IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 13, No. 4, December 2024, pp. 3772~3785
ISSN: 2252-8938, DOI: 10.11591/ijai.v13.i4.pp3772-3785

Impact of federated learning and explainable artificial intelligence for medical image diagnosis

Sivakumar Muthuramalingam[1], Padmapriya Thiyagarajan[2]
[1]Department of Computer Science and Engineering, Thiagarajar College of Engineering, Madurai, India
[2]Department of Applied Mathematics and Computational Science, Thiagarajar College of Engineering, Madurai, India

Article Info

Article history:

Received Jan 16, 2024
Revised Mar 27, 2024
Accepted Apr 18, 2024

Keywords:

Artificial intelligence
Data ethics
Explainable artificial intelligence
Federated learning
Machine learning
Medical image diagnostics
Medical image segmentation

ABSTRACT

Medical image recognition has enormous potential to benefit from the recent developments in federated learning (FL) and interpretable artificial intelligence (AI). The function of FL and explainable artificial intelligence (XAI) in the diagnosis of brain cancers is discussed in this paper. XAI and FL techniques are vital for ensuring data ethics during medical image processing. This paper highlights the benefits of FL, such as cooperative model training and data privacy preservation, and the significance of XAI approaches in providing logical justifications for model predictions. A number of case studies on the segmentation of medical images employing FL were reviewed to compares and contrasts various methods for assessing the efficacy of FL and XAI based diagnostic models for brain tumors. The relevance of FL and XAI to improve the accuracy and interpretability during medical image diagnosis have been presented. Future research directions are also described indicating as to integrate data from various modes, create standardised evaluation processes, and manage ethical issues. This paper is intended to provide a deeper insight on relevance of FL and XAI in medical image diagnosis.

Corresponding Author:

Sivakumar Muthuramalingam
Department of Computer Science and Engineering, Thiagarajar College of Engineering
Thirupparankundram, Madurai, India
Email: mskcse@tce.edu

1. INTRODUCTION

Analysis of medical images has far-reaching implications for many facets of modern patient care and professional practice. The diagnosis and identification of illness is one of its primary functions. This allows them to detect abnormalities and diseases when they are still treatable. Preventing diseases from progressing to more advanced stages not only increases the likelihood of successful treatment but also saves on healthcare costs [1]. Medical image interpretation is essential for treatment planning. Radiation oncologists and surgeons rely on accurate imaging to ensure procedures and treatments target affected areas while causing as little damage to healthy tissue as possible. This accuracy is critical for demanding surgeries and therapies such as brain surgery, where even small errors can have devastating consequences.

Analysis of medical images is crucial for monitoring disease progression. By comparing images collected at different points in time, doctors can study how diseases manifest and how effective treatments are. Better patient outcomes come from the ability to quickly change treatment approaches [2], [3]. Medical image analysis supports personalized healthcare. Because each patient is different and has unique anatomical and disease-specific characteristics, this is taken into account. By tailoring medication to specific patients based on their individual imaging data, doctors can increase the effectiveness of treatments while avoiding side effects or effects.

Medical complications include seizures, brain injuries, and increased intracranial pressure. Early detection and treatment can reduce the likelihood or even eliminate these side effects. Successful treatment and long-term survival are more likely when brain cancers are detected early [4], [5]. In some situations, early detection can even enable complete treatment of the cancer. A quick diagnosis of a medical picture can ease anxiety and confusion about the cause of symptoms. This allows patients to understand the situation better and make the best decision. Preserving neurological function and delaying or preventing the onset of symptoms including seizures, vision or hearing loss, and cognitive deficiencies, are possible because to early detection and treatment of disease [6].

Conventional image processing algorithms used to diagnose medical images have significant drawbacks that may affect their utility and accuracy. Because so many traditional image processing methods rely on human interpretation, the analysis may be vulnerable to subjectivity and variability [7]. It is possible for various radiologists or medical specialists to interpret the images differently, which might result in variations in diagnoses and treatment suggestions [8]. Although imaging tests are useful for finding and identifying medical images, image processing algorithms may not always be reliable at differentiating between different tumor types or finding small tumors that might be overlooked on imaging. Very small tumors, especially those that are slow-growing, may not be detectable by some image processing systems [9]. As a result, there may be tumors present but not recognised by the algorithm, producing false-negative results. Some image processing techniques might not be sufficiently precise to classify tumors as benign or malignant or to determine the tumour’s severity [10]. Motion artefacts, for example, might degrade the image quality and make it more challenging for image processing algorithms to correctly identify a medical image. The image processing methods used to diagnose medical images are not standardised, which can cause variation in the outcomes of various algorithms. The use of image processing algorithms to track a medical image's development over time or identify changes in the tumor's characteristics may not always be successful.

Although deep learning algorithms may improve the precision with which brain tumours are diagnosed, they are not without substantial downsides. In order to train and enhance their models, deep learning algorithms need access to massive amounts of data [11]. Nevertheless, getting such information might be difficult, especially for uncommon kinds of medical images. Black box models are sometimes used to describe deep learning algorithms because it's hard to figure out how they work within them. This makes it hard for doctors to figure out why the algorithm offers a certain treatment or diagnosis [12]. Deep learning algorithms could lead to incorrect diagnosis and inappropriate recommendations for therapy. If deep learning algorithms were only ever taught on data from a single institution or demographic, they might not be applicable to data from any other. This reduces their potential usefulness in healthcare settings. The training and optimization of a deep learning model can be computationally intensive, necessitating dedicated hardware and software. As a result, they may become more difficult and costly to implement in clinical settings [13].

Machine learning (ML) has the potential to enhance the efficiency and accuracy of diagnosing brain tumors. ML algorithms can detect subtle changes in brain imaging that are undetectable to the naked eye [14]. This could improve the precision of tumor characterization and diagnosis. Imaging and clinical data can be used with ML algorithms to help diagnose patients, make treatment decisions, and predict the type and grade of tumors [15]. The efficacy of treatment can be evaluated, and tumor growth can be tracked, with the use of ML algorithms applied to analyzed image data. This may allow doctors to make smarter decisions and improve their treatments. Invasive procedures such as brain biopsies can be avoided by using non-invasive diagnostic technologies such as magnetic resonance imaging (MRI) and computed tomography (CT) scans, which are more accurate [16].

Explainable artificial intelligence (XAI) is important to data ethics for several reasons. The ethical use of data is the subject of data ethics. It includes a set of principles and rules to ensure ethical data handling. However, XAI research focuses on creating artificial intelligence (AI) systems that can explain their decisions and actions. To ensure openness, accountability, and trustworthiness in AI decision-making, explainability must be integrated. XAI addresses ethical issues about biases, prejudice, and lack of interpretability in opaque AI models by giving accessible explanations for AI generated decisions. XAI helps AI systems become transparent by revealing a model's decision-making process. Transparency in AI is crucial for monitoring and assessing developers and companies. Transparency helps identify and attribute unethical or biased AI outputs to the guilty parties. Ethical issues in data and AI stem from bias and discrimination in algorithms and datasets.

XAI could help identify and reduce prejudice. XAI helps explain biased outcomes by revealing the elements and data points that influence AI decision-making. This enhanced exposure makes it easier to identify and correct injustice and discrimination. XAI may help promote fairness and equity in AI systems by illuminating bias-prone decision-making processes. Informed consent from data subjects is necessary for ethical data use. If AI systems can explain their decision-making processes, people are more likely to trust and agree to AI applications using their data. Certain data protection legislation, such as the EU's general data protection regulation (GDPR), allow individuals to be informed about automated decisions that affect them. XAI is crucial to meeting the legal requirement to provide clear and understandable reasons for AI system judgements. Ethical AI aims to reduce discrimination and negative effects. XAI can help identify AI systems making decisions based on irrelevant or discriminatory criteria. This feature allows businesses to prevent and fix such issues.

Federated learning's (FL) privacy-preserving framework for ML is crucial to data ethics. This strategy solves many ethical issues with centralised data gathering and processing. FL is important in data ethics for various reasons. First, it lets data contributors train ML models collaboratively while protecting their anonymity. This distributed strategy keeps sensitive data local, reducing the hazards of centralising massive databases. FL also increases fairness and inclusivity by using varied datasets from different sources for model training. This reduces biases from using a single, centralised dataset. ML traditionally uses centralised models, which require the storage of sensitive data. However, this practise may compromise privacy. FL is an innovative method for training ML models with user-owned data. This strategy reduces data exchange, addressing privacy issues and hazards. FL respects data ownership and governance. In a FL system, individuals and organisations own and govern their data. This decentralised approach lets people actively train models while maintaining data control. This is congruent with ethical ideals that value data autonomy and control.

