101150

48. Sajjad H, Kanwal T, Anjum A, Khan A, Khan A, Manzoor U. An efficient privacy preserving protocol for dynamic continuous data collection. *Comput Secur.* (2019) 86:358–71. doi: 10.1016/j.cose.2019.06.017

49. Ni C, Cang LS, Gope P, Min G. Data anonymization evaluation for big data and IoT environment. *Inf Sci.* (2022) 605:381–92. doi: 10.1016/j.ins.2022.05.040

50. Sei Y, Andrew Onesimu J, Ohsuga A. Machine learning model generation with copula-based synthetic dataset for local differentially private numerical data. *IEEE Access.* (2022) 1:101656–101671. doi: 10.1109/ACCESS.2022.3208715

51. Sun S, Ma S, Song JH, Yue WH, Lin XL, Ma T, et al. Experiments and analyses of anonymization mechanisms for trajectory data publishing. *J Comput Sci Technol.* (2022) 375:1026–48. doi: 10.1007/s11390-022-2409-x

52. Canbay Y, Sagiroglu S, Vural YA. new utility-aware anonymization model for privacy preserving data publishing. *Concurr Comput Pract Exp.* (2022) 34:e6808. doi: 10.1002/cpe.6808

53. Onesimu JA, Karthikeyan J, Sei Y. An efficient clustering-based anonymization scheme for privacy-preserving data collection in IoT based healthcare services. *Peer Peer Netw Appl.* (2021) 14:1629–49. doi: 10.1007/s12083-021-01077-7

54. Abbasi A, Mohammadi BA. Clustering-based anonymization approach for privacy-preserving in the healthcare cloud. *Concurr Comput Pract Exp.* (2022) 34:e6487. doi: 10.1002/cpe.6487

55. Onesimu JA, Karthikeyan J, Eunice J, Pomplun M, Dang H. Privacy preserving attribute-focused anonymization scheme for healthcare data publishing. *IEEE Access.* (2022) 10:86979–97. doi: 10.1109/ACCESS.2022.3199433

56. Avraam D, Jones E, Burton PA. deterministic approach for protecting privacy in sensitive personal data. *BMC Med Inform Decis Mak.* (2022) 22:1–17. doi: 10.1186/s12911-022-01754-4

57. Kanwal T, Anjum A, Khan A, Asheralieva A, Jeon GA. Formal adversarial perspective: Secure and efficient electronic health records collection scheme for multi-records datasets. *Trans Emerg Telecommun Technol.* (2021) 32:e4180. doi: 10.1002/ett.4180

58. Kanwal T, Anjum A, Malik SU, Sajjad H, Khan A, Manzoor U, et al. A robust privacy preserving approach for electronic health records using multiple dataset with multiple sensitive attributes. *Comput Secur.* (2021) 105:102224. doi: 10.1016/j.cose.2021.102224

59. Jayapradha J, Prakash M. f-Slip: an efficient privacy-preserving data publishing framework for 1:M microdata with multiple sensitive attributes. *Soft Comput.* (2022) 26:13019–36. doi: 10.1007/s00500-021-06275-2

60. Khan SI, Khan ABA, Hoque ASML. Privacy preserved incremental record linkage. *Big Data J.* (2022) 9:1–27. doi: 10.1186/s40537-022-00655-7

61. Raju NVSL, Naresh VS. Dynamic distributed KCi-slice data publishing model with multiple sensitive attributes. *Concurr Comput Pract Exp.* (2022) 34:e7064. doi: 10.1002/cpe.7064

62. Sarkar A, Senthil Murugan T. Cluster head selection for energy efficient and delay-less routing in wireless sensor network. *Wirel Networks.* (2019) 25:303–20. doi: 10.1007/s11276-017-1558-2

63. Dua D, Graff, C. *UCI Machine Learning Repository: Adult Data Set*. UCI (2017). Available online at: https://archive.ics.uci.edu/ml/datasets/adult (accessed March 2, 2019).

64. *Data – Informsdataminingcontest.* Available online at: https://sites.google.com/site/informsdataminingcontest/data (accessed July 12, 2020).

65. Kim S, Chung YD. An anonymization protocol for continuous and dynamic privacy-preserving data collection. *Futur Gener Comput Syst.* (2019) 93:1065–73. doi: 10.1016/j.future.2017.09.009

66. Oishi K, Sei Y, Tahara Y, Ohsuga A. Semantic diversity: privacy considering distance between values of sensitive attribute. *Comput Secur.* (2020) 94:101823. doi: 10.1016/j.cose.2020.101823

67. Kondepogu MD, Andrew J. Secure E-health record sharing using blockchain: a comparative analysis study. In: *Proc - 2022 6th Int Conf Intell Comput Control Syst ICICCS 2022.* (2022). 861–8.

68. Thapa C, Camtepe S. Precision health data: requirements, challenges and existing techniques for data security and privacy. *Comp Biol Med.* (2021) 129:104130. doi: 10.1016/j.compbiomed.2020.104130

69. Qiu H, Qiu M, Liu M, Memmi G. Secure health data sharing for medical cyber-physical systems for the Healthcare 4.0. *IEEE J Biomed Heal Inf.* (2020) 24:2499–505. doi: 10.1109/JBHI.2020.2973467

70. Zhuang Y, Sheets LR, Chen YW, Shae ZY, Tsai JJP, Shyu CRA, et al. Patient-centric health information exchange framework using blockchain technology. *IEEE J Biomed Heal Informatics.* (2020) 24:2169–76. doi: 10.1109/JBHI.2020.2993072

71. Huang H, Gong T, Ye N, Wang R, Dou Y. Private and secured medical data transmission and analysis for wireless sensing healthcare system. *IEEE Trans Ind Informatics.* (2017) 13:1227–37. doi: 10.1109/TII.2017.2687618

72. Zhang Y, Deng RH, Han G, Zheng D. Secure smart health with privacy-aware aggregate authentication and access control in internet of things. *J Netw Comput Appl.* (2018) 123:89–100. doi: 10.1016/j.jnca.2018.09.005

73. Andrew J, Kathrine GJW. An intrusion detection system using correlation, prioritization and clustering techniques to mitigate false alerts. *Adv Big Data Cloud Comp.* (2018) 645:257–68. doi: 10.1007/978-981-10-7200-0_23

frontiers | Frontiers in Public Health

TYPE Original Research
PUBLISHED 08 March 2023
DOI 10.3389/fpubh.2023.1055815

OPEN ACCESS

EDITED BY
Steven Fernandes,
Creighton University, United States

REVIEWED BY
Amir Faisal,
Sumatra Institute of Technology, Indonesia
Zekun Jiang,
Sichuan University, China

*CORRESPONDENCE
Cien Fan
✉ fce@whu.edu.cn
Meng Wu
✉ wb000713@whu.edu.cn

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 28 September 2022
ACCEPTED 17 February 2023
PUBLISHED 08 March 2023

CITATION
Ma W, Li X, Zou L, Fan C and Wu M (2023) Symmetrical awareness network for cross-site ultrasound thyroid nodule segmentation. *Front. Public Health* 11:1055815. doi: 10.3389/fpubh.2023.1055815

Symmetrical awareness network for cross-site ultrasound thyroid nodule segmentation

Wenxuan Ma[1], Xiaopeng Li[1], Lian Zou[1], Cien Fan[1]* and Meng Wu[2]*

[1]Electronic Information School, Wuhan University, Wuhan, China, [2]Department of Ultrasound, Zhongnan Hospital of Wuhan University, Wuhan, China

Recent years have seen remarkable progress of learning-based methods on Ultrasound Thyroid Nodules segmentation. However, with very limited annotations, the multi-site training data from different domains makes the task remain challenging. Due to domain shift, the existing methods cannot be well generalized to the out-of-set data, which limits the practical application of deep learning in the field of medical imaging. In this work, we propose an effective domain adaptation framework which consists of a bidirectional image translation module and two symmetrical image segmentation modules. The framework improves the generalization ability of deep neural networks in medical image segmentation. The image translation module conducts the mutual conversion between the source domain and the target domain, while the symmetrical image segmentation modules perform image segmentation tasks in both domains. Besides, we utilize adversarial constraint to further bridge the domain gap in feature space. Meanwhile, a consistency loss is also utilized to make the training process more stable and efficient. Experiments on a multi-site ultrasound thyroid nodule dataset achieve 96.22% for PA and 87.06% for DSC in average, demonstrating that our method performs competitively in cross-domain generalization ability with state-of-the-art segmentation methods.

KEYWORDS

thyroid nodule segmentation, thyroid nodule classification, domain adaptation, ultrasound image processing, medical image segmentation

1. Introduction

According to global Cancer statistics 2020 (1), thyroid cancer has become one of the fastest-growing cancers in the past 20 years, ranking in 9th place for incidence. Early symptoms manifest as thyroid nodules, and then as the disease progresses, patients gradually feel pain. If not promptly detected and treated in the early stage, thyroid cancer can cause significant harm to patients and even be life-threatening. Therefore, early and accurate assessment is of crucial importance for improving the chances of cure and survival for patients.

In thyroid diagnosis, ultrasound imaging technique (2) has become the preferred imaging modality due to many advantages such as convenience, good reproducibility, and low cost (3, 4). Usually, radiologists diagnose patients base on the ultrasound characteristics of the images, which requires physicians to have rich experience and superb technology. With the increasing number of thyroid patients year by year, the current demand for radiologists is increasing so fast that it is no longer sufficient to rely on manual diagnosis to meet the needs of society.

Since machine learning and deep learning (5) pervade medical image computing, they are becoming increasingly critical in the medical imaging field, including medical image segmentation (6–12). Leveraging learning-based techniques, multiple novel methods have been proposed to conduct medical image segmentation tasks. Compared with traditional mathematical morphology-based methods, learning-based ones have achieved impressive results. In practice, however, there remain several challenges for deep learning-based methods. One salient problem is that deep learning lacks generalization capability, resulting in models trained on data from one site can not achieve good results on the data from other sites. This is because the ultrasound images from different sites show discrepancies in appearance and contrast due to different imaging protocols, examination equipment and patient groups, which is called domain shift. Due to the existence of domain shift, the model obtained only through deep learning methods cannot be adapted to different sites. In addition, since it is difficult to label medical data, most medical centers still maintain a state where it is difficult to use deep neural network algorithms.

Domain adaptation (13) is a valid approach to address the domain shift problem. The core task of domain adaptation is to tackle the differences in probability distribution between the source domain and the target domain by learning robust knowledge. Existing researches on domain adaptation for medical imaging can be divided into two categories. The first category aims at the feature transfer (14). It learns transferable and discriminative features across different domains (15, 16). Specifically, it maps features from source and target domains to the same distribution *via* certain transforms (17). The second category aims at the model transfer (17). It learns transferable models by fine-tuning or other methods in the target domain (18, 19). One widely-used example is transferring parameters from the pre-trained model on ImageNet (20) to other tasks. Those two categories can greatly improve the generalization capacity of deep learning-based model by dealing with the domain heterogeneity. However, these methods still suffer from certain limitations. First, many methods inevitably require a few labeled target data for fine-tuning. This restricts their performance to unsupervised scenarios. Meanwhile, medical image annotations often require considerable efforts and time. Second, as for model transfer, dataset bias (21) deteriorates the transferring performance. Namely, when the source domain, e.g., ImageNet, differs too much from the target domain, e.g., medical images, this method achieves only average performance. Third, these methods only perform monodirectional domain shift, namely, source to target. Therefore, the image translation functions may lead to undesirable distortions.

In this paper, we propose a domain adaptation framework for medical image segmentation. Our architecture is composed of image translation module and image segmentation module. Medical images from different domains have different styles at the pixel level, and the rule also applies to our source and target domain. Inspired by Cyclegan (22), we use the image translation module to realize the translation between the source domain and the target domain, and guide the process with a pixel-level adversarial loss. Apart from pixel-level alignment, the alignment of semantic features also has a great impact on image segmentation tasks. Therefore, we further unify the style of latent vectors drawn from the segmentation network, and guide the process with a feature-level adversarial loss. The domain gap is well bridged through two-step alignment on both the pixel and feature level. Our segmentation module is constructed into two symmetrical parts to realize the task of segmentation in both the source and target domain, respectively. In each branch, we utilize Efficientnet (23), with strong feature extraction capabilities, to extract deep semantic features and build an image segmentation network based on the encoder-decoder structure. In order to enhance the feature fusion ability and improve the segmentation performance, we use the hybrid channel attention mechanism to concat the features between the encoder and the decoder. Ultimately, considering that the segmentation results from the two branches of the same image should be consistent, we introduce the segmentation consistency loss to further guide the unlabeled branch in an unsupervised manner. In short, our main contributions and novelty of the paper could be summarized as follows:

(1) We propose a domain adaptation framework for medical image segmentation which can narrow the domain gap between different data and effectively improve the generalization ability.
(2) We apply multi-level domain adaptation to simultaneously bridge the domain gap on both pixel-level and feature-level through adversarial learning, and obtain better adaptation results.
(3) Considering the invariance of the segmentation results of the same target in the domain adaptation process, we implement bidirectional symmetric awareness through segmentation consistency loss to further improve the stability and performance of our model.

1.1. Related works

1.1.1. Medical image segmentation

To tackle the medical image segmentation problem, traditional segmentation methods focus on the contour, shape and region properties of thyroid nodules (24–28), while mainstream researchers now focus more on deep learning-based methods. Wang et al. (6) apply multi-instance learning and attention mechanism to automatically detect thyroid nodules in three stages, the feature extraction network, the iterative selection algorithm, and the attention-based feature aggregation network. Peng et al. (7) propose an architecture that combines low-level and high-level features to generate new features with richer information for improving the performance of medical image segmentation. Zhang et al. (8) propose a multiscale mask region-based network to detect lung tumors, which trains multiple models and acquires the final results through weighted voting. Tong et al. (9) propose a novel generative adversarial network-based architecture to segment head and neck cancer. This method uses the shape representation loss and 3D convolutional autoencoder to strengthen the shape consistency of predictions of the segmentation network. Similarly, Trullo et al. (10) propose to use distance-aware adversarial networks to segment multiple organs. This method leverages the global localization information of each organ along with the spatial relationship between them to conduct the task. Li et al. (11) utilize the widely-anticipated transformer to process the medical image.

This method applies squeezed and expanded attention blocks to encode and decode features extracted from CNN. Also, inspired by transformer, Cao et al. (12) propose to conduct image segmentation using a modified transformer-based architecture to improve the performance by combining global branch and local one.

1.1.2. Domain adaptation for medical image analysis

As a promising solution to tackle domain heterogeneity among multi-site medical imaging datasets, domain adaptation has attracted adequate attention in the field. He et al. (29) propose to conduct the domain shift procedure using adversarial network. This method uses a label predictor and a domain discriminator to draw the domain distance closer. Li et al. (18) propose a modified subspace alignment method to diminish the disparity among different datasets, which aligns the sample points from separate feature spaces into the same subspace. Zhang et al. (30) propose the task driven generative adversarial networks to transfer CT images to X-ray images by leveraging a modified cycle-GAN sub-structure with add-on segmentation supervisions to learn the transferable knowledge. Chen et al. (31) propose an unsupervised domain adaptation framework, utilizing synergistic learning-based method to conduct domain translation from MR to CT. Ahn et al. (32) propose an unsupervised feature augmentation method. In this method, image features extracted from a pre-trained CNN are augmented by proportionally combining the feature representation of other similar images. Yoon et al. (33) propose to mitigate dataset bias by extending the classification and contrastive semantic alignment (CCSA) loss that aims to learn domain-invariant features. Dou et al. (34) propose to tackle the domain shift by aligning the feature spaces of source and target domains by utilizing the plug-and-play adaptation mechanism and adversarial learning. Perone et al. (35) propose to conduct domain adaptation in semi-supervised scenarios. Containing teacher models and student models, this method leverages the self-ensembling mechanism to improve the generalization of the models. Gao et al. (36) propose a lesion scale matching approach to use latent space search for bounding box size to resize the source domain images and then match the lesion scales between the two disease domains by utilizing the Monte Carlo Expectation Maximization algorithm. Kang et al. (37) propose intra- and inter-task consistent learning, where task inconsistency is restricted, to have a better performance on all tasks like thyroid nodule segmentation and classification. Gong et al. (38) design a thyroid region prior guided feature enhancement network (TRFEplus) for the purpose of utilizing prior knowledge of thyroid gland segmentation to improve the performance of thyroid nodule segmentation.

2. Materials and methods

2.1. Data acquisition

Our ultrasound thyroid nodule datasets consist of three domain data collected from different patients in different medical centers with different ultrasound systems. The first two datasets are private datasets, which contain 936 and 740 images, respectively, while the third dataset is the public dataset DDTI (39) containing 637 images.