Ethical data practises emphasise data minimization, or restricting data gathering and processing to what is necessary. FL uses decentralised data sources to update models without sharing raw data. FL emphasises informed consent and active engagement in ML model training. This observation follows ethical norms that value informed consent and data transparency. FL can improve data security and confidentiality by restricting data exposure. Data is stored on the user's device, preventing data breaches and unauthorised access. This technique efficiently addresses ethical issues related to data security and confidentiality. FL's capacity to use many data sources during model training might reduce bias and discrimination in AI models. FL's distributed nature allows the training process to reflect data source variety, minimising bias and prejudice in AI models. This technique may help overcome biased AI system ethics and promote justice and inclusion in AI applications. This strategy promotes justice and resolves bias-related AI ethical issues. FL follows data ethics principles like privacy, data ownership, permission, fairness, and data minimization. The approach being described is interesting for ML since it addresses ethical issues connected with typical data practises and centralised processing.

In section 2, we brief as how FL can help diagnose malignant brain tumors. In section 3, we outline the role of XAI in the diagnosis of brain tumors. Section 4 discusses the existing methods for diagnosing medical images using FL and XAI. Section 5 presents the evaluation of medical image diagnosis methods using FL and XAI. Section 6 discusses case studies of medical image analysis using XAI and FL. In section 7, we discuss about the results and possible directions for future study.

2. FEDERATED LEARNING FOR MEDICAL IMAGE DIAGNOSIS-SOME REFLECTIONS

ML method FL allows for collective model training on dispersed data. FL allows training models directly on data spread across numerous devices or locations while keeping the data local and ensuring privacy, unlike typical ML methods [17]. FL disperses the data so that it can be used in the model-training process, rather than having it all kept in one central location. It's important to keep this in mind if data sharing is restricted for any reason, whether it be due to privacy concerns, security concerns, or legal requirements. FL provides a workaround by enabling edge nodes to collaborate and collaboratively train a shared model without releasing their unique data [18]. Figure 1 depicts the implementation of FL in medical image analysis.

In FL, a central server sets up and communicates the model framework to all client devices. The distributed model is trained using local data collected from each device. Training can make use of deep neural networks and other ML techniques [19]. Devices update the central server with their refined model parameters or gradients after completing in-house training. Using methods like averaging and weighted averaging, the server collects model updates from numerous devices and merges them into a unified model. Iterative processes are used for both regional training and model aggregation [20].

FL's ability to maintain user anonymity is a major selling point. Local data is stored on each device, and only model updates are exchanged. This distributed approach helps guarantee personal data is kept hidden [21]. The trained model can then be deployed for inference on additional data, either on the central server or on the participating devices, after the FL process reaches the required level of model performance. The healthcare industry, where patient confidentiality and data security are of the utmost importance, is just one area where FL has been widely adopted and successfully put to use. While protecting patients' anonymity, it facilitates model-training collaboration between healthcare facilities [22]. Medical image analysis, disease prediction, and clinical decision assistance are just a few examples of successful applications of FL. However, FL also presents difficulties such as inefficient communication, data heterogeneity between devices, and dealing with data that is not independently and identically distributed. These issues are the topic of ongoing research aimed at making FL even more versatile and efficient [23].

Medical imaging is an area where FL presents both opportunities and obstacles. Protecting users' anonymity is a major perk. FL paves the way for cooperation across hospitals without jeopardizing patients' right to privacy. FL ensures compliance with data protection standards [24] by storing data locally and only communicating model updates. Additionally, FL allows access to diverse and large-scale data by aggregating information from multiple sources. This leads to more robust and generalizable models. Another advantage is the reduction in data transfer and storage requirements. By exchanging model updates instead of raw data, FL minimizes bandwidth usage and storage costs [25]. Furthermore, FL facilitates collaborative learning in resource-constrained environments. Local training on edge devices or distributed systems makes FL suitable for medical imaging applications in remote or low-resource settings. However, FL also faces challenges. Heterogeneity of data, arising from variations in imaging protocols, equipment, and patient populations, can impact model performance. Handling non-identically distributed (Non-IID) data remains an ongoing research challenge. Communication and computational overhead pose additional hurdles, particularly when dealing with large-scale medical imaging datasets. Efficient compression and transmission techniques are being explored to mitigate these challenges. Ensuring fairness, model interpretability, and addressing security concerns are also important considerations in FL for medical imaging. Despite these challenges, FL holds great promise in advancing medical imaging research and applications while safeguarding patient privacy [26]–[30].

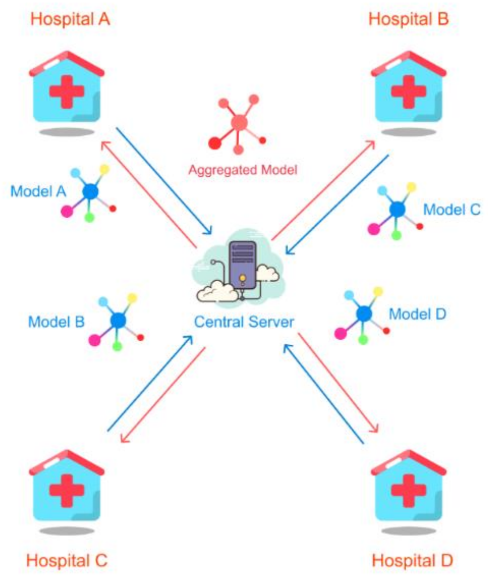

Figure 1. Implementation of FL in medical image analysis

3. EXPLAINABLE AI FOR MEDICAL IMAGE DIAGNOSIS- SOME REFELECTONS

Building ML or AI systems with human-comprehensible explanations for their predictions and actions is the focus of XAI study [31]. In contrast to the black box nature of complex AI models, XAI aims to make the decision-making process understandable to humans. This is achieved through interpretable model architectures, such as decision trees, or by generating post-hoc explanations that highlight the key factors influencing the model's outputs [32]. To account for the model's behaviour over the entire dataset, XAI considers both local and global explanations. Predictions can be improved with the use of local explanations. User-centric explanations are made specifically for the target group, making them meaningful and understandable. Users can judge the dependability and fairness of AI systems by evaluating the quality and credibility of explanations, which is a key component of XAI. When used to healthcare, XAI can improve doctors' ability to understand AI-based diagnoses, build trust, and streamline collaboration between humans and AI. XAI's overarching goal is to promote transparency, accountability, and trust in AI systems by bridging the gap between how they make decisions and how humans can understand them [33]. Figure 2 illustrates the importance of XAI in the detection of medical images.

Interpretability is of utmost importance in medical image diagnosis as it enhances the understanding and trust in AI systems' decisions. Accurate and trustworthy diagnoses are crucial in medical imaging for patient care and treatment planning. Radiologists and physicians benefit from being able to analyse AI models so that they can understand the thinking behind the factors used in these systems' classifications and forecasts [34]. Interpretability helps validate a diagnosis since it provides clear justifications for the model's decisions about which parts of a picture to focus on. Additionally, it allows doctors to assess the model's efficacy, spot any biases or errors, and make educated decisions based on the AI system's findings. Additionally, interpretability encourages collaboration between human specialists and AI systems, enabling radiologists to employ AI as a tool in their diagnostic workflow while guaranteeing a sound and credible final diagnosis [35].

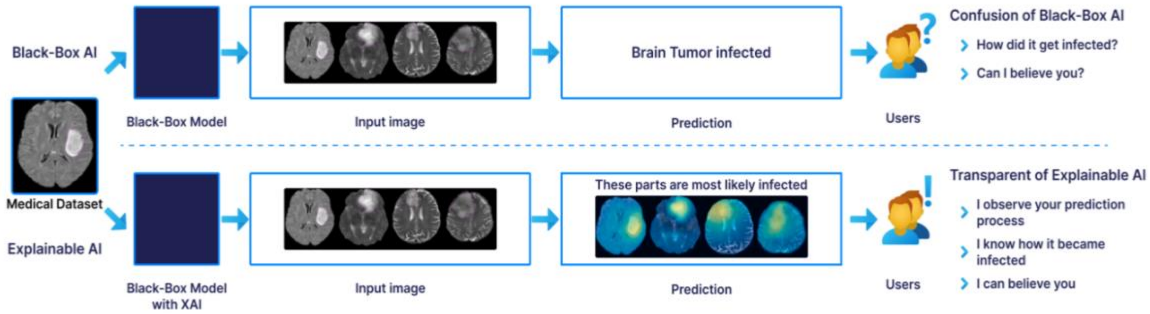

Figure 2. XAI in medical image analysis-a high level view

AI models can be utilized in a variety of ways to create explanations for medical imaging diagnosis. With the use of gradient-weighted class activation mapping (CAM), we can generate heat maps that show where in the input image the model's prediction was most strongly impacted [36]. Visualization is a powerful tool for learning more about the characteristics and potential sites of a brain tumor. Model-independent local interpretation is achieved by tweaking the input image and monitoring the model's predictions to create explanations. It identifies the most important regions or pixels for a prediction, providing insight into what drives the model's decision [37].

Shapley additive explanations (SHAP) provides relevance values to various aspects in the input image based on how much they contribute to the model's output. It provides a unifying framework for feature attribution, helping doctors to grasp the relative value of various picture attributes in medical image diagnosis. Rule-based explanations generate a set of interpretable rules based on the AI model's learnt decision boundaries [38]. These rules can specify explicit requirements or thresholds that aid in understanding how various elements or combinations of features affect the model's predictions. Feature visualisation techniques create visual representations of the model's learned characteristics. These visualisations assist physicians in comprehending the learnt representations and identifying relevant patterns or structures in imaging data that are symptomatic of medical image characteristics [39].

4. STATE-OF-THE-ART TECHNIQUES FOR MEDICAL IMAGE DIAGNOSIS USING FL AND XAI

In the field of medical imaging diagnostics, FL can be utilized to enable collaborative model training across several institutions, while simultaneously safeguarding the confidentiality of patient data. The following are some FL-based ways for analysing medical images and making a diagnosis. FL makes it possible for many institutions or hospitals to collaborate on the training of a medical image diagnosis model without releasing raw patient data. Instead, the model is trained at each individual institution using the institution's own data, and only model updates or gradients are sent to a centralized server for the purpose of aggregation. This decentralized technique safeguards data while simultaneously capitalizing on the accumulated knowledge contained across diverse datasets [40]. By utilizing pre-trained models from a variety of institutions, FL makes transfer learning possible in the field of medical picture identification. Each institution is able to train a base model using its own data, and then transfer the model to a shared, federated dataset in order to make additional adjustments to it. The model's convergence can be achieved more quickly with this strategy, and diagnostic performance is enhanced.

By training multiple models at different institutions and then pooling their predictions, FL can be used to build an ensemble of brain tumour diagnosis models. The ensemble is built by aggregating the predictions of individual models trained by different institutions [41]. The diagnostic system may be made more accurate and robust through the use of ensemble learning. Brain tumour imaging data can benefit from FL's ability to assist the extraction of collaborative features. Institutions are able to apply pre-trained deep learning models to extract features from their local data, and they can then just release the extracted features as opposed to the original, unfiltered imagegraphs. After that, a global model for analysing unexpected information can be constructed by using these traits as building blocks [42]. The term federated transfer learning refers to the process that results when FL and transfer learning are combined. Institutions collaborate in order to train a standardized baseline model with the use of their own data, and then they share this model with one another so that they can fine-tune it using their own data. The federated model that was produced as a result [43] draws upon sources from a wide range of organizations.

These FL-based methods improve medical image diagnosis by using the potential of distributed data while maintaining privacy. By employing FL, institutions can collectively train models with larger and more diversified datasets, resulting in enhanced diagnostic model accuracy and generalisation. Model-agnostic strategies seek answers that are independent of the exact AI model employed. Model behaviour and the impact of input features on model predictions are analysed using these methods [44], regardless of the model's underlying architecture. Note that the specific AI model, data characteristics, and the requirements of doctors or end-users will all play a role in deciding which explanation technique to adopt. To create thorough and contextually appropriate explanations for medical image diagnosis, many methodologies might be integrated or altered [45].

Transparent and easily interpretable XAI approaches can improve the efficacy of models used in the diagnosis of brain tumors. In this article, we show many XAI-based approaches for detecting brain tumors [46]. In order to produce maps that highlight the regions of the brain that contributed most to the model's conclusion, a heat map visualization approach such as grad-CAM might be used. These heat maps provide visual explanations for medical image identification by identifying which portions of the image were most influential in the diagnosis [47]. XAI approaches can evaluate the significance of various picture features or regions in the diagnosis of medical images. Clinicians can determine which features the model depends on for predictions by measuring the importance of specific features such as forms, textures, or intensity patterns. Rule-based techniques generate interpretable rules based on the AI model's learned decision limits [48]. These rules can be explicit conditions or thresholds that assist doctors in understanding how individual aspects of the image or combinations of features contribute to medical image diagnosis. Model-agnostic strategies seek answers that are independent of the exact AI model employed. These methods concentrate on analysing the model's behaviour and the impact of input variables on its predictions, allowing physicians to comprehend the decision-making process across different models or architectures [49]. Contextual explanations can be provided by XAI techniques by analysing the reasons behind the model's predictions within the framework of medical knowledge and guidelines. By relating the model's results to standard medical metrics, these justifications increase the diagnostic process's interpretability and reliability.

XAI algorithms can create natural language explanations of the reasoning and elements impacting medical image diagnosis [50]. These human-readable explanations can assist physicians in understanding and communicating the model's decisions to patients or other healthcare professionals. Clinicians can improve their understanding of AI model decision-making, the factors influencing forecasts, and the model's outputs by applying XAI methods into brain tumor diagnoses. This promotes openness, trust, and cooperation between AI systems and medical staff, leading to improved diagnostic accuracy and clinical interpretability of medical images [51].

Combining FL and XAI approaches can provide an effective approach for medical image identification. Here are some examples of how to integrate FL and XAI. Using FL, researchers from different institutions can work together to fine-tune a single XAI model. Any organization can participate in the training process by contributing local data while yet protecting user privacy. The resulting XAI model would explain the algorithm's reasoning for its predictions in language that clinicians could understand [52]. Instead of aggregating solely model updates or gradients in FL, the federated aggregation process might incorporate explanations given by XAI approaches. This will enable the aggregation server to aggregate not just the model parameters but also the generated explanations from other institutions, resulting in an intelligible and transparent final model.

Multiple XAI models can be trained in tandem across organizations using FL XAI models for hypothesis generation in brain tumour diagnosis are developed independently by each institution [53]. The ensemble, which is comprised of the individual models' forecasts and explanations, then makes the ultimate call. This method yields both accurate predictions and a variety of interpretable explanations. In this method, institutions work together to extract interpretable features from their local medical image imaging data using XAI approaches. Rather than distributing raw images or models, the extracted features are distributed and utilised to train a global XAI model. This enables the extraction of shared knowledge and the production of transparent and interpretable explanations for medical image diagnosis [54].

Advantages of FL and XAI in brain tumour diagnosis include interpretable decision-making, protection of patient privacy, and group learning. It combines the power of distributed data and model training with the capacity to offer explanations for physicians that improve transparency, trust, and understanding. Hybrid approaches, which combine the capabilities of FL and XAI, offer the potential to increase accurate and interpretable medical image diagnosis while resolving privacy concerns [55].

5. EVALUATION OF FL AND XAI-BASED MEDICAL IMAGE DIAGNOSIS TECHNIQUES

Quantitative assessments of accuracy, precision, recall, specificity, and overall performance of FL and XAI-based approaches in identifying brain tumours are provided by the performance evaluation metrics. The area under the ROC curve (AUC-ROC) and average precision metrics are frequently used to evaluate the discriminatory power and precision-recall tradeoff of models [56]. Table 1 depicts some of the evaluation metrics employed by FL and XAI-based diagnostic techniques.