2.2. Method overview

In this work, we aim to build a segmentation network with remarkable cross-domain generalization ability. Specifically, given a labeled dataset $X_S = \{x_s\}_{s=1}^{N_S}$ in source domain and an unlabeled dataset $X_T = \{x_t\}_{r=1}^{N_T}$ in target domain, where N_S and N_T denote the number of images, we assume that they obey the marginal distributions $P_S(x_s)$ and $P_T(x_t)$. The domain adaptation problem can be defined as mapping X_S and X_T to corresponding latent spaces *via* $F_S^{Enc}: X_S \rightarrow Z_S$, $F_T^{Enc}: X_T \rightarrow Z_T$, respectively. The representations Z_S and Z_T are desired to obey the same distribution, so that $Z_S \approx Z_T$. Consequently, we present a novel bidirectional symmetric segmentation framework, as is shown in Figure 1. Designed to close the domain gap on both the pixel level and feature level, our framework is divided into one bidirectional image translation module and two symmetric image segmentation modules. At the pixel level, we introduce the image translation module, i.e., $G_{S\rightarrow T}$ and $G_{T\rightarrow S}$, where $G_{S\rightarrow T}$ translates images from source domain to target domain while $G_{T\rightarrow S}$ performs image translation inversely. Considering of the fact that semantic information has a more profound impact on image segmentation, we propose to unify the style of latent vectors drawn from the segmentation network on the feature level. Given x_t and $x_{s\rightarrow t}$, the segmentation module is utilized to encode them into latent codes z_t and $z_{s\rightarrow t}$. And an adversarial discriminator is utilized to close their domain gap, encouraging them to obey the same distribution. Consistent with the bidirectional translation module, we introduce two symmetrical segmentation branches, i.e., F_S and F_T, to respectively achieve segmentation in the source and target images. In each branch, the Efficientnet (23) and the hybrid channel attention mechanism are introduced to enhance the feature extraction and fusion capability of our segmentation modules. Besides, the segmentation results of x_t and $x_{t\rightarrow s}$ should be the same because they represent the same content information. The rule also applies to the relationship between x_s and $x_{s\rightarrow t}$. Therefore, we introduce the segmentation consistency loss to further guide the network.

2.3. Network architecture

2.3.1. Translation module

The image translation modules are designed to close the domain gap on both the pixel and feature level. Each module consists of one encoder network and the corresponding decoder network. As is shown in Figure 2, the source image is first fed into the encoder to generate the latent codes, which is then decoded into generated image by the decoder. The process also applies to the translation of target image. Besides, the 8-layer residual blocks are utilized to improve the network's learning ability. The encoder consists of one convolution block mapping image to high-dimension space, two downsampling convolution blocks with stride 2 and residual blocks. In terms of the decoder, we

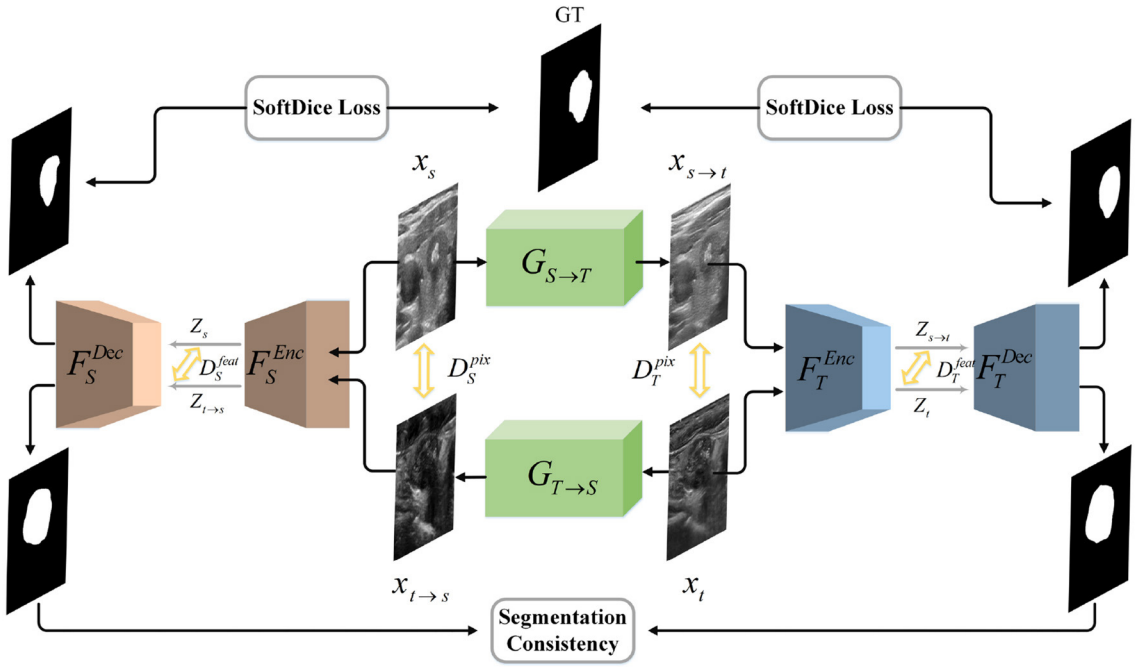


FIGURE 1
Illustration of our network architecture. We design a bidirectional translation module to respectively generate $x_{t\to s}$ and $x_{s\to t}$. We introduce adversarial loss to unify the style of translated and target domain on both the pixel and feature level. Then we utilize two symmetrical segmentation modules, i.e., F_S and F_T, to respectively achieve segmentation in two domains. SoftDice loss and segmentation consistency loss are proposed to further conduct the network.

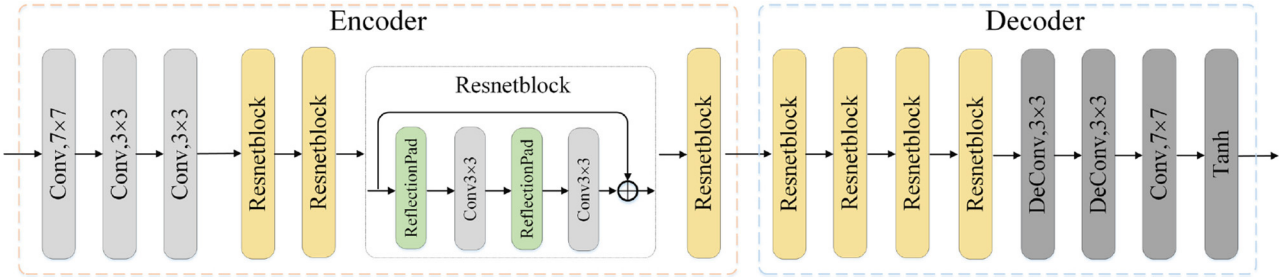


FIGURE 2
Illustration of one translation module architecture. The input is the image x of source or target domain and the output is the translated image. Conv: convolution layer, Deconv: deconvolution layer, Tanh: Hyperbolic Tangent.

utilize two trainable deconvolution layers instead of the traditional upsampling blocks to improve the translation performance.

2.3.2. Segmentation module

Based on the image segmentation framework UNet, our segmentation modules adopt the EfficientNet-B0 as feature extraction network, also the hybrid attention mechanism to enhance the expression ability of fusion feature and further improve the segmentation performance. The architecture of our segmentation modules is shown in Figure 3. We construct the modules with an encoder-decoder architecture, where the encoder is utilized to extract multi-scale feature maps while the decoder translates the low-resolution feature maps back into images of original size. The EfficientNet-B0 module in the encoder mainly uses the Mobile Inverted Bottleneck convolution layer to extract features related to the target, and alternately uses the MBConv modules with different convolution kernel sizes to expand the receptive field. To make full use of the location information contained in the shallow features, we integrate the skip connection mechanism into the encoder-decoder architecture to fuse the shallow features from encoder and the deep features from decoder. After that, we feed the fusion feature into the channel attention module and the spatial attention module, respectively, to obtain the feature maps whose channel and spatial semantic information are calibrated. By adding the two calibrated feature maps, new features with global dependence come into being.

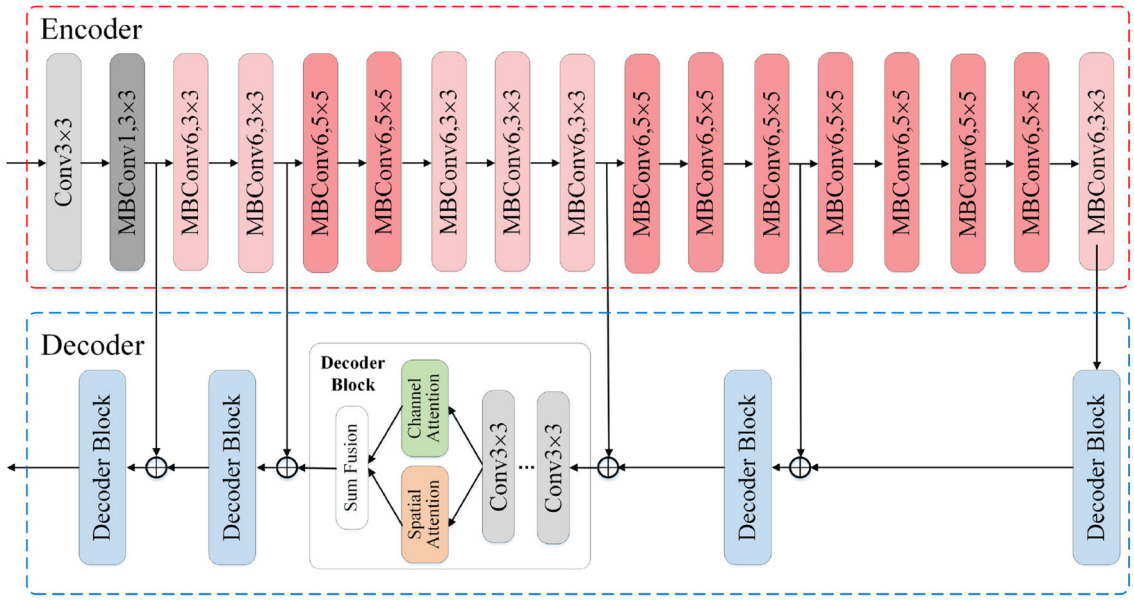


FIGURE 3
Illustration of the segmentation module architecture. The input is the original images or the ones translated by image translation module, and the output is the segmentation result. The encoder adopts EfficientNet-B0 as the feature extraction skeleton. In the decoder we incorporate the hybrid attention mechanism and skip connection mechanism.

2.4. Loss functions

2.4.1. Image translation loss

As is discussed above, we utilize the image translation modules to bridge the domain gap on both the pixel and feature level. The discriminators incorporated denote two feature-level discriminators (D_S^{feat} and D_T^{feat}), and two pixel-level discriminators (D_S^{pix} and D_T^{pix}). Take the source domain discriminators for example, the D_S^{feat} is adopted to narrow the gap between the feature map of the source image x_s and the translated image $x_{t\to s}$, while the D_S^{pix} is adopted to close the gap on the pixel level. The adversarial losses of source domain are shown as follows.

$$\begin{aligned}\mathcal{L}_S^{\text{adv}} &= \mathbb{E}\left[-\log\left(1-D_S^{\text{feat}}\left(x_{t\to s}\right)\right)\right]+\mathbb{E}\left[-\log\left(D_S^{\text{feat}}\left(x_s\right)\right)\right]\\ &+\mathbb{E}\left[-\log\left(1-D_S^{pix}\left(x_{t\to s}\right)\right)\right]+\mathbb{E}\left[-\log\left(D_S^{pix}\left(x_s\right)\right)\right].\end{aligned} \tag{1}$$

Similar to the source domain adversarial losses, the adversarial losses of target domain are shown as follows.

$$\begin{aligned}\mathcal{L}_T^{\text{adv}} &= \mathbb{E}\left[-\log\left(1-D_T^{\text{feat}}\left(x_{s\to t}\right)\right)\right]+\mathbb{E}\left[-\log\left(D_T^{\text{feat}}\left(x_t\right)\right)\right]\\ &+\mathbb{E}\left[-\log\left(1-D_T^{pix}\left(x_{s\to t}\right)\right)\right]+\mathbb{E}\left[-\log\left(D_T^{pix}\left(x_t\right)\right)\right].\end{aligned} \tag{2}$$

Besides, we introduce the cycle-consistency loss (22) to further conduct the translation module. Explicitly, if we feed the image x_s to $G_{S\to T}$ and then to $G_{T\to S}$, the result obtained should be the same as the original image x_s. The rule also applies to the image x_t. The cycle-consistency loss is shown as follows.

$$\mathcal{L}^{\text{cycle}} = \mathbb{E}\left[\left\|G_{T\to S}\left(x_{s\to t}\right)-x_s\right\|_1\right]+\mathbb{E}\left[\left\|G_{S\to T}\left(x_{t\to s}\right)-x_t\right\|_1\right]. \tag{3}$$

Moreover, we also adopt an identity mapping loss (22) to prevent the generators from producing undesired results. For instance, the result of feeding the image x_s to $G_{T\to S}$ should be indistinguishable from the original input x_s. The loss is shown as follows.

$$\mathcal{L}^{\text{iden}} = \mathbb{E}\left[\left\|G_{T\to S}\left(x_s\right)-x_s\right\|_1\right]+\mathbb{E}\left[\left\|G_{S\to T}\left(x_t\right)-x_t\right\|_1\right]. \tag{4}$$

2.4.2. Segmentation loss

The segmentation modules are divided into two parts to realize image segmentation in the source and target domain, respectively. Since the source domain image x_s is labeled, we train the segmentation process of x_s and $x_{s\to t}$ in a supervised manner. We utilize Dice loss as the supervised loss, which is defined as follows.

$$\mathcal{L}^{\text{seg}} = \mathcal{L}_{DICE}\left(\mathcal{F}_S\left(x_s\right),\text{GT}\right)+\mathcal{L}_{DICE}\left(\mathcal{F}_T\left(G_{S\to T}\left(x_s\right)\right),\text{GT}\right). \tag{5}$$

In terms of the target domain image x_t which is unlabeled, we train its segmentation process in an unsupervised manner and propose a consistency loss, which is shown as follows.

$$\mathcal{L}^{\text{consis}} = \mathcal{L}_{DICE}\left(\mathcal{F}_T\left(x_t\right),\mathcal{F}_S\left(G_{T\to S}\left(x_t\right)\right)\right). \tag{6}$$

Our consistency loss aims at guide the unlabeled branch with the supervised networks in labeled branch.

The overall loss function is defined as follows:

$$\begin{aligned}L &= \lambda_{adv}\left(\mathcal{L}_S^{adv}+\mathcal{L}_T^{\text{adv}}\right)+\lambda_{\text{cycle}}\mathcal{L}^{\text{cycle}}+\lambda_{\text{iden}}\mathcal{L}^{\text{iden}}+\lambda_{\text{seg}}\mathcal{L}^{\text{seg}}\\ &+\lambda_{\text{consis}}\mathcal{L}^{\text{consis}},\end{aligned} \tag{7}$$

Where the hyper-parameters $\lambda_{adv},\lambda_{cycle},\lambda_{iden},\lambda_{seg},\lambda_{consis}$ denote the weight of each term.

2.5. Metrics

In the task of medical image segmentation, each pixel in the image can be divided into *Positive*, which means that exact area belongs to Thyroid Nodule, and *Negative* with the opposite meaning. For any image segmentation method, there would be *TruePositive*, *TrueNegative*, *FalsePositive* and *FalseNegative* representing 4 types of relationship between the result and the ground truth, denoted by *TP*, *TN*, *FP*, *FN*. To verify how well our method can tackle the medical image segmentation problem, the following evaluating metrics are utilized.

Pixel Accuracy(*PA*): The most commonly utilized metric in image segmentation task. It can be seen as the accuracy in pixel level, defined as the percentage of the correctly predicted pixels in total pixels. Formally, *PA* is defined as follows.

$$PA = \frac{TP + TN}{TP + FP + FN + TN}. \tag{8}$$

Dice Similarity Coefficient(*DSC*) (40): Another wildly utilized metric in image segmentation problems. It can be used to evaluate the similarity of two groups, which represent predict result and ground truth in this situation. Formally, *DSC* is defined as follows.

$$DSC = \frac{2TP}{FP + 2TP + FN}. \tag{9}$$

True Positive Rate(*TPR*): Defined as the percentage of the correctly predicted pixels of positives in total positive pixels given by ground truth. In this case it can represent the ability of detecting positive area, thus known as Sensitivity. Formally, *TPR* is defined as follows.

$$TPR = \frac{TP}{TP + FN}. \tag{10}$$

True Negative Rate(*TNR*): Defined as the percentage of the correctly predicted pixels of negatives in total negative pixels given by ground truth. In this case it can represent the ability of not being confused by negative area, thus known as Specificity. Formally, *TNR* is defined as follows.

$$TNR = \frac{TN}{TN + FP}. \tag{11}$$

3. Experimental results and discussion

3.1. Implementation

3.1.1. Data processing

Considering that each image frame contains unnecessary text which would affect the image segmentation performance, we only preserve the part with content information and remove the remaining. After that, we unify the size of the cropped image to 400*400. From each dataset, we randomly choose 200 images as the test set, others as the train set.

3.1.2. Training details

During the training process, in order to ensure that our network can work effectively, we optimize our network step by step. The whole process is divided into three steps. Firstly, the image translation module is trained with adversarial loss to get optimized $G_{S\to T}$ and $G_{T\to S}$. Then given the inputs x_s and $G_{S\to T}(x_s)$, we train source domain segmentation network F_S and target domain segmentation network F_T, respectively in a supervised manner. Finally, under the constraint of the total loss L, we carry out a more refined optimization of the whole network. In the training process, we set $\lambda_{adv} = \lambda_{cycle} = \lambda_{iden} = 1, \lambda_{seg} = 10, \lambda_{consis} = 2$. The whole experiment is carried out with four 1080Ti.

3.1.3. Comparison methods

We compare our network with the following state-of-the-art methods: DeepLabV3 (41), PSPNet (42), FPN, PAN (43), and TRFEplus (38).

3.2. Ultrasound thyroid nodule segmentation

We carry out the experiment of ultrasound thyroid nodule segmentation to evaluate our proposed method. Three ultrasound thyroid nodule datasets with annotations are labeled as domain1, domain2, and domain3, respectively. Each time we select two of them, one of which is used as the source domain and the other is used as the target domain, then 6 sets of experiments are conducted.

We compare our method with several state-of-the-art methods using the four evaluating metrics above to compare the performance. For all the compared methods, we carry out comprehensive data augmentation to improve their generalization ability. Besides the commonly used geometric transformations, we also introduce noise, blur, occlusion, etc. to improve their performance as much as possible. Meanwhile, we exclude any data augmentation strategy in our method and only employ the domain adaptation architecture. For our method, we conduct two sets of experiment, one of which is applying data augmentation to our segmentation module only (namely SegM+AUG), and the other is our proposed domain adaptation framework. It should be noted that we exclude the data augmentation strategy from the latter scheme because the data augmentation will change the style of images and thus influence the translation process on pixel level.

Utilizing the metrics mentioned in Section 2.5, the results are presented in Tables 1–7. As can be seen from the tables, our method performs more favorably against other methods, especially in the most representative metric *DSC*, which confirms the feasibility of our domain adaptation method. Specifically, our method increases the average *PA* by 0.99%, the average *DSC* by 3.64% and the average *TNR* by 0.24%. It is worth noting that if the confusing pixels are judged as negative ones mostly, the *FP* will be greatly reduced, and the *FN* will be greatly increased as a cost. That is to say, this strategy improves the *TPR* by sacrificing *TNR*, which makes the results of TRFEplus (38) in Table 5 close to ours on *TNR* but far behind ours on *TPR*. From Tables 6, 7, we can see that the performance of the network applied domain adaptation is better than the one with data augmentation strategy.

TABLE 1 Results on ultrasound thyroid nodule datasets with DeepLabV3 (41) +AUG.

Source	Target	PA	DSC	TPR	TNR
Domain1	Domain2	0.9505	0.8461	0.9186	0.9531
Domain1	Domain3	0.9494	0.8579	0.9450	0.9477
Domain2	Domain1	0.8348	0.6991	0.9962	0.8003
Domain2	Domain3	0.7858	0.5444	0.9910	0.7596
Domain3	Domain1	0.9681	0.9071	0.8638	0.9939
Domain3	Domain2	0.9669	0.8623	0.8298	0.9892
Average		0.9093	0.7862	**0.9241**	0.9073

Bold values mean that the value is the best among all the comparison methods and our method.

TABLE 2 Results on ultrasound thyroid nodule datasets with PSPNet (42) +AUG.

Source	Target	PA	DSC	TPR	TNR
Domain1	Domain2	0.8581	0.6287	0.9336	0.8474
Domain1	Domain3	0.7955	0.5766	0.9220	0.7801
Domain2	Domain1	0.7835	0.6215	0.9610	0.7460
Domain2	Domain3	0.7263	0.4673	0.9582	0.6974
Domain3	Domain1	0.8740	0.9249	0.8903	0.9942
Domain3	Domain2	0.8902	0.3878	0.3037	0.9952
Average		0.8213	0.5678	0.8281	0.8434

TABLE 3 Results on ultrasound thyroid nodule datasets with FPN +AUG.

Source	Target	PA	DSC	TPR	TNR
Domain1	Domain2	0.9434	0.8190	0.8910	0.9512
Domain1	Domain3	0.9641	0.8993	0.9611	0.9607
Domain2	Domain1	0.8645	0.7425	0.9958	0.8360
Domain2	Domain3	0.8256	0.5947	0.9923	0.8047
Domain3	Domain1	0.9740	0.9249	0.8903	0.9942
Domain3	Domain2	0.9670	0.8582	0.7972	0.9942
Average		0.9231	0.8064	0.9213	0.9235

We further show the qualitative comparison of the methods in Figure 4. As can be seen, while other methods either fail to segment the nodules or over-segment a large portion of nodules, our method generates more accurate segmentation results.

3.3. Ablation study

To verify the advancement of our medical image translation module and the effectiveness of the consistency loss, we carry out a series of ablation experiments as follows: (a) **w/o translation module:** disabling the whole image translation module during training. (b) **w/o feature-level GAN:** disabling feature-level adversarial loss during training. (c) **w/o consistency loss:** disabling consistency loss during training; (d) **w/o DA:** only segmentation module.

TABLE 4 Results on ultrasound thyroid nodule datasets with PAN (43) +AUG.

Source	Target	PA	DSC	TPR	TNR
Domain1	Domain2	0.9270	0.8136	0.9463	0.9219
Domain1	Domain3	0.9449	0.8557	0.9663	0.9382
Domain2	Domain1	0.8915	0.7720	0.9931	0.8692
Domain2	Domain3	0.8378	0.6156	0.9900	0.8189
Domain3	Domain1	0.9691	0.9076	0.8661	0.9939
Domain3	Domain2	0.9618	0.8341	0.7633	0.9946
Average		0.9220	0.7998	0.9209	0.9228

TABLE 5 Results on ultrasound thyroid nodule datasets with TRFEplus (38) + AUG.

Source	Target	PA	DSC	TPR	TNR
Domain1	Domain2	0.9597	0.8633	0.9451	0.9602
Domain1	Domain3	0.9637	0.8609	0.8588	0.9783
Domain2	Domain1	0.9473	0.8408	0.7743	0.9855
Domain2	Domain3	0.9431	0.7641	0.7242	0.9779
Domain3	Domain1	0.9570	0.8756	0.8304	0.9874
Domain3	Domain2	0.9430	0.8007	0.8580	0.9611
Average		0.9523	0.8342	0.8318	0.9750

TABLE 6 esults on ultrasound thyroid nodule datasets with SegM + AUG.

Source	Target	PA	DSC	TPR	TNR
Domain1	Domain2	0.9518	0.8443	0.9109	0.9566
Domain1	Domain3	0.9749	0.9105	0.9482	0.9773
Domain2	Domain1	0.8747	0.7548	0.9939	0.8486
Domain2	Domain3	0.8260	0.5932	0.9921	0.8054
Domain3	Domain1	0.9714	0.9155	0.8726	0.9955
Domain3	Domain2	0.9656	0.8548	0.8075	0.9910
Average		0.9274	0.8122	0.9209	0.9291

TABLE 7 Results on ultrasound thyroid nodule datasets with ours.

Source	Target	PA	DSC	TPR	TNR
Domain1	Domain2	0.9730	0.8947	0.9147	0.9839
Domain1	Domain3	0.9744	0.9063	0.9573	0.9795
Domain2	Domain1	0.9676	0.9106	0.9558	0.9724
Domain2	Domain3	0.9394	0.7778	0.9691	0.9408
Domain3	Domain1	0.9614	0.8949	0.8243	0.9977
Domain3	Domain2	0.9576	0.8392	0.8132	0.9899
Average		**0.9622**	**0.8706**	0.9057	**0.9774**

Bold values mean that the value is the best among all the comparison methods and our method.

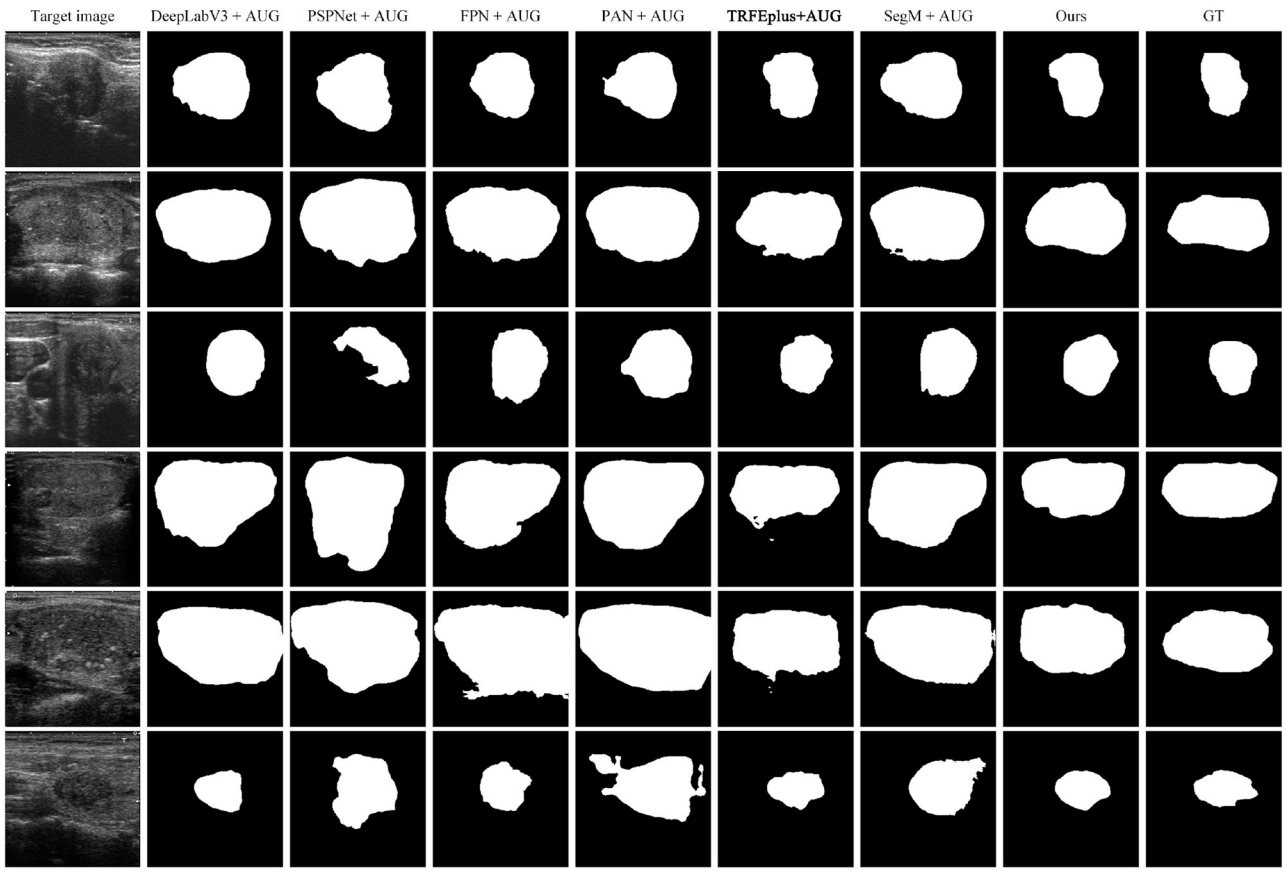


FIGURE 4
Comparison of cross-domain segmentation results on target domain. From left to right: target image, DeepLabV3+AUG, PSPNet+AUG, FPN+AUG, PAN+AUG, TRFEpuls+AUG, SegM+AUG, Ours, GT.

TABLE 8 Ablation study on various constraints.

Ablation (w/o)	Translation module	Feature-level GAN	Consistency loss	DA	Ours
PA	0.9632	0.9644	0.9163	0.8613	**0.9730**
DSC	0.8488	0.8541	0.7069	0.6141	**0.8947**

Bold value means that the value is the best in the ablation study.

The results of our ablation experiments are demonstrated in Table 8. From the results, we can summarize the following conclusions. (a) In the absence of translation module and feature-level GAN, *DSC* drops by 0.0459 and 0.0406, respectively, which proves that they play a certain role in improving the segmentation results and are of equal importance. (b) In the absence of consistency loss, *DSC* drops sharply by 0.18, which indicates that the segmentation consistency loss plays a decisive role in our network's performance. (c) When we only use our segmentation module to complete cross-domain tasks, the effect is not satisfactory, namely 0.6141 in *DSC*. It illustrates the effectiveness of our domain adaptation framework. In conclusion, our ablation experiments indicate that the proposed medical image translation module, the consistency loss and closing in the feature space are helpful to close the domain gap between source data and target data.

4. Conclusion

In this paper, we have presented a domain adaptation method for medical image segmentation. In order to alleviate the domain shift problem caused by the difference in data styles, we propose to bridge the domain gap between multi-site medical data on both the pixel and feature level. Meanwhile, we introduce two symmetrical hybrid-attention segmentation modules to segment the source domain data and target domain data, respectively. Besides, we construct the segmentation consistency loss to guarantee the model stability. Experimental results on Ultrasound Thyroid Nodules

datasets show the remarkable generalization ability of our proposed method.

Data availability statement

The private datasets presented in this article are not readily available due to policy. Requests to access the datasets should be directed to corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by Medical Ethics Committee, Zhongnan Hospital of Wuhan University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

WM, XL, CF, LZ, and MW: conceptualization, writing—original draft preparation, and writing—review and editing. WM, XL, and CF: methodology and investigation. WM, XL, CF, and LZ: validation. WM, XL, CF, and MW: formal analysis. WM, CF, LZ, and MW: resources. WM, XL, and MW: data curation. WM and XL: visualization. CF, LZ, and MW: supervision and project administration. MW: funding acquisition. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The numerical calculations in this article have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

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.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2021) 71:209–49. doi: 10.3322/caac.21660

2. Sarvazyan AP, Rudenko OV, Swanson SD, Fowlkes JB, Emelianov SY. Shear wave elasticity imaging: a new ultrasonic technology of medical diagnostics. *Ultrasound Med Biol*. (1998) 24:1419–35. doi: 10.1016/S0301-5629(98)00110-0

3. Faisal A, Ng SC, Goh SL, George J, Supriyanto E, Lai KW. Multiple LREK active contours for knee meniscus ultrasound image segmentation. *IEEE Trans Med Imaging*. (2015) 34:2162–71. doi: 10.1109/TMI.2015.2425144

4. Faisal A, Ng SC, Goh SL, Lai KW. Knee cartilage segmentation and thickness computation from ultrasound images. *Med Biol Eng Comput*. (2018) 56:657–69. doi: 10.1007/s11517-017-1710-2

5. Nielsen MA. *Neural Networks and Deep Learning. Vol. 25*. San Francisco, CA: Determination Press (2015).

6. Wang L, Zhang L, Zhu M, Qi X, Yi Z. Automatic diagnosis for thyroid nodules in ultrasound images by deep neural networks. *Med Image Anal*. (2020) 61:101665. doi: 10.1016/j.media.2020.101665

7. Peng D, Yu X, Peng W, Lu J. DGFAU-Net: global feature attention upsampling network for medical image segmentation. *Neural Comput Appl*. (2021) 33:12023–37. doi: 10.1007/s00521-021-05908-9

8. Zhang R, Cheng C, Zhao X, Li X. Multiscale mask R-CNN-based lung tumor detection using PET Imaging. *Mol Imaging*. (2019) 18:1536012119863531. doi: 10.1177/1536012119863531

9. Tong N, Gou S, Yang S, Cao M, Sheng K. Shape constrained fully convolutional DenseNet with adversarial training for multiorgan segmentation on head and neck CT and low-field MR images. *Med Phys*. (2019) 46:2669–82. doi: 10.1002/mp.13553

10. Trullo R, Petitjean C, Dubray B, Ruan S. Multiorgan segmentation using distance-aware adversarial networks. *J Med Imaging*. (2019) 6:014001. doi: 10.1117/1.JMI.6.1.014001

11. Li S, Sui X, Luo X, Xu X, Liu Y, Goh RSM. Medical image segmentation using squeeze-and-expansion transformers. *arXiv preprint arXiv:210509511*. (2021) doi: 10.24963/ijcai.2021/112

12. Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al. Swin-Unet: unet-like pure transformer for medical image segmentation. *arXiv preprint arXiv:210505537*. (2021). doi: 10.1007/978-3-031-25066-8_9

13. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*. (2009) 22:1345–59. doi: 10.1109/TKDE.2009.191

14. Liu B, Wang X, Dixit M, Kwitt R, Vasconcelos N. Feature space transfer for data augmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT: IEEE (2018). p. 9090–8.

15. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: *European Conference on Computer Vision*. Springer (2014). p. 818–33.

16. Oquab M, Bottou L, Laptev I, Sivic J. Learning and transferring mid-level image representations using convolutional neural networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH: IEEE (2014). p. 1717–24.

17. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. In: *International Conference on Artificial Neural Networks*. Springer (2018). p. 270–9.