Table 1. Metrics used in evaluation of medical image diagnosis techniques

Metrics	Formula	Description
Accuracy	(TP+TN) / (TP+TN+FP+FN)	Overall correctness of diagnostic predictions
Precision	TP / (TP+FP)	Proportion of correctly predicted positive cases
Recall (sensitivity)	TP / (TP+FN)	Proportion of correctly predicted positive cases
Specificity	TN / (TN+FP)	Proportion of correctly predicted negative cases
F1 Score	2×(precision×recall) / (precision+recall)	Harmonic mean of precision and recall
AUC-ROC	-	Measures the model's ability to distinguish between positive and negative cases
Average precision	-	Average precision across different recall levels based on the precision-recall curve

The evaluation techniques used for medical image diagnostic techniques based on FL and XAI also have certain challenges and limitations. Some common challenges and limitations associated with these evaluation techniques are limited ground truth data, data heterogeneity, and lack of evaluation protocols, interpretability and ethical considerations [57]. Obtaining a reliable and comprehensive ground truth for medical image diagnosis can be challenging. The availability of high-quality ground truth data is crucial to the reliability of the evaluation metrics. An incomplete or inaccurate ground truth can introduce biases and affect the performance evaluation. Variations in picture acquisition techniques, equipment, and patient demographics can introduce unavoidable discrepancies into FL data. Performance evaluations can be impacted by such data heterogeneity, making it hard to generalize results to other settings or datasets [58]. The absence of standardized evaluation protocols for FL and XAI based medical image diagnosis poses challenges in comparing and replicating results across different studies. The variations in evaluation methodologies and metrics used make it challenging to establish consistent benchmarks and assess the progress in the field.

XAI techniques aim to provide interpretable explanations for model decisions [59]. However, there can be a trade-off between interpretability and performance. While more complicated models may be more accurate, they may not be explainable, while simpler models with better interpretability may be less effective. Balancing interpretability and performance are a challenge in evaluating and selecting medical image diagnostic techniques. In some cases, the FL or XAI based medical image diagnostic models may utilize black-box models that are inherently complex and lack interpretability. Even though XAI methods can offer explanations after the fact, they might not be able to reveal all of the black-box models' inner workings.

It is common practice to evaluate FL and XAI based brain tumour diagnostic methods on predefined datasets that may not be representative of the complexity and variety of actual clinical settings [60]. Generalizing the performance results to diverse patient populations, imaging protocols, and clinical environments can be challenging. Evaluation techniques should also consider ethical aspects, such as fairness, bias, and accountability [61]. Ensuring that the evaluation metrics and methodologies address these concerns can be challenging, particularly in complex and sensitive domains like medical imaging. Addressing these challenges and limitations requires ongoing research and collaboration between the medical and AI communities. Developing standardized evaluation protocols, improving access to high-quality ground truth data, and advancing interpretability techniques are crucial for reliable and meaningful evaluation of FL and XAI based medical image diagnostic techniques [62].

Future research in FL and XAI for medical image diagnostic techniques holds significant potential for advancements in accuracy, interpretability, and clinical applicability. There are some promising directions for future research in this field. Developing more efficient and secure FL frameworks specifically tailored for medical imaging applications can further improve collaborative medical image diagnosis [63]. This includes addressing challenges related to data heterogeneity, communication efficiency, privacy preservation, and model aggregation techniques. Combining imaging modalities including MRI, CT, and positron emission tomography with clinical and genetic data improves our understanding of brain tumors [60]. In order to improve diagnostic precision and choice-making, future studies can investigate FL and XAI methods that enable the integration and analysis of multi-modal data.

Knowledge can be transferred from well-established datasets to domains with minimal labelled data by exploring transfer learning and domain adaption approaches within the FL and XAI frameworks [64]. This approach can improve model performance and generalization in medical image diagnosis across different healthcare institutions or regions. Advancing XAI techniques specific to deep learning models can provide more detailed and actionable explanations for medical image diagnosis. Within this framework, novel techniques for feature extraction, visualization, and comprehending the decision-making process of deep neural networks can be investigated. The interpretability and decision-making power of FL and XAI models can be enhanced with the incorporation of clinical knowledge and domain experience [65]. Incorporating medical guidelines, previous knowledge, and expert annotations can improve the diagnosis processes explain ability and reliability.

It is essential to examine methods for evaluating model robustness and calculating uncertainty in brain tumour detection using FL and XAI. Uncertainty estimate helps clinicians learn how confident they can be in model predictions, which improves their ability to make decisions and handle risks. To demonstrate the efficacy, reliability, and impact of FL and XAI-based medical image diagnosis approaches, comprehensive validation studies in real-world clinical settings are required [66]. Collaboration with healthcare institutions and clinicians is critical to evaluate the performance, usability, and integration of these techniques into routine clinical practice. Future research should address ethical and legal challenges associated with FL and XAI-based diagnostic techniques, including privacy, data ownership, bias, and accountability. Ensuring transparency, fairness, and compliance with regulatory requirements are crucial for the responsible deployment of these technologies. By exploring these research directions, FL and XAI-based medical image diagnostic techniques can advance the field of medical imaging, improving accuracy, interpretability, and the overall quality of patient care [67].

6. CASE STUDIES

6.1. Explainable artificial intelligence in healthcare

The novel method to Parkinson's disease diagnosis makes use of XAI. To make DaTscan images more easily understood, the authors employ local interpretable model-agnostic explanations (LIME). This research shows how crucial it is for medical AI to be open and explainable to both doctors and patients [65]. To deal intensively with the topic of XAI in biomedicine and emphasize the crucial importance of developing trustworthy AI systems for medical professionals and patients. The article examines various methods and tactics for establishing interpretability in AI-driven medical decision-making and emphasizes the importance of ethical and transparent AI. XAI to gain new insights into tumor microenvironmental factors associated with improved outcomes in breast cancer patients. The study not only proves the diagnostic capabilities of AI, but also its ability to explain medical data in an understandable way, thus leading to better patient care [66]. Developed an XAI model for glaucoma diagnosis. Aside from accurate diagnostic performance, the model's interpretability allows clinicians to understand the reasons behind its predictions. This strategy increases confidence in AI-powered medical decision-making and delivers actionable insights for healthcare professionals [67]. To be used XAI to diagnose biological mental illnesses. The study not only helps identify mental health problems by creating a model with explainable properties, but also provides interpretable insights into the diagnostic process, supporting more effective treatments and interventions [68].

To examine the ability of XAI to predict cardiovascular events using molecular data. Explainable models enable clinicians to gain an understanding of the aspects that contribute to risk assessment, enabling more informed patient treatment and prevention initiatives [69]. In-depth research on the uses of XAI in healthcare over the past decade was undertaken [70]. The comprehensive analysis examines the growth of XAI in healthcare including varied application cases and shows the revolutionary importance of interpretability in medical AI. The study examines numerous applications of XAI in medical diagnostics and surgical decision making. It highlights the importance of transparency in AI-driven healthcare, enabling physicians to not only trust but also understand AI-generated suggestions, leading to better patient outcomes [71].

To proposed an automatic diagnosis strategy for myocarditis diseases using depth transformers and XAI in cardiac MRI processing. This method helps doctors' better plan patient care by enhancing diagnostic precision and illuminating how AI arrives at its conclusions [72]. To be involvement of XAI in deep learning-based medical image processing. By making AI-generated data interpretable, the work contributes to the credibility of image-based diagnostics and makes them more accessible for clinical use [73]. To be developed a methodology to assess blood test parameters that may be helpful in COVID-19 diagnosis. By providing clear and understandable insights into diagnostic assessments, the interpretable AI models used in this work help address pandemic-related health challenges [74]. In advanced stage ovarian cancer, found that XAI can be used to predict complete surgical cytoreduction. The work enables individual treatment plans and surgical planning by integrating interpretability into the diagnostic process [75]. According to reported the use of XAI in lung cancer screening models. The interpretability of the AI system helps doctors not only diagnose diseases but also understand the underlying elements that impact diagnostic evaluations [76]. To categorize prostate cancer, created an XAI model using ultrasound and MRI data. The study increases trust in AI-powered healthcare workflows while improving diagnostic accuracy and providing insights into the variables that influence AI conclusions. Sadeghi *et al.* [58] performed a systematic analysis of XAI's applications in healthcare in 2023. It highlights XAI's promise to improve the openness and interpretability of AI systems, while describing the various applications of XAI in medical diagnosis and therapy [77]. A scoping review carried out in 2023 to examine the developments, advantages and possible uses of XAI in medicine. The increasing use of XAI in healthcare and its potential impact on clinical practice are highlighted in the review [78].