18. Li W, Zhao Y, Chen X, Xiao Y, Qin Y. Detecting Alzheimer's disease on small dataset: a knowledge transfer perspective. *IEEE J Biomed Health Inform*. (2018) 23:1234–42. doi: 10.1109/JBHI.2018.2839771

19. Cheng B, Liu M, Zhang D, Munsell BC, Shen D. Domain transfer learning for MCI conversion prediction. *IEEE Trans Biomed Eng*. (2015) 62:1805–17. doi: 10.1109/TBME.2015.2404809

20. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL: IEEE (2009). p. 248–55.

21. Torralba A, Efros AA. Unbiased look at dataset bias. In: *CVPR 2011*. Colorado Springs, CO: IEEE (2011). p. 1521–8.

22. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the IEEE International Conference on Computer Vision*. Venice: IEEE (2017). p. 2223–32.

23. Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. In: *International Conference on Machine Learning*. PMLR (2019). p. 6105–14.

24. Du W, Sang N. An effective method for ultrasound thyroid nodules segmentation. In: *2015 International Symposium on Bioelectronics and Bioinformatics (ISBB)*. Beijing: IEEE (2015). p. 207–10.

25. Mylona EA, Savelonas MA, Maroulis D. Automated adjustment of region-based active contour parameters using local image geometry. *IEEE Trans Cybern*. (2014) 44:2757–70. doi: 10.1109/TCYB.2014.2315293

26. Tuncer SA, Alkan A. Segmentation of thyroid nodules with K-means algorithm on mobile devices. In: *2015 16th IEEE International Symposium on Computational Intelligence and Informatics (CINTI)*. Budapest: IEEE (2015). p. 345–48.

27. Zhao J, Zheng W, Zhang L, Tian H. Segmentation of ultrasound images of thyroid nodule for assisting fine needle aspiration cytology. *Health Inf Sci Syst*. (2013) 1:1–12. doi: 10.1186/2047-2501-1-5

28. El Naqa I, Yang D, Apte A, Khullar D, Mutic S, Zheng J, et al. Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning a. *Med Phys*. (2007) 34:4738–49. doi: 10.1118/1.2799886

29. He C, Wang S, Kang H, Zheng L, Tan T, Fan J. Adversarial domain adaptation network for tumor image diagnosis. *Int J Approx Reason*. (2021) 135:38–52. doi: 10.1016/j.ijar.2021.04.010

30. Zhang Y, Miao S, Mansi T, Liao R. Task driven generative modeling for unsupervised domain adaptation: application to x-ray image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer (2018). p. 599–607.

31. Chen C, Dou Q, Chen H, Qin J, Heng PA. Synergistic image and feature adaptation: towards cross-modality domain adaptation for medical image segmentation. In: *Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33*;. (2019). p. 865–72.

32. Ahn E, Kumar A, Fulham M, Feng D, Kim J. Unsupervised domain adaptation to classify medical images using zero-bias convolutional auto-encoders and context-based feature augmentation. *IEEE Trans Med Imaging*. (2020) 39:2385–94. doi: 10.1109/TMI.2020.2971258

33. Yoon C, Hamarneh G, Garbi R. Generalizable feature learning in the presence of data bias and domain class imbalance with application to skin lesion classification. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer (2019). p. 365–73.

34. Dou Q, Ouyang C, Chen C, Chen H, Glocker B, Zhuang X, et al. Pnp-adanet: plug-and-play adversarial domain adaptation network with a benchmark at cross-modality cardiac segmentation. *arXiv preprint arXiv:181207907*. (2018) doi: 10.1109/ACCESS.2019.2929258

35. Perone CS, Ballester P, Barros RC, Cohen-Adad J. Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. *Neuroimage*. (2019) 194:1–11. doi: 10.1016/j.neuroimage.2019.03.026

36. Gao J, Lao Q, Kang Q, Liu P, Zhang L, Li K. Unsupervised cross-disease domain adaptation by lesion scale matching. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer (2022). p. 660–70.

37. Kang Q, Lao Q, Li Y, Jiang Z, Qiu Y, Zhang S, et al. Thyroid nodule segmentation and classification in ultrasound images through intra-and inter-task consistent learning. *Med Image Anal*. (2022) 79:102443. doi: 10.1016/j.media.2022.102443

38. Gong H, Chen J, Chen G, Li H, Chen F, Li G. Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules. *Comput Biol Med*. (2022) 106389:1–12. doi: 10.1016/j.compbiomed.2022.106389

39. Pedraza L, Vargas C, Narváez F, Durán O, Mu noz E, Romero E. An open access thyroid ultrasound image database. In: *10th International Symposium on Medical Information Processing Analysis. vol. 9287*. International Society for Optics Photonics (2015). p. 92870W.

40. Milletari F, Navab N, Ahmadi SA. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: *2016 fourth international conference on 3D vision (3DV)*. Stanford, CA: IEEE (2016). p. 565–71.

41. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. (2018) p. 801–18.

42. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid scene parsing network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI: IEEE (2017). p. 2881–90.

43. Li H, Xiong P, An J, Wang L. Pyramid attention network for semantic segmentation. *arXiv preprint arXiv:180510180*. (2018).

frontiers | Frontiers in Public Health

TYPE Opinion
PUBLISHED 31 March 2023
DOI 10.3389/fpubh.2023.1110088

OPEN ACCESS

EDITED BY
Xi Zhu,
Columbia University, United States

REVIEWED BY
Kathiravan Srinivasan,
Vellore Institute of Technology, India

*CORRESPONDENCE
J. Andrew
✉ andrew.j@manipal.edu

SPECIALTY SECTION
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

RECEIVED 28 November 2022
ACCEPTED 20 March 2023
PUBLISHED 31 March 2023

CITATION
Andrew J, Rudra M, Eunice J and Belfin RV (2023) Artificial intelligence in adolescents mental health disorder diagnosis, prognosis, and treatment.
Front. Public Health 11:1110088.
doi: 10.3389/fpubh.2023.1110088

Artificial intelligence in adolescents mental health disorder diagnosis, prognosis, and treatment

J. Andrew[1]*, Madhuria Rudra[2], Jennifer Eunice[3] and R. V. Belfin[4]

[1]Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India, [2]Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka, India, [3]Electronics and Communication Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India, [4]BRIC, School of Medicine, University of North Carolina, Chapel Hill, NC, United States

KEYWORDS

artificial intelligence, machine learning, deep learning, mental health, adolescent, neuroimaging

1. Introduction

Social, psychological, and emotional wellbeing are all considered to be components of one's mental health. It affects how someone thinks, feels, and responds to circumstances. When one has good mental health, it is easier to perform efficiently and reach their full potential (1). Preschool, adolescence, and adulthood are all included in the definition of mental health. Anxiety, social phobia, depression, panic disorder, substance dependence, and specific illnesses are factors that contribute to mental health problems that result in mental illness. The mental health status of adolescents in India is a topic of great concern and importance. Adolescents are children within the age group of 10–19. According to the National Mental Health Survey of India (2015–2016) (2), the prevalence of psychiatric disorders among adolescents of ages 13–17 is 7.3% and in the US it is 27.9% (3). This problem is further aggravated by stigma and lack of awareness surrounding mental health and a treatment gap of 95% in common mental disorders which is greater than the treatment gap for severe disorders (76%).

Early diagnosis of mental health issues is a crucial step for improving patient care and understanding of mental health diseases (4). However, the diagnosis of mental disorders is challenging because sometimes it is hard to distinguish between mental and physical health. The diagnosis of mental illnesses is reliance on a person's self-reporting to targeted interrogations used for the identification of particular patterns of public interactions and emotions. Unlike other chronic illnesses, this does not require laboratory testing or measures (5). While treating mental health problems, a challenge that is often faced is the lack of mechanistic models for psychiatric disorders. The complexity of the brain renders it difficult to have a clear model describing the development of a given mental illness (6).

The recent advancements in Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) are widely applied in many domains and have made a successful dent in the healthcare sector too. The performance of AI, ML, and DL and based on the availability of data. Since mental health data availability is increasing every day AI techniques can be applied to better understand mental health (7). Recently in healthcare sectors, DL techniques provide superior performance in diagnosis, prognosis, and treatments. In medicine, AI is used in the form of machine learning models to understand medical data to gain insights in order to improve health outcomes and patient experience. Mental health professionals employ AI to fulfill various tasks such as providing insights into therapy sessions for better quality, refining diagnosis and monitoring patient's condition and altering treatment as deemed necessary (8).

2. Deep learning techniques

Deep Learning (DL) techniques are popular for their layered architecture that can learn the complex features of the data. DL techniques generally contains an input layer that takes input features, output layer that gives the learnt output and multiple hidden layers finds the relationship between the input and output classes. All layers perform the mathematical computations to find the relationship between various patterns of the input and output data. In this section, some of the popular DL techniques are introduced.

2.1. Neural networks

Neural networks are also called Artificial Neural Networks (ANN) that mimic human brain neurons. Each neuron is can perform a mathematical operation. The human brain has millions of neurons that are interconnected together. The main idea of ANN is to mimic the human brain like activity. Artificial neurons forward the weighted sum of the inputs through an activation function such as sigmoid, rectified linear units (ReLU) and so on for the non-linear transformation of the data. A simple ANN consists of a single hidden layer where DNN (Deep Neural Networks) consists of multiple hidden layers. The number of hidden layers increase the complexity of the network and it is helpful when the data is complex. The other variants of neural networks are feed forward neural networks, and back propagation neural networks.

2.2. Convolutional neural networks

Convolutional Neural Network (CNN) is a type of neural network, that comprises of convolutional operations and neural network layers (9). CNN is basically designed for image classification where it takes the image pixels as input and performs convolution and pooling operations to learn the low and high level features from the images and correspond them to the output-label. The basic architecture of a CNN contains a convolution layer, pooling layers, and fully connected (FC) layers. Pooling layers are used for edge detection and feature detection through max, min and average pooling operations. FC layer is a neural network where each neuron is connected to all the neurons in the next layers and forms a network. CNN models yield superior performance with image data hence, CNN can be used in the diagnosis of mental disorder through medical image analysis or through facial recognition. The variants of CNN models are Conv1D, Conv2D, and Conv3D.

2.3. Recurrent neural networks

Recurrent Neural Networks (RNN) are specifically designed for data that has temporal features. RNN maintains the temporal relationship between the features by feeding them to a recurrent neuron. RNN can extract features and long-term dependencies from sequential and time-series data (10). LSTM (Long Short Term Memory) and GRU (Gated Recurrent Units) are some of the well-known RNN models. LSTM can be used to capture the dynamic temporal features of brain signals for the diagnosis of mental disorders. GRU is less complex compared to LSTM so based on the size of the dataset the type of RNN model can be selected.

2.4. Generative networks

Generative models have gained attention for their ability to create synthetic data. As the success of the DL techniques are purely based on the availability of the data, generative networks help in creating synthetic data when there is less data availability, or to treat missing values and useful in data augmentation. Generative Adversarial Network (GAN), Variational Autoencoders (VAE) are some of the popular generative models (11). GAN is a combination of generator and discriminator networks where the model can recreate a look-a-like data from the learned features. VAE has encoder and decoders in the network, the encoder learns the latent spatial information from the input data and the decoder reconstructs the original data from the learnt representation.

3. Mental disorders in adolescents

Among adolescents, anxiety disorders and mood disorders are the most common. While anxiety and mood issues are two to three times more prevalent in female adolescents than in male adolescents, attention deficit disorder affects male and female adolescents differently. Adolescent patients of obstetricians and gynecologists are more likely to have one or more mental health issues than younger patients (12). The following are some of the common mental disorders.

3.1. Anxiety disorders

The most common anxiety disorders are General Anxiety Disorder (GAD) and panic disorder, which affect 3.6% of 10–14-year-olds and 4.6% of 15–19-year-olds, respectively. Excessive anxiety and disproportionate worry over a variety of things are symptoms of GAD.

3.2. Mood disorders

According to estimates, mood disorders such Major Depressive Disorder (MDD) and bipolar disorder impact 2.8% of teenagers aged 15–19 and 1.1% of adolescents aged 10–14. MDD is defined as a minimum 2-week period marked by either a loss of interest in previously enjoyed activities or a melancholy mood.

3.3. Attention deficit hyperactivity disorder

Attention Deficit Hyperactivity Disorder (ADHD) occurs among 3.1% of 10–14 year-olds and 2.4% of 15–19 year-olds. It is

characterized by maladaptive and inconsistent with developmental level symptoms of inattention, hyperactivity, or impulsivity.

3.4. Obsessive compulsive disorder

An individual's obsessions or compulsions can lead to obsessive compulsive disorder (OCD). Obsessions are characterized as unwelcome intrusive ideas that recur frequently in a person's head. Compulsions, on the other hand, are compulsive mental patterns of thought or behavior that one feels compelled to engage in.

4. Diagnostic tools

Typically, mental diseases are identified based on a person's self-report on certain questionnaires made to look for certain emotional or social interaction patterns. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) is the diagnostic instrument that is most frequently used. A youngster or teenager may find it challenging to maintain concentration for more than 2–3 h during such conventional diagnostic interviews. Therefore, neuroimaging, which captures brain activity, serves as an alternative to such interviews. The diagnosis of various types of mental illnesses in adolescents is aided by understanding of brain activity.

4.1. MRI

Magnetic resonance imaging (MRI), a non-invasive imaging procedure, produces precise, three-dimensional anatomical images (13). Functional MRI (fMRI) and Structural MRI (sMRI) are the two types of neuroimages that are used for diagnosing mental disorders. Human brain activity changes are captured in fMRI data as Blood Oxygenation Level Dependent (BOLD) signals. The structured representation of the brain is shown in 3D in sMRI data together with its spatial configurations and voxel intensities. In state-of-the-art research, DL techniques are successful in analyzing neuroimage data.

4.2. EEG data

Electroencephalogram (EEG) signals are electrical signals that represent brain activity. EEG signals can be recorded using small low-cost sensors. The sensors record the electrical signals generated by the brain. EEG signals are instrumental in diagnosing various types of health disorders. It also plays a major role in diagnosing mental disorders.

4.3. PET

A functional imaging technique called a positron emission tomography (PET) scan employs radioactive materials called radiotracers to track and evaluate changes in physiological processes like blood flow, regional chemical composition, and absorption. PET helps in the development of diagnostic biomarkers for mental health diseases.

4.4. MEG

Magnetoencephalography (MEG) is similar to EEG that records brain activity through electrical signals. MEG is superior to EEG which has better spatial resolution with less artifacts. MEG can also be used for the diagnosis of psychiatric disorders.

4.5. Genetic data

Numerous common and uncommon genetic variations, such as single nucleotide polymorphisms (SNPs), have been linked to mental health issues using conventional statistical research in genetics and genomics, such as genome-wide association studies. Recent advances in next-generation sequencing technologies have generated a large amount of high-throughput genome or exome sequencing data, enabling researchers to examine patients with mental health illnesses by examining a variety of genetic changes across an individual's genome.

5. AI for mental disorder diagnosis

Artificial Intelligence has been found to be advantageous in the process of diagnosis of medical disorders. With the application of various techniques involving machine learning such as Boltzmann machine, support vector machine (SVM), K-Nearest Neighbor (kNN), diseases are detected and diagnosed (14, 15). Machine learning and computer vision, which are subsets of AI, are widely used due to their imaging, segmentation and predictability capabilities. Machine learning approaches are successful due to their focus on the performance of the models on new data which is also known as generalizability. Computer vision is applied in the process of diagnosis by detecting, segmenting and classifying images e.g., segmentation of radiological images and then further classifying them into diagnostic categories, detection of presence of metastases etc. Schizophrenia is serious mental disorder where a person is in hallucinations or delusion and interprets reality differently. To diagnose such disorder, Khan et al. (16) proposed a deep neural network which takes the genome sequencing data as the input and learns the feature representation of the data to diagnose the Schizophrenia disorder. The proposed model has attained an area under the curve (AUC) of 0.57.

AI is also widely applicable in diagnosis of mental health conditions (17). The diagnosis of a new patient is predicted using the training dataset of the diagnosis of the previous patients. Furthermore, artificial intelligence can also differentiate between diagnosis of diseases with similar symptoms but divergent methods of treatment. This is observed in the instance when bipolar or unipolar depression (18) is to be identified based on brain imaging features and types of dementia are to be differentiated based on structural MRI scans (19). Data-driven AI methods based on

TABLE 1 AI techniques for mental disorder diagnosis, prognosis, and treatment.

Reference	Application	AI technique	Mental disorder type	Dataset type	Performance	Strength/weakness
Mikolas et al. (31)	Diagnosis	Support vector machine (SVM)	Schizophrenia	sMRI	ACC: 62.34% ($p = 0.005$)	Early detection of mental health disorders
Uyulan et al. (32)		CNN	Depression	EEG	ACC: 92.66% AUC: 0.9	Potential biomarker for confirming mood disorders
Drysdale et al. (33)		Clustering	Depression	fMRI	ACC: 89.2%	Helps to identify individuals who can benefit from neurostimulation therapies
Luo et al. (34)		Ensemble model	ADHD	Multimodal	ACC: 78.3% AUC: 0.89	Potential neurobiological markers for neurodevelopmental disorders
Zou et al. (35)		3D CNN	ADHD	fMRI	ACC: 65.67%	Potential neurobiological markers for neurodevelopmental disorders.
Khan et al. (16)		DNN	Schizophrenia	Genome sequencing data	AUC: 0.57	Identification of genes that causes mental disorder
Yan et al. (36)		LSTM	Schizophrenia	fMRI	ACC: 83.2%	Potential deep-chronnectome-learning with time courses
Koutsouleris et al. (21)	Prognosis	SVM	Depression	Multimodal	ACC: 85.9%	Prediction of psychosis transition
Geng and Xu (23)		Autoencoder and CNN	Depression	fMRI	ACC: 0.95	Considered 8 causality measures for depression prognosis
Sheynin et al. (37)		DNN	Post-traumatic stress disorder (PTSD)	fMRI	ACC: 81.3% AUC: 0.84	Single patient characterization is lacking
Nieuwenhuis et al. (24)		CNN	Schizophrenia	sMRI	ACC: 89% ($p < 0.001$)	Lack of large homogeneous data for multi-center prognosis
Smucny et al. (25)		Neural networks	Schizophrenia	fMRI	ACC: 70% MSE: 9.47	Potential biomarker for schizophrenia and psychosis treatment
Koutsouleris et al. (28)	Treatment	SVM	Psychosis	sMRI	ACC: 82%	Potential biomarker for prodromal phase of psychosis
Chang et al. (26)		SVM	Depression	fMRI and sMRI	ACC: 87.4% ($p < 0.001$)	Lack of multimodal data consideration that can improve the model efficiency
Acharya et al. (29)		CNN	Depression	EEG	ACC: 0.935	Different stages of depression is not explored
Zou et al. (30)		CNN	ADHD	fMRI and sMRI	ACC: 0.692	Potential biomarker for ADHD disorders
Pinaya et al. (38)		Autoencoder	Schizophrenia	sMRI	ACC: 0.639–0.707	Efficient in calculating neuroanatomical deviations in neuropsychiatric populations

various factors such as demographic features, neurocognitive and biomarker profiles can aid in identifying novel disease subtypes (20). Moreover, AI methodologies can decipher patterns from data stemming from a long time span, which is critical for validating the accuracy of diagnoses based on the evolution of psychiatric conditions over a span of time. Sen et al., utilized a generative network called autoencoder that learns the spatial latent representation of the data. The model worked with fMRI and sMRI data for the diagnosis of ADHD. The model used multimodal features of the input to improve the classification accuracy up to 67%.

6. AI for mental disorder prognosis

The strong predictive power and generalization ability of AI is useful in the prognosis of mental disorders. AI techniques can utilize psychiatric patient's longitudinal data that is observed over a period of time for accurate prognoses. The longitudinal data can be neuroimaging data, genetic genomic data, and electronic health records (EHR). Koutsouleris et al. (21) collected MRI and genotyping data of psychiatric patients with high-risk syndromes and recent onset depression for the prognosis of psychosis mental disorder. For the prediction of psychosis disorder SVM technique is

utilized which performs the classification of high-risk, recent-onset, and normal control from the learnt features from multimodal data.

Another type of brain data that can be used to perform the prognosis of mental disorder is EEG signals. In Ref. (22), Zhdanov et al., have used resting state EEG signals to predict the effectiveness of escitalopram treatment for the patients with depression. The authors used radial basis function (RBF) SVM to estimate the treatment effectiveness there by performing a prognosis study of the patients. The proposed model had achieved an accuracy of 82.4% for the classification among different severities of the disorders. Geng and Xu (23) and Nieuwenhuis et al. (24) analyzed the neuroimaging data using DL techniques such as autoencoder and CNN. They utilized fMRI and sMRI data to perform prognosis study of patients with depression and Schizophrenia disorders. Smucny et al. (25) studied the clinical improvement of psychosis patients from the fMRI data. A variety of ML and DL models are used to compare the performance of the techniques on the prognostic study is presented. The author has observed that multilayer perceptron or neural networks have performed well on the fMRI data for the accurate prognosis of psychosis patients with an accuracy of 70% and MSE (Mean Squared Error) of 9.47.

7. AI for mental disorder treatment

AI can be effectively used to measure the response to various treatments. Also, it can play a role in predicting the response of various drug combination which can help to develop precision medicines. Chang et al. (26) have measured the effectiveness of methylphenidate for treating ADHD from neuroimaging biomarkers. A multivariate pattern recognition approach is developed to measure the differences in the volumetric information in the sMRI. SVM is used as the classifier to perform the binary classification between the two classes of different volumetric measures. Li et al. (27) measures the gray matter volumetric (GMV) correlates in adolescents ADHD. Machine learning models are utilized to classify the normal and malignant samples of sMRI. Koutsouleris et al. (28) have used neuroanatomical pattern classification to predict the mental disorder transition. Psychosis (at-risk mental states) subjects sMRI data are collected and classified using SVM. The different classes for the classifiers are early at-risk, late at-risk, and healthy controls. The model achieved an accuracy of 82%in classification.

CNN-based DL models are efficient in classifying neuroimaging data. Acharya et al. (29) utilized EEG signals for automated depression screening. The authors have proposed a deep CNN model to classify the EEG signals of normal and depressive subjects. The model considered EEG signals of left and right hemisphere of the brain. The performance of the model was 93% and 95% respectively. Zou et al. (30) proposed a 3D CNN model to learn the low level and complex features form fMRI and sMRI neuroimaging data for ADHD treatments. The model was able to achieve an accuracy of 69.15% with limited training samples. Table 1 shows an overview of the various AI techniques in mental disorder diagnosis, prognosis, and treatment for neuroimaging data.

8. Limitations of AI in mental health

The clinical validity and readiness for implementation of AI applications in clinical decision-making and patient care are both constrained. The performance of any AI application is affected by the quality and size of the data, an example of which is overfitting which is caused due to small and limited sample sizes. The generalizability of ML models is further restricted by testing them only using data from the same sample and not out-of-sample. The input features such as clinical data and demographics also restrict the predictive capability of such researches. Since no study is exhaustive enough to consider all factors, the clinical efficiency of the features used to derive the models must be considered. The validity of the outputs of the algorithms can be applicable only to a certain group of people or in specific situations (39).

Binary classifiers are more commonly used instead of regressors due to the ease of training them. However, using this approach neglects the severity of a condition. Hence, studies in future should take into consideration the severity of the condition. Another challenge that is faced by studies is modeling rare events or illnesses of highly imbalanced datasets. For such instances, classifiers predict the majority. However, this can be overcome by methods such as over-sampling and under-sampling. Over-sampling involves matching ratios of the minority and majority by duplicating minority samples meanwhile under-sampling involves reducing the number of majority samples. There are also ensemble learning methods which are combining several methods in order to reduce variance and improve predictions.

Moreover, mental health treatment relies on soft skills like rapport and relationship building with the patients and observing their emotions and behavior. The absence of crucial components like human compassion and empathy is also a hindrance in the treatment process. Further research is necessary on AI applications to prevent it from working in an unpredictable manner. Extensive risk assessment and administrative oversight is required before being put into medical use as it should be able to handle unusual situations.

Patient privacy is at risk due to the increased use of online servers. Online databases and the devices they are stored at are vulnerable to hacking and unauthorized monitoring. In India, the Ministry of Health and Family Welfare released a draft of Healthcare Security Act, which penalizes breach of data and promotes maintenance of electronic records. However, there is still a lack of guidelines regarding the assistance received by mental health professionals on delivering AI services. There are no laws defined to hold the software developer accountable for an technological glitches. Its other drawbacks include, most prominently the lack of human empathy and compassion, which are crucial components when treating patients who have suffered mental trauma or are experiencing a mental condition.

8.1. Future directions

AI struggles to self-reflect and account for the diversity amongst people, their thoughts, perspectives and morality. To deal with the human compassion and empathy limitations of AI, the concept

of Artificial Wisdom has been introduced. Wisdom is associated with greater societal and individual wellbeing. The idea of creating AI that has human societal values and wise is something that is yet to be explored. It is unlikely that human wisdom will ever be fully programmed into machines, but partial instances are present in the form of robots being physical therapists and social workers along with providing cognitive assistance to the elderly (40). These developments are along the direction of creating machines who employ wise principles to make wise decisions. The advancements of AI wisdom can play a major role mental healthcare in the future.

Author contributions

JA and MR conceived the idea. JA, MR, JE, and RVB devised the work, the main conceptual ideas, the proof outline, and worked on the manuscript. JA and MR worked on the technical details. All authors contributed to the article and approved the submitted version.

Conflict of interest

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

Publisher's note

References

1. Artificial intelligence in prediction of mental health disorders induced by the COVID-19 pandemic among health care workers. *Croat Med J.* (2020) 61:279. doi: 10.3325/cmj.2020.61.279

2. Murthy RS. National mental health survey of India 2015-2016. *Indian J Psychiatry.* (2017) 59:21–26. doi: 10.4103/psychiatry.IndianJPsychiatry_102_17

3. Kessler RC, Avenevoli S, McLaughlin KA, Green JG, Lakoma MD, Petukhova M, et al. Lifetime co-morbidity of DSM-IV disorders in the US National Comorbidity Survey Replication Adolescent Supplement (NCS-A). *Psychol Med.* (2012) 42:1997–2010. doi: 10.1017/S0033291712000025

4. Du M, Liu S, Wang T, Zhang W, Ke Y, Chen L, et al. Depression recognition using a proposed speech chain model fusing speech production and perception features. *J Affect Disord.* (2023) 323:299–308. doi: 10.1016/j.jad.2022.11.060

5. Su C, Xu Z, Pathak J, Wang F. Deep learning in mental health outcome research: a scoping review. *Transl Psychiatry.* (2020) 10:1–26. doi: 10.1038/s41398-020-0780-3

6. Rosenfeld A, Benrimoh D, Armstrong C, Mirchi N, Langlois-Therrien T, et al. *Big Data Analytics and AI in Mental Healthcare*. Amsterdam: Elsevier (2019), p. 137–7. doi: 10.48550/arxiv.1903.12071

7. Khan DM, Yahya N, Kamel N, Faye I. A novel method for efficient estimation of brain effective connectivity in EEG. *Comput Methods Programs Biomed.* (2023) 228:107242. doi: 10.1016/j.cmpb.2022.107242

8. Lillo E, Mora M, Lucero B. Automated diagnosis of schizophrenia using EEG microstates and Deep Convolutional Neural Network. *Expert Syst Appl.* (2022) 209:118236. doi: 10.1016/j.eswa.2022.118236

9. Mhathesh TSR, Andrew J, Martin Sagayam K, Henesey L. A 3D convolutional neural network for bacterial image classification. In: Peter J, Fernandes S, Alavi A, editors. *Advances in Intelligent Systems and Computing*. New York, NY: Springer (2021), p. 419–31. doi: 10.1007/978-981-15-5285-4_42

10. Mikolov T, Karafiát M, Burget L, Jan C, Khudanpur S. Recurrent neural network based language model. In: *Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010*. Chiba (2010). p. 1045–48. doi: 10.21437/Interspeech.2010-343

11. Andrew J, Mhatesh TSR, Sebastin RD, Sagayam KM, Eunice J, Pomplun M, Dang H. Super-resolution reconstruction of brain magnetic resonance images *via* lightweight autoencoder. *Informatics Med Unlocked.* (2021) 26:100713. doi: 10.1016/j.imu.2021.100713

12. Malhotra S, Chakrabarti S, Shah R, Mehta A, Gupta A, Sharma M, et al. novel screening and diagnostic tool for child and adolescent psychiatric disorders for telepsychiatry. *Indian J Psychol Med.* (2015) 37:288 Psy doi: 10.4103/0253-7176.162921

13. Magnetic Resonance Imaging (MRI). Available online at: https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri (accessed November 26, 2022).

14. Lee EE, Torous J, De Choudhury M, Depp CA, Graham SA, Kim H-C, et al. Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biol Psychiatry Cogn Neurosci Neuroimaging.* (2021) 856–864. doi: 10.1016/j.bpsc.2021.02.001

15. Mahendran N, Durai Raj Vincent PM, Srinivasan K, Sharma V, Jayakody DK. Realizing a stacking generalization model to improve the prediction accuracy of major depressive disorder in adults. *IEEE Access.* (2020) 8:49509–950 doi: 10.1109/ACCESS.2020.2977887

16. Khan A, Liu Q, Wang K. iMEGES: integrated mental-disorder GEnome score by deep neural network for prioritizing the susceptibility genes for mental disorders in personal genomes. *BMC Bioinformatics.* (2018) 19:501. doi: 10.1186/s12859-018-2469-7

17. Uusitalo S, Tuominen J, Arstila V. Mapping out the philosophical questions of AI and clinical practice in diagnosing and treating mental disorders. *J Eval Clin Pract.* (2021) 27:478–84. doi: 10.1111/jep.13485

18. Srinivasan K, Mahendran N, Vincent DR, Chang CY, Syed-Abdul S. Realizing an integrated multistage support vector machine model for augmented recognition of unipolar depression. *Electron.* (2020) 9:647. doi: 10.3390/electronics9040647

19. Nichols JA, Herbert Chan HW, Baker MAB. Machine learning: applications of artificial intelligence to imaging and diagnosis. *Biophys Rev.* (2019) 11:111–118. doi: 10.1007/s12551-018-0449-9

20. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med.* (2019) 49:1426–426 doi: 10.1017/S0033291719000151

21. Koutsouleris N, Dwyer DB, Degenhardt F, Maj C, Urquijo-Castro MF, Sanfelici R, et al. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry.* (2021) 78:195–209. doi: 10.1001/jamapsychiatry.2020.3604

22. Zhdanov A, Atluri S, Wong W, Vaghei Y, Daskalakis ZJ, Blumberger DM, et al. Use of Machine learning for predicting escitalopram treatment outcome from electroencephalography recordings in adult patients with depression. *JAMA Netw Open.* (2020) 3:e1918377. doi: 10.1001/jamanetworkopen.2019.18377

23. Geng X-F, Xu J-H. Application of autoencoder in depression diagnosis. *DEStech Trans Comput Sci Eng.* (2017) 146–51. doi: 10.12783/dtcse/csma2017/17335

24. Nieuwenhuis M, Schnack HG, van Haren NE, Kahn RS, Lappin J, Dazzan P, et al. Multi-center MRI prediction models: predicting sex and illness course in first episode psychosis patients. *Neuroimage.* (2017) 145:246–253. doi: 10.1016/j.neuroimage.2016.07.027

25. Smucny J, Davidson I, Carter CS. Comparing machine and deep learning-based algorithms for prediction of clinical improvement in psychosis with functional magnetic resonance imaging. *Hum Brain Mapp.* (2021) 42:1197–197: doi: 10.1002/hbm.25286

26. Chang JC, Lin HY, Lv J, Tseng WYI, Gau SSF. Regional brain volume predicts response to methylphenidate treatment in individuals with ADHD. *BMC Psychiatry.* (2021) 21:1–14 doi: 10.1186/s12888-021-03040-5

27. Li CS, Chen Y, Ide JS. Gray matter volumetric correlates of attention deficit and hyperactivity traits in emerging adolescents. *Sci Rep*. (2022) 12:2022.05.16.492088. doi: 10.1038/s41598-022-15124-7

28. Koutsouleris N, Meisenzahl EM, Davatzikos C, Bottlender R, Frodl T, Scheuerecker J, et al. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch Gen Psychiatry*. (2009) 66:700. doi: 10.1001/archgenpsychiatry.2009.62

29. Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H, Subha DP. Automated EEG-based screening of depression using deep convolutional neural network. *Comput Methods Programs Biomed*. (2018) 161:103. doi: 10.1016/j.cmpb.2018.04.012

30. Zou L, Zheng J, Miao C, McKeown MJ, Wang ZJ. 3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI. *IEEE Access*. (2017) 5:23626–362. doi: 10.1109/ACCESS.2017.2762703

31. Mikolas P, Hlinka J, Skoch A, Pitra Z, Frodl T, Spaniel F, Hajek T. Machine learning classification of first-episode schizophrenia spectrum disorders and controls using whole brain white matter fractional anisotropy. *BMC Psychiatry*. (2018) 18:1–7. doi: 10.1186/s12888-018-1678-y

32. Uyulan C, Erginka J, Skoch A, Pitra Z, Frodl T, Spaniel F, et al. Major depressive disorder classification based on different convolutional neural network models: deep learning approach. *Clin EEG Neurosci*. (2021) 52:38G. doi: 10.1177/15500594209 16634

33. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med*. (2017) 23:28–38. doi: 10.1038/nm.4246

34. Luo Y, Alvarez TL, Halperin JM, Li X. Multimodal neuroimaging-based prediction of adult outcomes in childhood-onset ADHD using ensemble learning techniques. *NeuroImage Clin*. (2020) 26:102238. doi: 10.1016/j.nicl.2020.1 02238

35. Zou L, Zheng J, Mckeown MJ. Deep learning based automatic diagnoses of attention deficit hyperactive disorder. In: *2017 IEEE Glob Conf Signal Inf Process Glob 2017 - Proc*. Montreal, QC (2018). p. 962–6. doi: 10.1109/GlobalSIP.2017.8309103