6.2. Federated learning in healthcare

To address privacy and security issues, developed a FL paradigm for edge-based analysis of healthcare data. It represents a ground breaking method for data protection and collaborative analysis [79]. For FL, address dynamic contracts in smart healthcare applications. The study focuses on resource-efficient model training that enables healthcare companies to operate efficiently while limiting data sharing [80]. To examined the future of digital health from a FL perspective. They emphasize the opportunities for collaborative research and privacy protection in healthcare, laying the foundation for a more secure and data-driven healthcare ecosystem [81]. Li *et al.* [82] examined FL applications in the context of the internet of things (IoT) in depth, with a particular focus on healthcare. In addition to discussing the potential of FL for IoT applications in healthcare, the poll also covers privacy and security concerns. Clinical outcomes in patients with COVID-19 were predicted using FL. The benefits of FL for healthcare decision support are highlighted by this real-world example, which is especially relevant in the context of pandemics [83]. Used biological data to conduct a systematic review of FL applications. The potential of FL in health research is demonstrated, and the most important results and successes in this sector are summarized in the overview [84]. To be presented FL and fine-grained privacy for use in medical image analysis. They stress the need for confidentiality safeguards even as medical imaging diagnostics benefit from the pooled resources of multiple data sources [85].

Propose a FL strategy for protecting healthcare data in big data environments. The study addresses data security concerns while allowing healthcare companies to collaborate on data-driven research [86]. Developed work using FL to diagnose heart problems in a hospital setting. In this context, AI techniques for protecting privacy are crucial, and research suggests that the potential for secure analysis of medical data is being discussed [87]. Liu *et al.* [88] investigated the development of intelligent healthcare systems based on FL that are both secure and efficient. Their efforts ensure that AI-driven healthcare solutions are trustworthy by bolstering data security and model accuracy [88]. Privacy-protecting FL algorithms in healthcare systems were analysed. The evaluation highlights potential privacy concerns related to collaborative health research and provides recommendations for enhancing data protection [89].

To offered a reinforced FL technique for healthcare IoT devices using particle swarm optimization. This method improves model performance and data security as well as the effectiveness of AI-driven healthcare applications [90]. Patients' lengths of stay in hospitals can be predicted via FL, researchers are able to improve healthcare resource allocation and patient care planning while keeping private medical information secure [91]. To investigated XAI's function in chronic wound categorization and showed its potential applications beyond diagnosis. Wound treatment is the centre of the study, although the broad applications of XAI in medicine are also discussed [92]. Introduced a better, more understandable AI tool for hospital recommendations. By selecting hospitals based on transparent and interpretable criteria, this platform improves healthcare and patient outcomes [93]. FL to estimate the length of stay of hospital patients, resulting in more efficient resource management and patient care planning [94]. The study shows the value of FL in healthcare. A method for detection and forecasting based on AI was presented [95].

6.3. Lessons learned

In medical image diagnosis, grad-CAM and LIME offer significant advantages that can be exploited depending on the specific diagnostic goals and the type of underlying deep learning models. Grad-CAM excels at spatial localization and visual explanations, making it particularly valuable for cases where it is critical to understand where the model focuses within an image. This is particularly beneficial in scenarios such as radiology, where precise location of an abnormality or lesion in an X-ray or MRI image is essential for an accurate diagnosis. Radiologists and healthcare professionals can benefit from grad-CAM's clear visualizations that highlight areas of interest and provide insight into the model's decision-making process.

On the other hand, LIME's model-agnostic approach proves advantageous when dealing with a variety of deep learning architectures and data types commonly encountered in medical image analysis. Its ability to provide local, case-specific explanations can help understand why a particular diagnosis was made for a particular image. In cases where model interpretability is critical to building trust and ensuring the ethical use of AI in healthcare, LIME's ability to explain individual predictions contributes to transparency and accountability. Additionally, LIME can be valuable for uncovering the reasons for unexpected model behaviour, such as misclassifications or cases where the confidence level of the model is low. Ultimately, the choice between grad-CAM and LIME in medical image diagnosis should depend on the specific requirements of the diagnostic task, the type of medical image data to be analysed, and the level of interpretability required for medical professionals to confidently incorporate AI-driven insights integrate their clinical workflows. In some scenarios, combining both methods can provide a comprehensive solution that enables both spatial localization and instance-specific explanations to improve the diagnostic process and improve patient care.

7. DISCUSSION

This study investigated the effects of FL and XAI for medical image diagnosis. While earlier studies have explored the impact of FL and XAI in various domains, they have not explicitly addressed its influence on the specific challenges and nuances associated with medical image diagnosis. Existing research has primarily focused on general applications of FL and XAI, overlooking the unique requirements and considerations essential for the accurate and reliable diagnosis of medical images. This research gap highlights the need for a dedicated investigation into the tailored implementation and effectiveness of FL and XAI techniques in the context of medical image diagnosis, addressing issues such as interpretability, transparency, and trustworthiness in the healthcare domain.

In our investigation into the impact of FL and XAI for medical image diagnosis, we identified a noteworthy correlation between the application of FL and XAI techniques and improved diagnostic outcomes. The integration of FL demonstrated a significant enhancement in collaborative learning across decentralized medical data sources, leading to heightened accuracy and robustness in image diagnosis. Additionally, the incorporation of XAI not only contributed to accurate predictions but also provided valuable insights into the decision-making process, fostering transparency and interpretability in the diagnostic outcomes. Our findings underscore the potential of FL and XAI in revolutionizing medical image diagnosis by addressing issues of data privacy, model transparency, and diagnostic reliability in a collaborative healthcare environment.

While this study delves into the impact of FL and XAI for medical image diagnosis, offering a comprehensive analysis of their influence on diagnostic outcomes, it is crucial to acknowledge certain limitations into consideration. The scope of our investigation may not encompass the full spectrum of medical imaging conditions or diverse patient populations, potentially impacting the generalizability of our findings. The dynamic nature of healthcare practices and evolving technology landscapes introduce an inherent challenge in capturing real-world variations that may affect the application of FL and XAI in different clinical settings. The interpretability of XAI models, while improved, may still pose challenges in complex medical scenarios.

Our study on the impact of FL and XAI for medical image diagnosis reveals promising outcomes in terms of diagnostic accuracy and interpretability. To further advance this field, future research could delve into exploring the optimal configurations of FL and XAI models for specific medical imaging modalities. Investigating the integration of real-time feedback mechanisms to continuously improve model performance and interpretability in dynamic clinical environments would be valuable. Examining the ethical considerations and addressing the challenges associated with deploying these technologies in real-world healthcare settings should be a focal point for future investigations. By addressing these aspects, future research can contribute to refining and optimizing the practical implementation of FL and XAI, ultimately enhancing their impact on medical image diagnosis in a meaningful and responsible manner.

8. CONCLUSION

There have been important discoveries made in the field of medical image diagnosis as a result of the use of FL and XAI. FL allows for the training of collaborative models across several institutions without the need to share private patient information. FL-based medical image diagnostic models have demonstrated improved accuracy in detecting and classifying tumors, benefiting from the aggregation of diverse datasets. XAI approaches provide interpretable explanations for model predictions, improving doctors' understanding of decision-making processes and allowing them to make more informed therapeutic judgements. FL and XAI's transparency aids in model validation and develops trust in their diagnostic skills. Furthermore, FL protects data privacy by keeping patient data decentralised, assuring regulatory compliance. FL-based models have demonstrated the ability to generalise across institutions, incorporating differences in imaging methods and patient groups. When FL and XAI are used in clinical treatment, ethical considerations including fairness and bias are essential for ensuring responsible and ethical use. Brain tumour identification relies heavily on the accuracy, interpretability, cooperation, and privacy afforded by FL and XAI when applied to medical imaging.

9. FUTURE RESEARCH

There are important clinical and future research implications for using FL and XAI systems to detect brain tumours. FL and XAI, when used to the diagnosis of brain tumors in a clinical setting, can improve both model accuracy and interpretability, leading to more informed treatment decisions. Due of FL's collaborative character, knowledge sharing across institutions is encouraged while patient privacy is maintained, leading to more reliable and generalizable models. Transparent explanations for model predictions are provided by XAI techniques, enhancing physician acceptance and fostering greater trust. FL and XAI have applications in healthcare that include better patient outcomes, individualized treatment strategies, and more efficient use of available resources. Future research should concentrate on creating standardised evaluation procedures, addressing problems with data heterogeneity, and enhancing the readability of FL and XAI models. Furthermore, for the discipline to advance, study is needed on the ethical implications of FL and XAI, transfer learning, and the integration of multimodal data. Overall, FL and XAI have revolutionary implications for medical practise and upcoming research in medical image diagnosis, promising improvements in precision, interpretability, collaboration, and patient-specific care.