36. Yan W, Calhoun V, Song M, Cui Y, Yan H, Liu S, et al. Discriminating schizophrenia using recurrent neural network applied on time courses of multi-site FMRI data. *EBioMedicine*. (2019) 47:543–552. doi: 10.1016/j.ebiom.2019.08.023

37. Sheynin S, Wolf L, Ben-Zion Z, Sheynin J, Reznik S, Keynan JN, et al. Deep learning model of fMRI connectivity predicts PTSD symptom trajectories in recent trauma survivors. *Neuroimage*. (2021) 238:118242. doi: 10.1016/j.neuroimage.2021.118242

38. Pinaya WHL, Mechelli A, Sato JR. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: a large-scale multi-sample study. *Hum Brain Mapp*. (2019) 40:944–954. doi: 10.1002/hbm.24423

39. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, et al. Artificial intelligence for mental health and mental illnesses: an overview. *Curr Psychiatry Rep*. (2019) 21:1–8. doi: 10.1007/s11920-019-1094-0

40. Graham SA, Lee EE, Jeste DV, Van Patten R, Twamley EW, Nebeker C, et al. Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: a conceptual review. *Psychiatry Res*. (2020) 284:112732. doi: 10.1016/j.psychres.2019.112732

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 26 June 2023
DOI 10.3389/fonc.2023.1105224

OPEN ACCESS

EDITED BY
Chuanming Li,
Chongqing University Central Hospital, China

REVIEWED BY
Kathiravan Srinivasan,
Vellore Institute of Technology, Vellore, India
Yinlong Yuan,
Nantong University, China

*CORRESPONDENCE
Hu Luo
luohucy@163.com
Xiangdong Zhou
xiangdongzhou@126.com

RECEIVED 22 November 2022
ACCEPTED 06 June 2023
PUBLISHED 26 June 2023

CITATION
Yi X, Xu W, Tang G, Zhang L, Wang K, Luo H and Zhou X (2023) Individual risk and prognostic value prediction by machine learning for distant metastasis in pulmonary sarcomatoid carcinoma: a large cohort study based on the SEER database and the Chinese population. *Front. Oncol.* 13:1105224. doi: 10.3389/fonc.2023.1105224

Individual risk and prognostic value prediction by machine learning for distant metastasis in pulmonary sarcomatoid carcinoma: a large cohort study based on the SEER database and the Chinese population

Xinglin Yi[1], Wenhao Xu[2], Guihua Tang[1], Lingye Zhang[1], Kaishan Wang[3], Hu Luo[1]* and Xiangdong Zhou[1]*

[1]Department of Respiratory Medicine, Southwest Hospital of Third Military Medical University, Chongqing, China, [2]Department of Urinary Medicine Center, Southwest Hospital of Third Military Medical University, Chongqing, China, [3]Department of Neurosurgery Department, Southwest Hospital of Third Military Medical University, Chongqing, China

Background: This study aimed to develop diagnostic and prognostic models for patients with pulmonary sarcomatoid carcinoma (PSC) and distant metastasis (DM).

Methods: Patients from the Surveillance, Epidemiology, and End Results (SEER) database were divided into a training set and internal test set at a ratio of 7 to 3, while those from the Chinese hospital were assigned to the external test set, to develop the diagnostic model for DM. Univariate logistic regression was employed in the training set to screen for DM-related risk factors, which were included into six machine learning (ML) models. Furthermore, patients from the SEER database were randomly divided into a training set and validation set at a ratio of 7 to 3 to develop the prognostic model which predicts survival of patients PSC with DM. Univariate and multivariate Cox regression analyses have also been performed in the training set to identify independent factors, and a prognostic nomogram for cancer-specific survival (CSS) for PSC patients with DM.

Results: For the diagnostic model for DM, 589 patients with PSC in the training set, 255 patients in the internal and 94 patients in the external test set were eventually enrolled. The extreme gradient boosting (XGB) algorithm performed best on the external test set with an area under the curve (AUC) of 0.821. For the prognostic model, 270 PSC patients with DM in the training and 117 patients in the test set were enrolled. The nomogram displayed precise accuracy with AUC of 0.803 for 3-month CSS and 0.869 for 6-month CSS in the test set.

Conclusion: The ML model accurately identified individuals at high risk for DM who needed more careful follow-up, including appropriate preventative therapeutic strategies. The prognostic nomogram accurately predicted CSS in PSC patients with DM.

KEYWORDS

machine learning, SEER, Pulmonary sarcomatoid carcinoma, risk, prognosis

1 Introduction

Pulmonary sarcomatoid carcinoma (PSC) is rare among all lung malignancies, with an incidence of 0.1–0.5% (1, 2). According to the World Health Organization classification guidelines in 2021, PSC consists of giant cell carcinoma, pleomorphic carcinoma, spindle cell carcinoma, pulmonary blastoma, and carcinosarcoma (3). Clinically, newly diagnosed patients with PSCs are more likely to be male (77%), smokers (84%), and of advanced age, with a median age of 68 years at diagnosis (4). Similar to other subtypes of non-small cell lung carcinoma (NSCLC), the initial symptoms of PSC include cough, chest pain, weight loss, and dyspnea (5).

PSC is a highly aggressive malignancy. Recent studies have reported that the 5-year survival rate ranges from approximately 15% to 20% (6, 7), which is considerably lower than that in other subtypes of NSCLC. Targeted drugs, especially focusing on mesenchymal-epithelial transition (MET) exon 14 skipping mutations, have been shown to be beneficial in improving the median survival time to 10 months (8). However, distant metastasis (DM) still results in most patients having a shorter survival time (9–11). Owing to the rapid invasion into vasculature, 40–60% of PSC patients are diagnosed with DM at first presentation (6, 9, 12). The most common sites of DM are bone (16%), lungs (15%), brain (12%), and liver (8%), while 62% of patients present with multiple DM sites (13). The 1- and 3-year overall survival (OS) probabilities of PSC patients with DM are reported to be only 14.1% and 5.5%, respectively, which are significantly lower than those of patients without DM (58.2% and 38.1%, respectively) (13). Therefore, it is of great clinical significance to identify those at risk of developing DM upon initial stage of diagnosis.

Computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography-computed tomography (PET-CT) are commonly used diagnostic modalities for screening for DM in newly hospitalized patients with PSC. However, these screening methods are difficult to apply to all hospitalized PSC patients due to their high cost, including time-related costs. Meanwhile, due to its rare incidence, the prognostic factors of PSC patients with DM remain unclear. Consequently, research investigating the factors influencing survival is important as it can inform and guide clinical strategies. Since the emergence of big data analysis and machine learning (ML), it is possible to provide an alternative option for factors screening. There have been several predictive models with outstanding performance being built to apply in clinical practice by using big data and ML (14–16). The Surveillance, Epidemiology, and End Results (SEER) database (https://seer.cancer.gov/) covers geographically diverse patients with detailed information on the patients' clinicopathological statistics and follow-up visits, providing an abundance of medical cases to analyze. This real-world-based database offers a golden opportunity to develop or examine ML models in this field. However, to our knowledge, no studies have focused on establishing models for the prediction and prognosis of DM in PSC patients.

Therefore, this study aimed to establish and validate diagnostic and prognostic models based on ML algorithms and Cox regression through large population and ML. Additionally, this study aimed to offer personalized predictors that could serve as effective tools for clinicians to preliminarily evaluate the risk and prognosis of PSC with DM, saving patients from unnecessary costs.

2 Materials and methods

2.1 Data source and characteristics

Clinicopathological information of patients with PSC diagnosed between 2004 and 2018 was collected from the SEER database. Additionally, clinicopathological data from Southwest Hospital (2004–2022) in China was retrospectively obtained using an electronic medical record system. Inclusion criteria comprised the primary disease site and morphology, based on the 6th Edition of the American Joint Committee on Cancer (AJCC) "Lung," with the following International Classification of Diseases for Oncology 3rd Edition (ICD-O-3) codes: 8022/3, 8031/3, 8032/3, 8072/3, or 8980/3. Exclusion criteria comprised patients whose information on laterality, primary site, tumor-node-metastasis (TNM) stage, histology, marital status, and tumor size was unknown; patients with more than one primary malignancy; patients aged <18 years; and patients whose survival time and therapeutic information were unknown. In addition, patients with unknown survival months were excluded from the survival analysis. Finally, a total of 844 patients from the SEER database and 94 patients from Southwest Hospital in China were included in the analysis. For the diagnostic model for DM, 844 patients in the SEER cohort were divided randomly into training and internal test sets in a 7:3 ratio with 589 and 255 cases, respectively. 94 patients from the Southwest Hospital cohort were assigned as the external validation set. For the prognostic model for cancer-specific survival (CSS) of PSC patients

with DM, we randomly assigned PSC patients with DM into training and test groups in a 7:3 ratio. As a result, 270 PSC patients with DM were included in the training set, and 117 were included in the validation set. External examination for the prognostic model was not performed for two reasons: first, considerable prognostic information about the PSC patients with DM was censored in our hospital; second, the sample size was too small to satisfy the minimal sample size needed for analysis.

Clinicopathological factors included age, sex, histology, ethnicity, TNM stage, laterality, primary site, marital status, survival time, surgery, chemotherapy, and radiation therapy. Primary T and M staging was adjusted based on tumor size and extension according to the AJCC 8th Edition TNM staging system. Due to the sample limitations, ethnicity was categorized into 'European' and 'other' groups and marital status into 'married' and 'other' groups.

2.2 Statistical analysis

The overall statistical analysis was performed using software (version 4.2.1). Chi-square tests were used to discern differences in categorical variables, and t-tests were used to compare discrepancies between the continuous variables. All variables were included in the correlation analysis using the Spearman method, which was performed to determine which variables were significant and to exclude confounding variables.

2.3 Development and evaluation of ML-based diagnostic models for DM

In the preliminary analysis, variables with $P < 0.05$ in the univariate logistic analysis in the training set were included in the model construction process, which involved the logistic regression (LR), random forest (RF), support vector machine (SVM), extreme gradient boosting (XGB), decision tree (DT), and artificial neural network (ANN) models. The hyperparameters were tuned using 10-fold cross-validation and a grid search procedure to improve the predictive performance and enhance the stability of the ML models. The "tidymodels" package in R software completed all these development procedures.

Model performance was comprehensively evaluated using the area under the curve (AUC), sensitivity, specificity, and accuracy. In addition, we performed calibration curve and decision curve analysis (DCA) to check the discriminative ability and practical clinical value. The best-performing model was then used to build a web-based calculator to allow access to the model.

2.4 Establishment and validation of the prognostic nomogram

In the training cohort, a univariate Cox model was used to identify CSS-related independent factors for PSC patients with DM. Variables with $P < 0.05$ were included in a multivariate Cox analysis. The variables with $P < 0.05$ in the multivariate Cox regression were incorporated into the prognostic nomogram development to estimate survival probability at 3 and 6 months. The AUC, calibration, and DCA estimators were used to evaluate the nomogram.

3 Results

3.1 Baseline characteristics and correlation analysis

A total of 589 patients with PSC were included into the training set while 255 PSC patients were included into in the internal test set. 270 patients in the training set and 117 patients in the internal test set were accompanied with DM. In addition, 94 patients whose clinical information was recorded from the Southwest Hospital were assigned to an external test set, among whom 37 patients suffered from DM. The detailed characteristics and discrepancies between the DM and non-DM groups are presented in Table 1. Patients with DM were observed to be more likely male with advanced T and N stages. In addition, the histology of codes 8031 and 8032, namely giant cell carcinoma and spindle cell carcinoma, respectively, were found to be correlated with a higher proportion of DM. Spearman's correlation analysis suggested that T-stage, N-stage, radiation therapy, and histology were positively correlated with DM, whereas surgery and survival months were negatively correlated. In terms of survival, surgery and chemotherapy were observed to positively influence survival time. In contrast, older age, T stage, N stage, and M stage negatively contributed to survival (Figure 1).

3.2 Establishment and performance of diagnostic ML models for DM

The univariate logistic regression analysis (Supplementary Figure 1), showed that T-stage, N-stage, and histology were variables with P-values < 0.05 and were therefore included in the ML models. In addition, the multivariate analysis (Supplementary Figure 2) showed that the N3 stage, T4 stage, N2 stage, histology of 8031, T3 stage, and histology of 8032 (arranged from high to low according to odds ratios [ORs]) were identified as significant factors contributing to DM. Six ML learning algorithms were established by incorporating the above selection of variables, comprising logistic regression (LR), extreme gradient boosting (XGB), random forest (RF), support vector machine (SVM), decision tree (DT), and artificial neural network (ANN) algorithms. Hyperparameters were fine-tuned by performing 10-cross validation and grid searches. Finally, LR, ANN, and XGB were found to have the highest AUC in the internal test set (Figure 2) while the XGB algorithm outperformed the others with the highest AUC of 0.821 in the external test set. The best hyperparameter metric was eta, 0.007; max_depth, 1L; gamma, 0.011; colsample_bytree, 1; colsample_bynode, 0.231; min_child_weight, 8L; and subsample, 0.997.

10.3389/fonc.2023.1105224

TABLE 1 Baseline characteristics.

Variables	Train set		Internal test set		External test set	
	Non-DM (N = 319)	DM (N = 270)	Non-DM (N = 138)	DM (N = 117)	Non-DM (N = 57)	DM (N = 37)
T stage						
T1	45 (14.1%)	14 (5.2%)	30 (21.7%)	8 (6.8%)	16 (28.1%)	3 (8.1%)
T2	95 (29.8%)	45 (16.7%)	37 (26.8%)	22 (18.8%)	22 (38.6%)	14 (37.8%)
T3	86 (27.0%)	69 (25.6%)	31 (22.5%)	26 (22.2%)	9 (15.8%)	11 (29.7%)
T4	93 (29.2%)	142 (52.6%)	40 (29.0%)	61 (52.1%)	10 (17.5%)	9 (24.3%)
N stage						
N0	202 (63.3%)	86 (31.9%)	83 (60.1%)	41 (35.0%)	40 (70.2%)	9 (24.3%)
N1	42 (13.2%)	31 (11.5%)	22 (15.9%)	14 (12.0%)	7 (12.3%)	1 (2.7%)
N2	67 (21.0%)	109 (40.4%)	31 (22.5%)	43 (36.8%)	10 (17.5%)	14 (37.8%)
N3	8 (2.5%)	44 (16.3%)	2 (1.4%)	19 (16.2%)	0 (0%)	13 (35.1%)
Histology						
8022	155 (48.6%)	75 (27.8%)	81 (58.7%)	25 (21.4%)	14 (24.6%)	4 (10.8%)
8031	57 (17.9%)	96 (35.6%)	24 (17.4%)	39 (33.3%)	9 (15.8%)	9 (24.3%)
8032	90 (28.2%)	95 (35.2%)	29 (21.0%)	50 (42.7%)	20 (35.1%)	15 (40.5%)
8972	1 (0.3%)	0 (0%)	1 (0.7%)	0 (0%)	5 (8.8%)	4 (10.8%)
8980	16 (5.0%)	4 (1.5%)	3 (2.2%)	3 (2.6%)	9 (15.8%)	5 (13.5%)
Sex						
Female	135 (42.3%)	101 (37.4%)	60 (43.5%)	44 (37.6%)	9 (15.8%)	9 (24.3%)
Male	184 (57.7%)	169 (62.6%)	78 (56.5%)	73 (62.4%)	48 (84.2%)	28 (75.7%)
Laterality						
Center	136 (42.6%)	116 (43.0%)	54 (39.1%)	57 (48.7%)	31 (54.4%)	14 (37.8%)
Right	183 (57.4%)	154 (57.0%)	84 (60.9%)	60 (51.3%)	26 (45.6%)	23 (62.2%)
Primary site						
Lower	95 (29.8%)	81 (30.0%)	29 (21.0%)	40 (34.2%)	18 (31.6%)	14 (37.8%)
Middle	14 (4.4%)	18 (6.7%)	10 (7.2%)	5 (4.3%)	1 (1.8%)	1 (2.7%)
Other	20 (6.3%)	15 (5.6%)	4 (2.9%)	7 (6.0%)	/	/
Upper	190 (59.6%)	156 (57.8%)	95 (68.8%)	65 (55.6%)	38 (66.7%)	22 (59.5%)
Ethnicity						
Other	23 (7.2%)	23 (8.5%)	12 (8.7%)	9 (7.7%)	57 (100%)	37 (100%)
European	296 (92.8%)	247 (91.5%)	126 (91.3%)	108 (92.3%)	/	/
Age						
Mean (SD)	67.4 (11.5)	67.4 (11.4)	66.7 (10.6)	64.5 (12.0)	63.2 (11.8)	65.4 (11.0)
Marital status						
Married	174 (54.5%)	138 (51.1%)	71 (51.4%)	64 (54.7%)	56 (98.2%)	37 (100%)
Other	145 (45.5%)	132 (48.9%)	67 (48.6%)	53 (45.3%)	1 (1.8%)	0 (0%)

DM, distant metastasis; SD, standard deviation.