REFERENCES

[1] G. S. Madhuri, T. R. Mahesh, and V. Vivek, "A novel approach for automatic brain tumor detection using machine learning algorithms," in *Big Data Management in Sensing*, 1st ed., New York: River Publishers, 2022, pp. 87–101.

[2] B. Alther, V. Mylius, M. Weller, and A. Gantenbein, "From first symptoms to diagnosis: initial clinical presentation of primary brain tumors," *Clinical and Translational Neuroscience*, vol. 4, no. 2, 2020, doi: 10.1177/2514183x20968368.

[3] M. L. -Velazquez *et al.*, "Advances in brain tumor surgery for glioblastoma in adults," *Brain Sciences*, vol. 7, no. 12, 2017, doi: 10.3390/brainsci7120166.

[4] J. R. M. -Figueroa and E. Q. Lee, "Brain tumors," *The American Journal of Medicine*, vol. 131, no. 8, pp. 874–882, Aug. 2018, doi: 10.1016/j.amjmed.2017.12.039.

[5] M. Toğaçar, B. Ergen, and Z. Cömert, "BrainMRNet: brain tumor detection using magnetic resonance images with a novel convolutional neural network model," *Medical Hypotheses*, vol. 134, Jan. 2020, doi: 10.1016/j.mehy.2019.109531.

[6] K. Aldape *et al.*, "Challenges to curing primary brain tumours," *Nature Reviews Clinical Oncology*, vol. 16, no. 8, pp. 509–520, 2019, doi: 10.1038/s41571-019-0177-5.

[7] T. Hossain, F. S. Shishir, M. Ashraf, M. A. A. Nasim, and F. M. Shah, "Medical image detection using convolutional neural network," in *2019 1st international conference on advances in science, engineering and robotics technology (ICASERT) IEEE*, 2019, pp. 1–6.

[8] J. Amin, M. Sharif, M. Yasmin, and S. L. Fernandes, "A distinctive approach in brain tumor detection and classification using MRI," *Pattern Recognition Letters*, vol. 139, pp. 118–127, Nov. 2020, doi: 10.1016/j.patrec.2017.10.036.

[9] L. Kapoor and S. Thakur, "A survey on brain tumor detection using image processing techniques," in *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence*, Jan. 2017, pp. 582–585, doi: 10.1109/CONFLUENCE.2017.7943218.

[10] M. Gurbina, M. Lascu, and D. Lascu, "Tumor detection and classification of MRI brain image using different wavelet transforms and support vector machines," *2019 42nd International Conference on Telecommunications and Signal Processing, TSP 2019*, pp. 505–508, 2019, doi: 10.1109/TSP.2019.8769040.

[11] A. ARI and D. HANBAY, "Deep learning based brain tumor classification and detection system," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 26, no. 5, pp. 2275–2286, Sep. 2018, doi: 10.3906/elk-1801-8.

[12] M. A. Gulum, C. M. Trombley, and M. Kantardzic, "A review of explainable deep learning cancer detection models in medical imaging," *Applied Sciences*, vol. 11, no. 10, 2021, doi: 10.3390/app11104573.

[13] E. Q. Lee *et al.*, "Optimizing eligibility criteria and clinical trial conduct to enhance clinical trial participation for primary brain tumor patients," *Neuro-Oncology*, vol. 22, no. 5, pp. 601–612, Jan. 2020, doi: 10.1093/neuonc/noaa015.

[14] J. Kang, Z. Ullah, and J. Gwak, "MRI-based brain tumor classification using ensemble of deep features and machine learning classifiers," *Sensors*, vol. 21, no. 6, Mar. 2021, doi: 10.3390/s21062222.

[15] T. T. Tang, J. A. Zawaski, K. N. Francis, A. A. Qutub, and M. W. Gaber, "Image-based classification of tumor type and growth rate using machine learning: a preclinical study," *Scientific Reports*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.1038/s41598-019-48738-5.
[16] T. A. Soomro *et al.*, "Image segmentation for MR medical image detection using machine learning: a review," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 70–90, 2023, doi: 10.1109/RBME.2022.3185292.
[17] B. Camajori Tedeschini *et al.*, "Decentralized federated learning for healthcare networks: a case study on tumor segmentation," *IEEE Access*, vol. 10, pp. 8693–8708, 2022, doi: 10.1109/ACCESS.2022.3141913.
[18] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. Vincent Poor, "Federated learning for internet of things: a comprehensive survey," *IEEE Communications Surveys and Tutorials*, vol. 23, no. 3, pp. 1622–1658, 2021, doi: 10.1109/COMST.2021.3075439.
[19] K. Hu *et al.*, "Federated learning: a distributed shared machine learning method," *Complexity*, pp. 1–20, 2021, doi: 10.1155/2021/8261663.
[20] M. M. Amiri, D. Gunduz, S. R. Kulkarni, and H. V. Poor, "Federated learning with quantized global model updates," *arXiv-Computer Science*, pp. 1-16, 2020.
[21] V. Mugunthan, A. P. -Bueno, and L. Kagal, "Privacyfl: a simulator for privacy-preserving and secure federated learning," in *International Conference on Information and Knowledge Management*, 2020, pp. 3085–3092, doi: 10.1145/3340531.3412771.
[22] M. J. Sheller *et al.*, "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020, doi: 10.1038/s41598-020-69250-1.
[23] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Machine Intelligence*, vol. 2, no. 6, pp. 305–311, 2020, doi: 10.1038/s42256-020-0186-1.
[24] E. Darzidehkalani, M. G. -Rad, and P. M. A. V. Ooijen, "Federated learning in medical imaging: part II: methods, challenges, and considerations," *Journal of the American College of Radiology*, vol. 19, no. 8, pp. 975–982, 2022, doi: 10.1016/j.jacr.2022.03.016.
[25] Prayitno *et al.*, "A systematic review of federated learning in the healthcare area: from the perspective of data properties and applications," *Applied Sciences*, vol. 11, no. 23, 2021, doi: 10.3390/app112311191.
[26] S. Banabilah, M. Aloqaily, E. Alsayed, N. Malik, and Y. Jararweh, "Federated learning review: fundamentals, enabling technologies, and future applications," *Information Processing and Management*, vol. 59, no. 6, 2022, doi: 10.1016/j.ipm.2022.103061.
[27] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable AI: a review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, Dec. 2020, doi: 10.3390/e23010018.
[28] A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018, doi: 10.1109/ACCESS.2018.2870052.
[29] S. Muneer and M. A. Rasool, "AA systematic review: explainable Artificial Intelligence (XAI) based disease prediction," *International Journal of Advanced Sciences and Computing*, vol. 1, no. 1, pp. 1–6, 2022.
[30] J. H. Thrall, D. Fessell, and P. V. Pandharipande, "Rethinking the approach to artificial intelligence for medical image analysis: the case for precision diagnosis," *Journal of the American College of Radiology*, vol. 18, no. 1, pp. 174–179, 2021, doi: 10.1016/j.jacr.2020.07.010.
[31] A. B. -Montero *et al.*, "Artificial intelligence and machine learning for medical imaging: a technology review," *Physica Medica*, vol. 83, pp. 242–256, 2021, doi: 10.1016/j.ejmp.2021.04.016.
[32] F. A. Özbay and E. Özbay, "Brain tumor detection with mRMR-based multimodal fusion of deep learning from MR images using Grad-CAM," *Iran Journal of Computer Science*, vol. 6, no. 3, pp. 245–259, 2023, doi: 10.1007/s42044-023-00137-w.
[33] M. U. Ansari, "ULIME: uniformly weighted local interpretable model-agnostic explanations for image classifiers," *M.Sc. Thesis*, School of Computing, Queen's University, Ontario, Canada, 2022.
[34] S. Ahmed, S. N. Nobel, and O. Ullah, "An effective deep CNN model for multiclass brain tumor detection using MRI images and SHAP explainability," in *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Feb. 2023, pp. 1–6, doi: 10.1109/ECCE57851.2023.10101503.
[35] M. Esmaeili, R. Vettukattil, H. Banitalebi, N. R. Krogh, and J. T. Geitung, "Explainable artificial intelligence for human-machine interaction in brain tumor localization," *Journal of Personalized Medicine*, vol. 11, no. 11, Nov. 2021, doi: 10.3390/jpm11111213.
[36] N. Rieke *et al.*, "The future of digital health with federated learning," *npj Digital Medicine*, vol. 3, no. 1, Sep. 2020, doi: 10.1038/s41746-020-00323-1.
[37] S. Denissen *et al.*, "Transfer learning on structural brain age models to decode cognition in MS: a federated learning approach," *medRXiv*, pp. 1-6, 2023, doi: 10.1101/2023.04.22.23288741.
[38] D. Gandhi, J. Mehta, N. Shah, and R. Mangrulkar, "Federated learning for brain tumor segmentation on the cloud," in *Cloud Computing Technologies for Smart Agriculture and Healthcare*, Boca Raton: Chapman and Hall/CRC, pp. 261–278, 2021.
[39] R. Chelghoum, A. Ikhlef, A. Hameurlaine, and S. Jacquir, "Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images," *IFIP Advances in Information and Communication Technology*, pp. 189–200, 2020, doi: 10.1007/978-3-030-49161-1_17.
[40] D. Truhn *et al.*, "Encrypted federated learning for secure decentralized collaboration in cancer image analysis," *Medical Image Analysis*, vol. 92, 2024, doi: 10.1016/j.media.2023.103059.
[41] R. K. Gupta, S. Bharti, N. Kunhare, Y. Sahu, and N. Pathik, "Brain tumor detection and classification using cycle generative adversarial networks," *Interdisciplinary Sciences – Computational Life Sciences*, vol. 14, no. 2, pp. 485–502, 2022, doi: 10.1007/s12539-022-00502-6.
[42] J. Qian, H. Li, J. Wang, and L. He, "Recent advances in explainable artificial intelligence for magnetic resonance imaging," *Diagnostics*, vol. 13, no. 9, Apr. 2023, doi: 10.3390/diagnostics13091571.
[43] T. Jiao *et al.*, "Deep learning with an attention mechanism for differentiating the origin of brain metastasis using MR images," *Journal of Magnetic Resonance Imaging*, vol. 58, no. 5, pp. 1624–1635, 2023, doi: 10.1002/jmri.28695.
[44] B. Taşcı, "Attention deep feature extraction from brain MRIs in explainable mode: DGXAINet," *Diagnostics*, vol. 13, no. 5, 2023, doi: 10.3390/diagnostics13050859.
[45] S. Maqsood, R. Damaševičius, and R. Maskeliūnas, "Multi-modal brain tumor detection using deep neural network and multiclass SVM," *Medicina*, vol. 58, no. 8, Aug. 2022, doi: 10.3390/medicina58081090.
[46] R. A. Zeineldin *et al.*, "Explainability of deep neural networks for MRI analysis of brain tumors," *International Journal of Computer Assisted Radiology and Surgery*, vol. 17, no. 9, pp. 1673–1683, Apr. 2022, doi: 10.1007/s11548-022-02619-x.
[47] M. Eder, E. Moser, A. Holzinger, C. J. -Quartier, and F. Jeanquartier, "Interpretable machine learning with brain image and survival data," *BioMedInformatics*, vol. 2, no. 3, pp. 492–510, 2022, doi: 10.3390/biomedinformatics2030031.
[48] D. H. Mahlool and M. H. Abed, "Distributed brain tumor diagnosis using a federated learning environment," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 6, pp. 3313–3321, 2022, doi: 10.11591/eei.v11i6.4131.