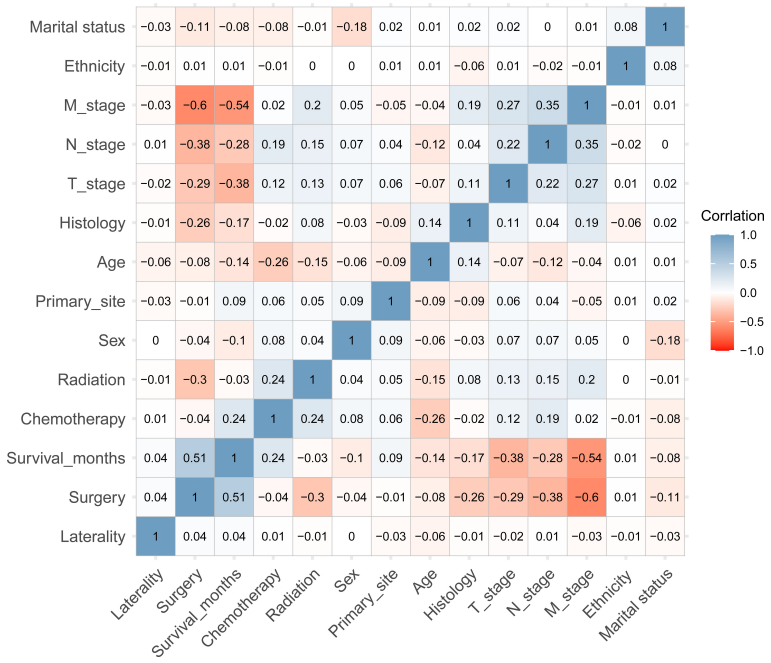


FIGURE 1
The heatmap of Spearman's correlation analysis of the baseline characteristics. The correlation index ranges from -1.0 to 1.0, with a brighter color indicating a stronger correlation.

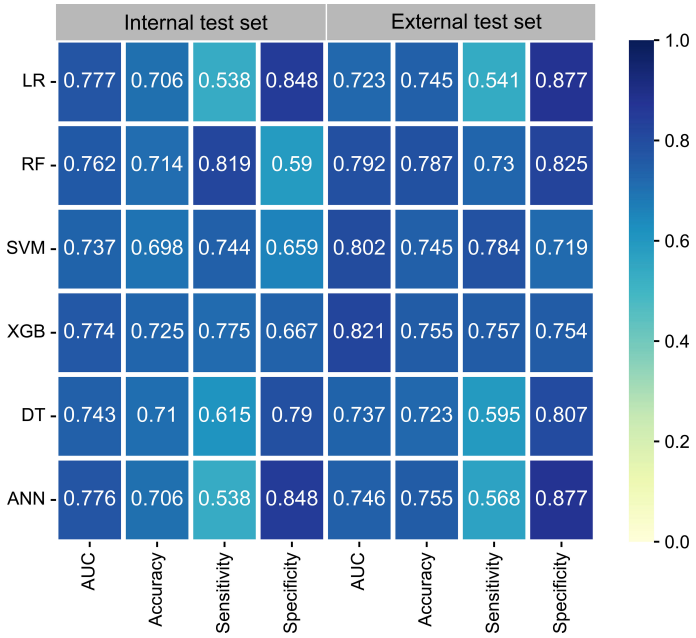


FIGURE 2
The performance of six ML models in terms of AUC, accuracy, sensitivity, and specificity. ANN, artificial neural network; AUC, area under the curve; DT, decision tree; LR, logistic regression; ML, machine learning; RF, random forest; SVM, support vector machine; XGB, extreme gradient boosting.

As illustrated in Figures 3A–C, the AUC differed slightly in the training set and the internal test set among the six ML algorithms; however, in the external test set, XGB performed the best. Calibration plots (Figures 3D–F) showed that ML algorithms had a good fitting ability, except for the ANN model, which meant that the ML algorithms accurately predicted the outcome. DCA curves (Figures 3G–I) suggested that ML algorithms had a high clinical application value and could effectively help diagnose DM.

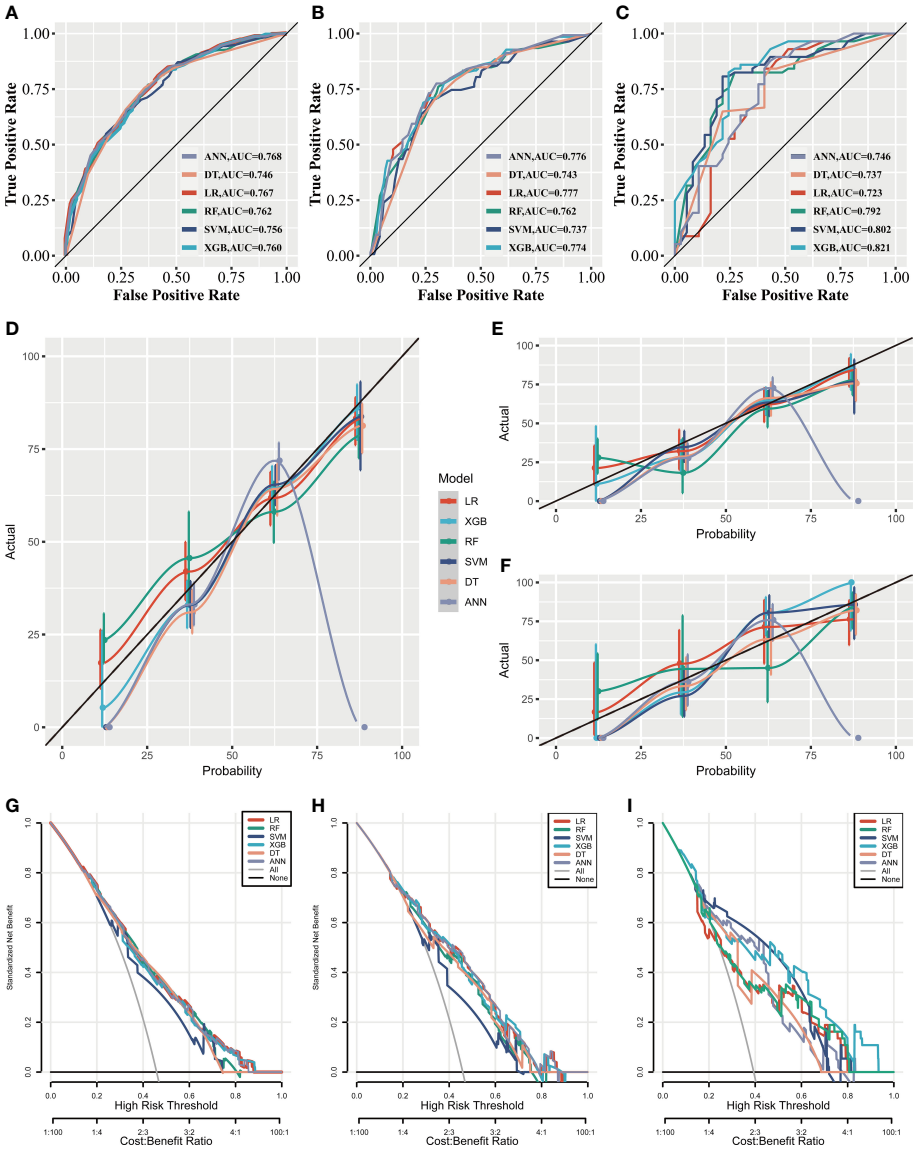


FIGURE 3
The clinical applicative performance of six ML models. Receiver operating characteristic curves of six ML models in the training set **(A)**, the internal test set **(B)**, and the external test set **(C)**. Calibration curves of six ML models in the training set **(D)**, the internal test set **(E)** and the external test set **(F)**. Decision curve analysis of six ML models in the training set **(G)**, the internal test set **(H)**, and the external test set **(I)**.

3.3 Web-based predicator publication

An online calculator based on the XGB model was successfully built (Figure 4), which performed the best among the other models. This web-based tool can be readily accessed at https://onepiece.shinyapps.io/PSCdistant/. This tool encompassed three simple clinical variables and could help clinicians accurately and conveniently identify those at risk for DM among patients with PSC.

3.4 Prognostic nomogram establishment and performance

In the prognostic analysis, 270 PSC patients with DM in the training and 117 patients in the test set were enrolled. As shown in Supplementary Figure 3, the univariate Cox regression analysis indicated that the T4 stage (P = 0.013), N2 stage (P = 0.048), N3 stage (P = 0.001), upper site (P = 0.023), surgery (P = 0.013), chemotherapy (P < 0.001), and radiation therapy (P < 0.001) were significantly associated with prognosis in PSC patients with DM. After performing multivariate Cox analysis (Supplementary Figure 4), stage T4 (P = 0.005), T3 (P = 0.034), N3 (P = 0.002), N2 (P = 0.005), and upper site (P = 0.01) were found to be independent adverse prognostic factors, whereas radiation therapy (P = 0.004), surgery (P = 0.004), and chemotherapy (P < 0.001) were protective prognostic factors for PSC patients with DM. We developed a prognostic nomogram based on these independent variables, to predict the CSS-related survival probability at 3 and 6 months (Figure 5). The CSS-related AUC values at 3 and 6 months were 0.810 and 0.822 in the training set (Figure 6A) and 0.803 and

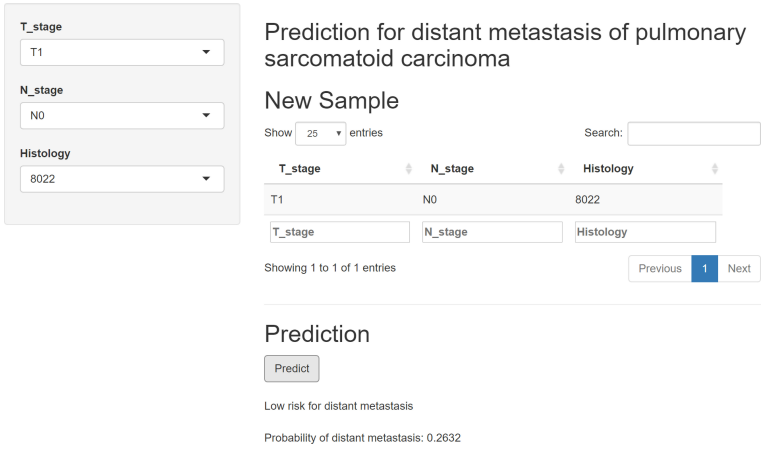


FIGURE 4
Web-based calculator online for free, based on the XGB model. XGB, extreme gradient boosting.

0.869 in the test set (Figure 6B), respectively, suggesting high predictive accuracy. In addition, calibration (Figures 6C–F) and DCA curves (Figures 6G, H) showed the prognostic nomogram's good fitting ability and clinical application.

4 Discussion

PSC presents a rare lung malignancy with a high probability for DM. Several studies having reported that DM is an independent risk factor for PSC (6, 17, 18). In this study, approximately 45.2% patients had DM at initial presentation, similar to that reported by Zombori-Tóth and Terra (19, 20). The high rate of DM may be attributed to its capacity to invade into vasculature and the high likelihood of DM recurrence. Lococo et al. analyzed the pathological results of 143 patients and found the high incidences of lymphatic (30%) and vascular emboli (68%) (21). Liang et al. found that MET mutations were commonly expressed in PSCs with a high rate of 16% (22), whereas the incidence of KRAS mutations was 22%. Similarly, Liu et al. have identified that eight out of 36 (22%) patients harboring MET mutations (23). This incidence was considerably more frequent than that in other NSCLC subtypes (3%) (24). A study on 77 Chinese patients with PSC indicated that positive METex14 skipping mutations rate was 20.8%, leading to worse disease-free survival (DFS) (25). In recent years, immune checkpoint inhibitors (ICIs) and targeted drugs for MET exon 14 skipping mutations have exhibited satisfying benefits in improving patient survival (6, 26, 27). However, patients who developed DM still had a significantly worse prognosis. Thus, early attention to DM is important to improve prognosis and help clinicians make the most optional targeted therapeutic decision.

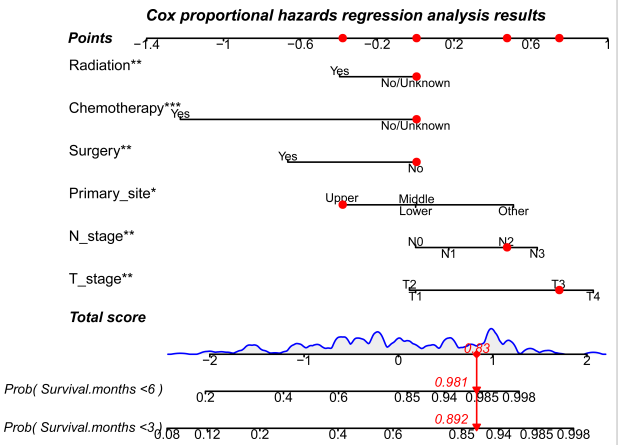


FIGURE 5
A prognostic nomogram to evaluate 3- and 6-month CSS-related survival probabilities for PSC patients with DM. Note: Drawing a vertical line from each line of the parameters to the "points" axis can be used to acquire the points for each variable. Then, the total score can be obtained by adding all the points for each independent variable. Finally, by drawing a vertical line from the "total score" axis to the survival probability line, the individual survival probability at 3 and 6 months can be calculated, and the risk level can be obtained. DM, distant metastases; PSC, pulmonary sarcomatoid carcinoma. *P<0.05; **P<0.01; ***P<0.001.

CT, MRI, and PET-CT are conventional radiological screening methods for patients with NSCLC. However, performing all these radiological examinations is costly for newly diagnosed patients. In addition, time-consuming screening processes and potential side effects can delay the patients' course from diagnosis to therapy. Therefore, there is an urgent need for a noninvasive, precise, and rapid method to assist in identifying potential DM at the initial hospitalization stage and to estimate the prognosis for PSC patients with DM. In this study, six ML algorithms for predicting DM in patients with PSC and a nomogram for evaluating the prognosis for those with DM were developed. An automatic calculator based on the best-performing algorithm was created and published online for public access. Moreover, the prognostic nomogram could accurately identify risk factors in PSC patients with DM and help clinicians evaluate survival.

Few studies have identified the risk factors for DM in patients with PSC due to its rarity. Logistic regression analysis was performed and found that histologic subtype, T stage, and N

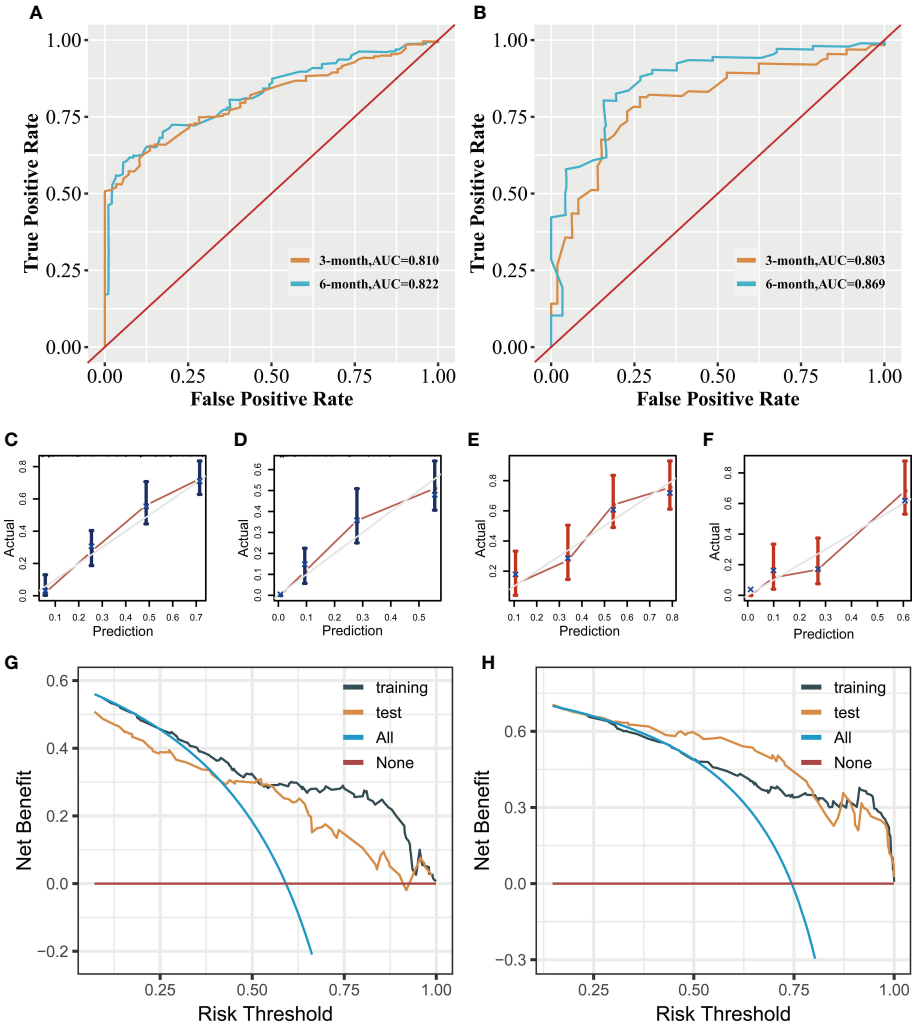


FIGURE 6
Clinical applicative performance of the prognostic nomogram. Receiver operating characteristic curves at 3 and 6 months in training **(A)** and test sets **(B)**. Calibration plots at 3 months **(C)** and 6 months **(D)** in the training set. Calibration plots at 3 months **(E)** and 6 months **(F)** in the test set. Decision clinical analysis at 3 and 6 months in the training **(G)** and test sets **(H)**.

stage were independent risk factors for DM in patients with PSC. Among these factors, the N3 stage was most correlated with DM. Patients with the N3 stage had the highest OR of 9.72 ($P < 0.001$), followed by the T4 stage (OR 4.30, $P < 0.001$), N2 stage (OR 3.09, $P < 0.001$), giant cell carcinoma (OR 2.56, $P < 0.001$), T3 stage (OR 2.36, $P = 0.022$), and spindle cell carcinoma (OR 2.12, $P < 0.001$). These findings were similar to those reported by Xiao et al., who retrospectively analyzed 934 PSC patients using the SEER program database and reported that spindle cell carcinoma (OR 3.151, $P < 0.001$) and giant cell carcinoma (OR 4.023, $P < 0.001$) were independent risk factors for DM (13).