[49] L. Gaur, M. Bhandari, T. Razdan, S. Mallik, and Z. Zhao, "Explanation-driven deep learning model for prediction of brain tumour status using MRI image data," *Frontiers in Genetics*, vol. 13, Mar. 2022, doi: 10.3389/fgene.2022.822666.
[50] D. Saraswat *et al.*, "Explainable AI for healthcare 5.0: opportunities and challenges," *IEEE Access*, vol. 10, pp. 84486–84517, 2022, doi: 10.1109/ACCESS.2022.3197671.
[51] S. Hossain, A. Chakrabarty, T. R. Gadekallu, M. Alazab, and M. J. Piran, "Vision transformers, ensemble model, and transfer learning leveraging explainable AI for brain tumor detection and classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 3, pp. 1261–1272, 2024, doi: 10.1109/JBHI.2023.3266614.
[52] S. Ullah, M. Ahmad, S. Anwar, and M. I. Khattak, "An intelligent hybrid approach for brain tumor detection," *Pakistan Journal of Engineering and Technology*, vol. 6, no. 1, pp. 42–50, 2023, doi: 10.51846/vol6iss1pp34-42.
[53] A. Rahman *et al.*, "Federated learning-based AI approaches in smart healthcare: concepts, taxonomies, challenges, and open issues," *Cluster Computing*, vol. 26, no. 4, pp. 2271–2311, Aug. 2023, doi: 10.1007/s10586-022-03658-4.
[54] P. Saxena, A. Maheshwari, and S. Maheshwari, "Predictive modeling of brain tumor: a deep learning approach," *Advances in Intelligent Systems and Computing*, vol. 1189, pp. 275–285, 2021, doi: 10.1007/978-981-15-6067-5_30.
[55] M. Ghassemi, L. Oakden-Rayner, and A. L. Beam, "The false hope of current approaches to explainable artificial intelligence in health care," *The Lancet Digital Health*, vol. 3, no. 11, pp. 745–750, 2021, doi: 10.1016/S2589-7500(21)00208-9.
[56] K. Aggarwal, M. Manso Jimeno, K. S. Ravi, G. Gonzalez, and S. Geethanath, "Developing and deploying deep learning models in brain magnetic resonance imaging: a review," *NMR in Biomedicine*, vol. 36, no. 12, Dec. 2023, doi: 10.1002/nbm.5014.
[57] P. N. Srinivasu, N. Sandhya, R. H. Jhaveri, and R. Raut, "From blackbox to explainable AI in healthcare: existing tools and case studies," *Mobile Information Systems*, pp. 1–20, Jun. 2022, doi: 10.1155/2022/8167821.
[58] Z. Sadeghi *et al.*, "A brief review of explainable artificial intelligence in healthcare," *arXiv-Computer Science*, pp. 1-23, 2023.
[59] T. Senevirathna, V. H. La, S. Marchal, B. Siniarski, M. Liyanage, and S. Wang, "A survey on XAI for beyond 5G security: technical aspects, use cases, challenges, and research directions," *arXiv-Computer Science*, pp. 1-49, 2022.
[60] S. I. Nafisah, G. Muhammad, M. S. Hossain, and S. A. AlQahtani, "A comparative evaluation between convolutional neural networks and vision transformers for COVID-19 detection," *Mathematics*, vol. 11, no. 6, 2023, doi: 10.3390/math11061489.
[61] L. Dao and N. Q. Ly, "A comprehensive study on medical image segmentation using deep neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 3, pp. 167–184, 2023, doi: 10.14569/IJACSA.2023.0140319.
[62] A. Kaur, G. Dong, and A. Basu, "GradXcepUNet: explainable AI based medical image segmentation," in *Smart Multimedia*, pp. 174–188, 2022, doi: 10.1007/978-3-031-22061-6_13.
[63] D. Dave, H. Naik, S. Singhal, and P. Patel, "Explainable AI meets healthcare: a study on heart disease dataset," *arXiv-Computer Science*, pp. 1-23, 2020.
[64] P. R. Magesh, R. D. Myloth, and R. J. Tom, "An explainable machine learning model for early detection of Parkinson's disease using LIME on DaTSCAN imagery," *Computers in Biology and Medicine*, vol. 126, Nov. 2020, doi: 10.1016/j.compbiomed.2020.104041.
[65] J. Lötsch, D. Kringel, and A. Ultsch, "Explainable artificial intelligence (XAI) in biomedicine: making AI decisions trustworthy for physicians and patients," *BioMedInformatics*, vol. 2, no. 1, pp. 1–17, 2022, doi: 10.3390/biomedinformatics2010001.
[66] D. Chakraborty *et al.*, "Explainable artificial intelligence reveals novel insight into tumor microenvironment conditions linked with better prognosis in patients with breast cancer," *Cancers*, vol. 13, no. 14, Jul. 2021, doi: 10.3390/cancers13143450.
[67] S. Oh, Y. Park, K. J. Cho, and S. J. Kim, "Explainable machine learning model for glaucoma diagnosis and its interpretation," *Diagnostics*, vol. 11, no. 3, 2021, doi: 10.3390/diagnostics11030510.
[68] A. M. Hilal *et al.*, "Modeling of explainable artificial intelligence for biomedical mental disorder diagnosis," *Computers, Materials and Continua*, vol. 71, no. 2, pp. 3853–3867, 2022, doi: 10.32604/cmc.2022.022663.
[69] A. M. Westerlund, J. S. Hawe, M. Heinig, and H. Schunkert, "Risk prediction of cardiovascular events by exploration of molecular data with explainable artificial intelligence," *International Journal of Molecular Sciences*, vol. 22, no. 19, 2021, doi: 10.3390/ijms221910291.
[70] H. W. Loh, C. P. Ooi, S. Seoni, P. D. Barua, F. Molinari, and U. R. Acharya, "Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022)," *Computer Methods and Programs in Biomedicine*, vol. 226, 2022, doi: 10.1016/j.cmpb.2022.107161.
[71] Y. Zhang, Y. Weng, and J. Lund, "Applications of explainable artificial intelligence in diagnosis and surgery," *Diagnostics*, vol. 12, no. 2, 2022, doi: 10.3390/diagnostics12020237.
[72] M. Jafari *et al.*, "Automatic diagnosis of myocarditis disease in cardiac MRI modality using deep transformers and explainable artificial intelligence," *arXiv-Computer Science*, pp. 1-27, 2022.
[73] B. H. M. van der Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," *Medical Image Analysis*, vol. 79, Jul. 2022, doi: 10.1016/j.media.2022.102470.
[74] L. M. Thimoteo, M. M. Vellasco, J. Amaral, K. Figueiredo, C. L. Yokoyama, and E. Marques, "Explainable artificial intelligence for COVID-19 diagnosis through blood test variables," *Journal of Control, Automation and Electrical Systems*, vol. 33, no. 2, pp. 625–644, 2022, doi: 10.1007/s40313-021-00858-y.
[75] A. Laios *et al.*, "Explainable artificial intelligence for prediction of complete surgical cytoreduction in advanced-stage epithelial ovarian cancer," *Journal of Personalized Medicine*, vol. 12, no. 4, 2022, doi: 10.3390/jpm12040607.
[76] K. Kobylińska, T. Orłowski, M. Adamek, and P. Biecek, "Explainable machine learning for lung cancer screening models," *Applied Sciences*, vol. 12, no. 4, 2022, doi: 10.3390/app12041926.
[77] M. R. Hassan *et al.*, "Prostate cancer classification from ultrasound and MRI images using deep learning based explainable artificial intelligence," *Future Generation Computer Systems*, vol. 127, pp. 462–472, 2022, doi: 10.1016/j.future.2021.09.030.
[78] R. González-Alday, E. García-Cuesta, C. A. Kulikowski, and V. Maojo, "A scoping review on the progress, applicability, and future of explainable artificial intelligence in medicine," *Applied Sciences*, vol. 13, no. 19, 2023, doi: 10.3390/app131910778.
[79] S. Hakak, S. Ray, W. Z. Khan, and E. Scheme, "A framework for edge-assisted healthcare data analytics using federated learning," in *2020 IEEE International Conference on Big Data, Big Data 2020*, 2020, pp. 3423–3427, doi: 10.1109/BigData50022.2020.9377873.
[80] W. Y. B. Lim *et al.*, "Dynamic contract design for federated learning in smart healthcare applications," *IEEE Internet of Things Journal*, vol. 8, no. 23, pp. 16853–16862, 2021, doi: 10.1109/JIOT.2020.3033806.
[81] Q. Li, B. He, and D. Song, "Model-contrastive federated learning," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10708–10717, doi: 10.1109/CVPR46437.2021.01057.
[82] L. Li, Y. Fan, M. Tse, and K. Y. Lin, "A review of applications in federated learning," *Computers and Industrial Engineering*, vol. 149, 2020, doi: 10.1016/j.cie.2020.106854.

[83] I. Dayan *et al.*, “Federated learning for predicting clinical outcomes in patients with COVID-19,” *Nature medicine*, vol. 27, no. 10, pp. 1735–1743, 2021.
[84] M. G. Crowson *et al.*, “A systematic review of federated learning applications for biomedical data,” *PLOS Digital Health*, vol. 1, no. 5, 2022, doi: 10.1371/journal.pdig.0000033.
[85] M. Adnan, S. Kalra, J. C. Cresswell, G. W. Taylor, and H. R. Tizhoosh, “Federated learning and differential privacy for medical image analysis,” *Scientific Reports*, vol. 12, no. 1, Feb. 2022, doi: 10.1038/s41598-022-05539-7.
[86] G. Dhiman *et al.*, “Federated learning approach to protect healthcare data over big data scenario,” *Sustainability*, vol. 14, no. 5, 2022, doi: 10.3390/su14052500.
[87] M. M. Yaqoob *et al.*, “Modified artificial bee colony based feature optimized federated learning for heart disease diagnosis in healthcare,” *Applied Sciences*, vol. 12, no. 23, 2022, doi: 10.3390/app122312080.
[88] W. Liu, Y. Zhang, G. Han, J. Cao, H. Cui, and D. Zheng, “Secure and efficient smart healthcare system based on federated learning,” *International Journal of Intelligent Systems*, 2023, doi: 10.1155/2023/8017489.
[89] X. Gu, F. Sabrina, Z. Fan, and S. Sohail, “A review of privacy enhancement methods for federated learning in healthcare Systems,” *International Journal of Environmental Research and Public Health*, vol. 20, no. 15, 2023, doi: 10.3390/ijerph20156539.
[90] E. H. Houssein and A. Sayed, “Boosted federated learning based on improved particle swarm optimization for healthcare IoT devices,” *Computers in Biology and Medicine*, vol. 163, 2023, doi: 10.1016/j.compbiomed.2023.107195.
[91] S. Rani, A. Kataria, S. Kumar, and P. Tiwari, “Federated learning for secure IoMT-applications in smart healthcare systems: a comprehensive review,” *Knowledge-Based Systems*, vol. 274, Aug. 2023, doi: 10.1016/j.knosys.2023.110658.
[92] S. Sarp, M. Kuzlu, E. Wilson, U. Cali, and O. Guler, “The enlightening role of explainable artificial intelligence in chronic wound classification,” *Electronics*, vol. 10, no. 12, 2021, doi: 10.3390/electronics10121406.
[93] Y.-C. Wang, T.-C. T. Chen, and M.-C. Chiu, “An improved explainable artificial intelligence tool in healthcare for hospital recommendation,” *Healthcare Analytics*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2023.100147.
[94] M. M. Rahman, D. Kundu, S. A. Suha, U. R. Siddiqi, and S. K. Dey, “Hospital patients’ length of stay prediction: a federated learning approach,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 7874–7884, 2022, doi: 10.1016/j.jksuci.2022.07.006.
[95] F. Ahmed, M. Asif, M. Saleem, U. F. Mushtaq, and M. Imran, “Identification and prediction of brain tumor using VGG-16 empowered with explainable artificial intelligence,” *International Journal of Computational and Innovative Sciences*, vol. 2, no. 2, pp. 24–33, 2023.

BIOGRAPHIES OF AUTHORS

Sivakumar Muthuramalingam is currently working as Associate Professor in Department of Computer Science and Engineering at Thiagarajar College of Engineering, Madurai. He received his Ph.D. from Anna University in 2023. His research focuses on artificial intelligence, machine learning, deep learning, and medical image processing. He can be contacted at email: mskcse@tce.edu or sivsiva.kumar.21@gmail.com.

Padmapriya Thiyagarajan is currently working as Assistant Professor in Department of Applied Mathematics and Computational Sciences at Thiagarajar College of Engineering, Madurai. She received her Ph.D. in the Department of Computer Science and Applications, Gandhigram Rural InstituteDeemed University, Gandhigram in 2022. She received her M.Phil. in Computer Science (2018-2019) from Gandhigram Rural Institute-Deemed University, Gandhigram. She received her M.E. in Computer Science and Engineering (2011-2013) from Sethu Institute of Technology, Virudhunagar. She received her M.Sc. in Software Engineering from RVS College of Engineering and Technology (2011), Dindigul. Her research focuses on artificial intelligence, machine learning, deep learning, medical image processing, software engineering, and data analytics. She can be contacted at email: stpca@tce.edu or stpadmapriya@gmail.com.