Similarly, T-stage and N-stage have been reported as risk factors for DM in lung adenocarcinoma and squamous cell carcinoma (28, 29). The two variables have also been identified as important factors for the development of DM in patients with PSC (28, 29). These three clinical features were then incorporated into the building procedures of the six ML models. The XGB model performed better than the other algorithms, with an AUC of 0.821, showing excellent predictive ability for DM in patients with PSC. In addition, calibration and DCA curves indicated that the model was highly consistent with actual observations and showed better clinical applicability (30).

Carcinoembryonic antigen levels, cytokeratin 19 fragment (CYFRA21-1), and other serologic indicators which were associated with DM in patients with NSCLC (31, 32) have not been included in this study due to missing data in the SEER program. However, the relatively high accuracy of the XGB algorithm still helped identify those PSC patients with a high risk for DM. To facilitate the use of the ML predictor, a web-based calculator was developed, which is likely to help clinicians detect DM early in patients with PSC in a convenient manner.

Compared to other subtypes of NSCLC, PSC is less sensitive to chemoradiotherapy. However, studies have confirmed that chemoradiotherapy significantly prolonged the CSS of PSC patients with DM. Xiao et al. analyzed the prognostic risk factors of 512 patients with metastatic PSC diagnosed between 1975 and

2016 in the SEER program (13). They found that having received chemotherapy (HR 0.308, P < 0.001) was an independent prognostic risk factor for PSC patients with DM, which was similar to our study findings (HR 0.29, P < 0.001). However, they did not identify radiation therapy and surgical resection as independent prognostic factors for PSC patients with DM, while these two therapies were indicated to prolong CSS significantly in our study. We speculate that this discrepancy was due to diagnosis time of the enrolled patients were prior to the year 2000 while surgical and radiation techniques were not so effective. Our study identified T-stage and N-stage as important prognostic risk factors for PSC patients with DM. Advanced T stage (T4, HR 2.52, P = 0.005; T3, HR 2.10, P = 0.03) and N-stage (N3, HR 1.88, P = 0.002; N2, HR 1.60, P = 0.005) correlated with poor CSS in PSC patients suffered from DM.

Interestingly, we have also found that the nodule site appeared to affect the prognosis for PSC patients with DM. CSS was longer in patients whose primary site was located on an upper rather than a lower lobe (HR, 0.68; P = 0.011). The reason may be that the nodules on the lower lung will develop more DM lesions leading a worse prognosis (13). Notably, we established a nomogram for predicting the prognosis for PSC patients with DM, with a relative higher AUC at 3 months and 6 months respectively. Its high consistency to actual observations indicates that this nomogram could precisely predict CSS in PSC patients with DM.

There are some limitations in our present study. First, this was a retrospective study, and selective bias could not be avoided. Second, characteristic information about DM was collected during initial hospitalization, which may have led to an underestimated percentage of DM in patients with PSC. Lastly, the predictive model was externally validated using patients' information from Southwest Hospital in China. However, since this hospital mainly treats Chinese patients, the model's application value should be further validated in cohorts involving differing ethnicities.

5 Conclusion

Our study indicated that PSC patients with advanced T-stage, N-stage, histology of giant cell carcinoma and spindle cell carcinoma were risk factors for DM and should receive more attention in terms of preventative therapeutic strategies. The AUC, accuracy, sensitivity, and specificity of the XGB model reached 0.821, 0.755, 0.757, and 0.754, respectively. A diagnostic model for DM based on the ML algorithm and a web-based predictor was then established, which could conveniently and precisely predict the risk of DM in PSC patients. Our study provided initial predictive and prognostic models for PSC patients with DM. Future studies may focus on further improving the models by adding other potential variables and developing more detailed models to predict the risk and prognosis for specific metastatic sites.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Ethics statement

The studies involving human participants were reviewed and approved by Southwest Hospital, Third Military Medicine University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

XY and WX designed the study. GT, LZ, and KW collected the primary data. XY drafted the manuscript. HL and XZ reviewed and edited the paper. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by Chongqing Medical Scientific Research Project (No. 2020FYYX012 and No. 2020jstg016).

Acknowledgments

We sincerely acknowledge all authors who contributed to establishing effective R packages and to Editors who helped to check for and correct grammatical errors in this work.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1105224/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The Forest plot of the univariate logistic regression analysis. OR, odds ratio.

SUPPLEMENTARY FIGURE 2
Forest plot of multivariate logistic regression analysis. The OR value decreases from top to bottom.

SUPPLEMENTARY FIGURE 3
The Forest plot of the univariate Cox proportional hazards regression analysis. HR, hazard ratio.

SUPPLEMENTARY FIGURE 4
The Forest plot of the multivariate Cox proportional hazards regression analysis. The HR value is in descending order from top to bottom.

References

1. Travis WD, Brambilla E, Burke AP, Marx A, Nicholson AG. Introduction to the 2015 world health organization classification of tumors of the lung, pleura, thymus, and heart. *J Thorac Oncol* (2015) 10(9):1240–2. doi: 10.1097/JTO.000-0000000000663

2. Li Y, Zhang L, Jiang J, Yang H, Cao L, Gu Q, et al. Clinical characteristics and prognostic analysis of 38 patients with pulmonary sarcomatoid carcinoma. *Chin J Lung Cancer* (2015) 18(9):537–42. doi: 10.3779/j.issn.1009-3419.2015.09.02

3. Ung M, Rouquette I, Filleron T, Taillandy K, Brouchet L, Bennouna J, et al. Characteristics and clinical outcomes of sarcomatoid carcinoma of the lung. *Clin Lung Cancer* (2016) 17(5):391–7. doi: 10.1016/j.cllc.2016.03.001

4. Sun L, Dai J, Chen Y, Duan L, He W, Chen Q, et al. Pulmonary sarcomatoid carcinoma: experience from SEER database and shanghai pulmonary hospital. *Ann Thorac Surg* (2020) 110(2):406–13. doi: 10.1016/j.athoracsur.2020.02.071

5. Vieira T, Antoine M, Ruppert A-M, Fallet V, Duruisseaux M, Giroux Leprieur E, et al. Blood vessel invasion is a major feature and a factor of poor prognosis in sarcomatoid carcinoma of the lung. *Lung Cancer* (2014) 85:276–81. doi: 10.1016/j.lungcan.2014.06.004

6. Liang L, Liu Z, Wang C, Xie S. Adjuvant chemotherapy is not a decisive factor in improving the overall survival of pulmonary sarcoma: a population-based study. *Front Oncol* (2022) 12:940839. doi: 10.3389/fonc.2022.940839

7. Zheng Y, Fu Y, Zhong Q, Deng R, Zhang Y. The treatment of advanced pulmonary sarcomatoid carcinoma. *Future Oncol* (2022) 18:727–38. doi: 10.2217/fon-2021-1071

8. Gong C, Xiong H, Qin K, Wang J, Cheng Y, Zhao J, et al. MET alterations in advanced pulmonary sarcomatoid carcinoma. *Front Oncol* (2022). doi: 10.3389/fonc.2022.1017026

9. Park JS, Lee Y, Han J, Kim HK, Choi YS, Kim J, et al. Clinicopathologic outcomes of curative resection for sarcomatoid carcinoma of the lung. *Oncology* (2011) 81:206–13. doi: 10.1159/000333095

10. Weissferdt A. Pulmonary sarcomatoid carcinomas: a review. *Adv Anat Pathol* (2018) 25:304–13. doi: 10.1097/PAP.0000000000000202

11. Yendamuri S, Caty L, Pine M, Adem S, Bogner P, Miller A, et al. Outcomes of sarcomatoid carcinoma of the lung: a surveillance, epidemiology, and end results database analysis. *Surgery* (2012) 152:397–402. doi: 10.1016/j.surg.2012.05.007

12. Smadhi H, Boudaya MS, Abdannadher M, BenAbdelghaffar H, Kamoun H, Ayadi A, et al. Pulmonary sarcomatoid carcinoma: a surgical diagnosis and prognostic factors. *Tunis Med* (2019) 97:128–32.

13. Xiao C, Yang X, Hao J, Guo C, Pu Q, Liu L. Clinicopathological features and prognostic analysis of metastatic pulmonary sarcomatoid carcinoma: a SEER analysis. *J Thorac Dis* (2021) 13:893–905. doi: 10.21037/jtd-20-2826

14. Du M, Haag DG, Lynch JW, Mittinty MN. Comparison of the tree-based machine learning algorithms to cox regression in predicting the survival of oral and pharyngeal cancers: analyses based on SEER database. *Cancers (Basel)* (2020) 12:2802. doi: 10.3390/cancers12102802

15. Lee C, Light A, Alaa A, Thurtle D, van der Schaar M, Gnanapragasam VJ. Application of a novel machine learning framework for predicting non-metastatic prostate cancer-specific mortality in men using the surveillance, epidemiology, and end results (SEER) database. *Lancet Digit Health* (2021) 3:e158–65. doi: 10.1016/S2589-7500(20)30314-9

16. Li T, Huang H, Zhang S, Zhang Y, Jing H, Sun T, et al. Predictive models based on machine learning for bone metastasis in patients with diagnosed colorectal cancer. *Front Public Health* (2022) 10:984750. doi: 10.3389/fpubh.2022.984750

17. Gu L, Xu Y, Chen Z, Pan Y, Lu S. Clinical analysis of 95 cases of pulmonary sarcomatoid carcinoma. *BioMed Pharmacother* (2015) 76:134–40. doi: 10.1016/j.biopha.2015.10.009

18. Huang S-Y, Shen S-J, Li X-Y. Pulmonary sarcomatoid carcinoma: a clinicopathologic study and prognostic analysis of 51 cases. *World J Surg Oncol* (2013) 11:252. doi: 10.1186/1477-7819-11-252

19. Terra SBSP, Aubry MC, Yi ES, Boland JM. Immunohistochemical study of 36 cases of pulmonary sarcomatoid carcinoma–sensitivity of TTF-1 is superior to napsin. *Hum Pathol* (2014) 45:294–302. doi: 10.1016/j.humpath.2013.09.005

20. Zombori-Tóth N, Kiss S, Oštarijaš E, Alizadeh H, Zombori T. Adjuvant chemotherapy could improve the survival of pulmonary sarcomatoid carcinoma: a systematic review and meta-analysis. *Surg Oncol* (2022) 44:101824. doi: 10.1016/j.suronc.2022.101824

21. Lococo F, Rapicetta C, Cardillo G, Stefani A, Margaritora S, Leuzzi G, et al. Pathologic findings and long-term results after surgical treatment for pulmonary sarcomatoid tumors: a multicenter analysis. *Ann Thorac Surg* (2017) 103:1142–50. doi: 10.1016/j.athoracsur.2016.08.114

22. Liang X, Li Q, Xu B, Hu S, Wang Q, Li Y, et al. Mutation landscape and tumor mutation burden analysis of Chinese patients with pulmonary sarcomatoid carcinomas. *Int J Clin Oncol* (2019) 24:1061–8. doi: 10.1007/s10147-019-01454-6

23. Liu X, Jia Y, Stoopler MB, Shen Y, Cheng H, Chen J, et al. Next-generation sequencing of pulmonary sarcomatoid carcinoma reveals high frequency of actionable MET gene mutations. *J Clin Oncol* (2016) 34:794–802. doi: 10.1200/JCO.2015.62.0674

24. Awad MM, Oxnard GR, Jackman DM, Savukoski DO, Hall D, Shivdasani P, et al. MET exon 14 mutations in non-Small-Cell lung cancer are associated with advanced age and stage-dependent MET genomic amplification and c-met overexpression. *J Clin Oncol* (2016) 34:721–30. doi: 10.1200/JCO.2015.63.4600

25. Li Y, Gao L, Ma D, Qiu T, Li W, Li W, et al. Identification of MET exon14 skipping by targeted DNA- and RNA-based next-generation sequencing in pulmonary sarcomatoid carcinomas. *Lung Cancer* (2018) 122:113–9. doi: 10.1016/j.lungcan.2018.06.001

26. Drusbosky LM, Dawar R, Rodriguez E, Ikpeazu CV. Therapeutic strategies in METex14 skipping mutated non-small cell lung cancer. *J Hematol Oncol* (2021) 14:129. doi: 10.1186/s13045-021-01138-7

27. Lu S, Fang J, Li X, Cao L, Zhou J, Guo Q, et al. Once-daily savolitinib in Chinese patients with pulmonary sarcomatoid carcinomas and other non-small-cell lung cancers harbouring MET exon 14 skipping alterations: a multicentre, single-arm, open-label, phase 2 study. *Lancet Respir Med* (2021) 9:1154–64. doi: 10.1016/S2213-2600(21)00084-9

28. Cai H, Wang H, Li Z, Lin J, Yu J. The prognostic analysis of different metastatic patterns in extensive-stage small-cell lung cancer patients: a large population-based study. *Future Oncol* (2018) 14:1397–407. doi: 10.2217/fon-2017-0706

29. Gao H, He Z-Y, Du X-L, Wang Z-G, Xiang L. Machine learning for the prediction of synchronous organ-specific metastasis in patients with lung cancer. *Front Oncol* (2022) 12:817372. doi: 10.3389/fonc.2022.817372

30. Chen M, Yang Q, Xu Z, Luo B, Li F, Yu Y, et al. Survival analysis and prediction model for pulmonary sarcomatoid carcinoma based on SEER database. *Front Oncol* (2021) 11:630885. doi: 10.3389/fonc.2021.630885

31. Cacho-Díaz B, Cuapaténcatl LD, Rodríguez JA, Garcilazo-Reyes YJ, Reynoso-Noverón N, Arrieta O. Identification of a high-risk group for brain metastases in non-small cell lung cancer patients. *J Neurooncol* (2021) 155:101–6. doi: 10.1007/s11060-021-03849-w

32. Park S, Lee SM, Ahn Y, Kim M, Suh CH, Do K-H, et al. Identification of predictors for brain metastasis in newly diagnosed non-small cell lung cancer: a single-center cohort study. *Eur Radiol* (2022) 32:990–1001. doi: 10.1007/s00330-021-08215-y

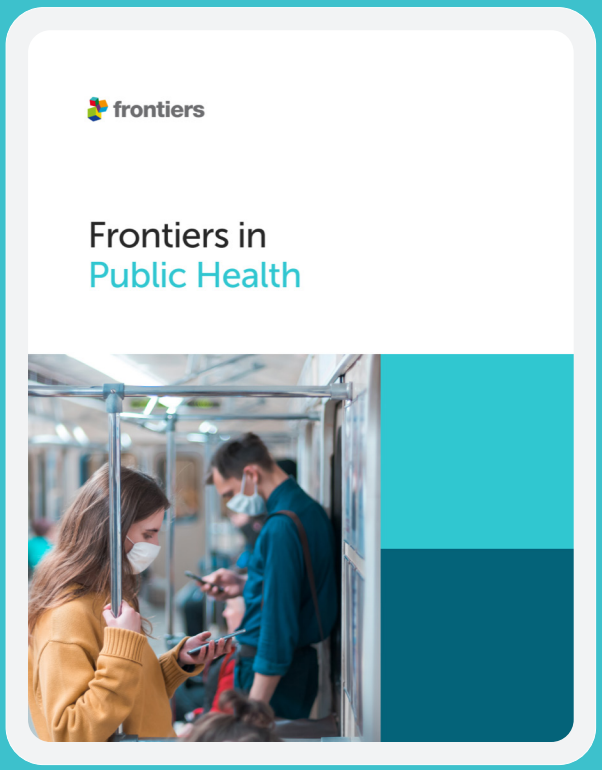
Frontiers in Public Health

Explores and addresses today's fast-moving healthcare challenges

One of the most cited journals in its field, which promotes discussion around inter-sectoral public health challenges spanning health promotion to climate change, transportation, environmental change and even species diversity.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact