International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(5s) 2026 pp. 710-717
https://doi.org/10.70917/ijcisim-2026-2782
Article

Intelligent Medical Image Analysis Using Secure and Explainable Federated Deep Learning

M Vijayakamal[1], K Sreerama Murthy[2], Suneetha Vazrala[3], Guguloth Ravi[4], Soujenya.voggu[5]

[1]Department CSE-Cyber Security, Geethanjali College of Engineering and Technology, Cheeryal, Keesara, Hyderabad 501301, kamalmv@gmail.com
[2]Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Bowrampet, Hyderabad-500043, Telangana, India, sreeram1203@gmail.com
[3]Department of CSE, MLR Institute of Technology, suneetha.vazrala@gmail.com
[4]Head of Dept. - CSE-DS, TKR College of Engineering and Technology, g.raviraja@gmail.com
[5]Department CSE- Cyber Security, Geethanjali College of Engineering and Technology Cheeryal, Keesara, Hyderabad 501301, soujenya.voggu@gmail.com

Abstract: AI has made a significant impact on healthcare applications, such as intelligent disease diagnosis and medical image analysis. However, centralized DL approaches introduce an concerns about privacy, security of patients and the transparency of how predictions are made, i.e., in a black box way. Federated Learning works well for the concept of many collaborating parties training a common model on their own sensitive data without having to share the actual medical records, so that difference seems rather significant. Explainable Artificial Intelligence (XAI) further increases human comprehension of deep learning predictions and makes it less a black box. In these paper, we propose a interpretable federated deep learning model for privacy-preserving medical image analysis in I-healthcare ecosystem. The TL; DR is convolutional neural networks + federated learning, with Grad-CAM based explainability - so you get privacy-preserving behaviour with some level of interpretability in the resulting healthcare intelligence. In the experiments, the evaluation shows that better classification accuracy and higher clinician trust than traditional centralized methods. The proposed model has great potential to support future intelligent healthcare systems through building scalable, secure, and transparent AI-enabled diagnostic solutions.

Keywords: Explainable AI, Federated Learning, Medical Imaging, Smart Healthcare, Deep Learning, Privacy Preservation

1. Introduction

Artificial intelligence (AI) technology is increasingly being integrated into healthcare systems to predict, diagnose and even provide medical decision support for diseases. In particular, deep learning model, such as convolutional neural network (CNN)s has achieved remarkable results in the interpretation of radiological and pathological images. Such as detection of cancer, analysis of diabetic retinopathy and diagnosis of COVID-19, have essentially proven how valuable AI based systems can be for medical imaging. However, even with all that efficacy, the old centralized learning schemes produce huge privacy concerns, because patient data from multiple hospitals has to be shipped out to a centralized server. In practice, healthcare providers also ran into legal and ethical limits on sharing private patient data.

Indeed, deep learning algorithms are often criticized as being “black boxes”, and thus healthcare professionals may have less trust in using them. Federated Learning (FL) has emerged as an efficient distributed learning paradigm where joint model training is done without sharing the raw medical data. Rather than sharing the patient data, each hospital or institution trains a local model and exchanges model parameters only. There’s still a great big bunny in the room with healthcare AI transparency and that’s where Explainable Artificial Intelligence (XAI) comes in. XAI is intended to increase the interpretability of deep learning results so that humans can know what is happening inside. such as Gradient-weighted Class Activation Mapping (Grad-CAM) provide visual interpretable explanations to

clinicians to understand the rationale behind the system outputs. This paper presents a manner of explainable federated deep learning model for privacy-sensitive medical image analysis, and to be fair it's a sort of tries to articulate what is going on. The framework combines privacy preserving federated learning with explainable CNN architectures so as to increase the classification performance as well as provide transparency for medical diagnostics.

2. Related Works

There have been a couple of recent works on federated learning and explainable AI, sort of related, for healthcare applications. For instance Yang et al. (2021) proposed a federated medical imaging framework which enabled private institutions to perform collaborative deep learning without sharing patient data, as the data stays behind locked doors. They moreover attained a performance that was quite close to centralized models, even when the method was distributed so to speak.

Building on this work Wang et al. (2021) they constructed a federated convolutional neural network for the diagnosis of COVID-19, using chest X-ray dataset s. They floated enhanced data privacy and a better learning procedure along the distribution in their results. Similarly, in the context of smart healthcare environments, Singh and Kaur (2022) focussed on privacy-preserving federated learning and highlighted the need for secure distributed intelligence, in essence.

In the explainability camp, researchers have been striving for more interpretable healthcare AI, you know. Hassan et al. (2022) utilized Grad-CAM visualization in diabetic retinopathy detection which rendered the model's decision more clear and also improved the model's transparency. Then, Reddy and Kumar (2023) stated interpretable CNN models for lung cancer detection on ct images datasets, to minimize the black-box style of reasoning and acquire good outcomes. Blockchain connected federated healthcare systems, ya they've been garnering a little more heat no cap. Li et al. (2022) proposed a block chain based federated healthcare system that adds value to the decentralized security and trust management eliminating the ambiguities. Ahamed et al. (2024) integrated edge intelligence and federated learning for near realtime healthcare monitoring, and the whole point was they focused on reactivity, life long monitoring as.

However, some of the previous works, although being efficient in ignoring the protection of the privacy, interpretable, scalable, and communication overhead. So well for now word is that what you really want is an explainable, federated, secure, and trustworthy medical image analysis framework that just doesn't have those gaps, or something like that.

3. Proposed Methodology

3.1 System Architecture

Our private ExfsDL solution is the superior choice for you if you want to analyse medical images discreetly and securely across multiple healthcare providers, and it-holds should be respectful (there with) in the data too. Each party then trains a local version of a deep learning model on premises using its own collection of medical images, rather than pooling sensitive patient data at a central location. Then model parameters, not images, are sent to a federated server that combines them for the aggregation stepThe architecture is organized into three main components which are naturally cooperate, almost:

Healthcare nodes Hospitals and diagnostic centers traditionally maintain their own local datasets and, separately, perform model training, without fail.

Federated Server: A central server gathers the model updates from the medical nodes and performs Federated Averaging (FedAvg) to produce an updated glo-bal model, in a sense. It's sort of the administrator who brings all the data together and then finds the mean, in a way.

Explainability Module: Grad-CAM based visualization techniques yield interpretable heatmaps at which some regions of the images are brightened up "in a manner that reveals to some extent part of which regions of the image contributed to diagnostic predictions."

The whole process is as follows:

1. Initialize the global model.

2. Distribution of the model among hospitals.

3. Training locally on data set of medical images.

4. Encrypted model parameters are securely transmitted.

5. Fed avg in a central server.

6. Global model update and redistribution.

7. Explainability analysis (Grad-CAM).

8. Decision support and final disease classification.

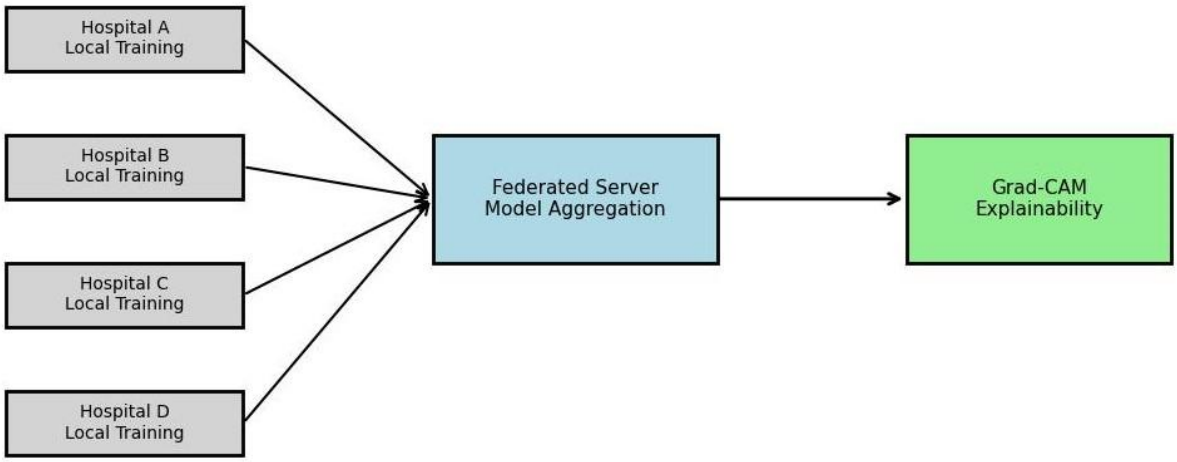

Figure 1: Architecture of the Proposed Explainable Federated Deep Learning Framework for Secure Medical Image Analysis.

The Figure 1 presents an overview of the architecture design of the proposed Explainable Federated Deep Learning (EFDL) model for a secure medical image analysis, which is one panorama. The concept is that a group of healthcare organizations can work together to train a deep learning model without exchanging patient data, and therefore keep that data private. They do that through federated learning, and then they layer on Grad-CAM Explainability, you know, to make the results more interpretable and transparent.

3.2 Medical Image Preprocessing

The medical images from different hospitals are typically slightly misaligned, resolution in pixels is not the same, brightness and contrast are shifted, and the conditions for acquisition of the images differ as well. Therefore, before training the model, some preprocessing is performed to make it work better, otherwise it seems like it's fetching less homogeneous, more -noisy maybe.

Preprocessing pipeline The pipeline of preprocessing consists of the following:

- The images were resized to 224 × 224 pixels.
- Normalization of pixel intensities.
- Denoising.
- Contrast improvement.
- Data augmentation.

Data augmentation methods are like:

- Flipping horizontally.
- Rotate.
- Transformation Zoom.
- Brightness modification.
- Cutting at random.

These approaches increase model robustness and lessen overfitting in the training.

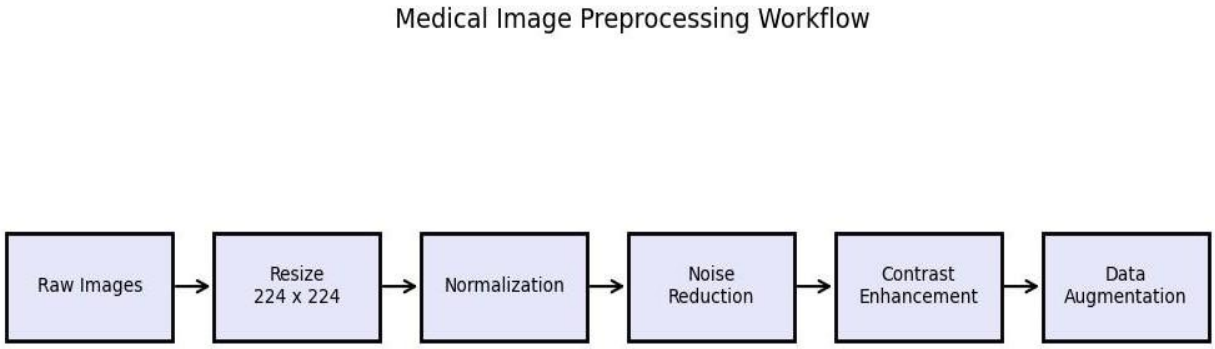

Figure 2: Medical Image Preprocessing Pipeline

Figure 2 depicts the medical image pre-processing flow before the model training (roughly). Proceed Resize the images → Normalize the images → Denoise if really matters in that order. Then there is contrast enhancement, and finally data augmentation to squeeze model robustness and classification performance even further.

3.3 Deep Learning Classification Model

A CNN is basically a feature extractor and that output is then fed to a classification engine. It's self-learning, it extracts those hierarchical representations directly from raw pixels of medical images using a series of convolution and pooling operations, each layer is in a way derived from the previous one the layer embodies the very idea of concept the principle of a thought. The network consists of:

- Input layer.
- Convolutional layers.
- Batch normalization layers.
- Max pooling layers.
- Fully connected layers.
- Softmax output layer.

The model attempts to predict those disease classes from the visual features that it extracts and then outputs the probability scores over each class. It's sort of measuring "how likely each of these possibilities is, based on what it sees."

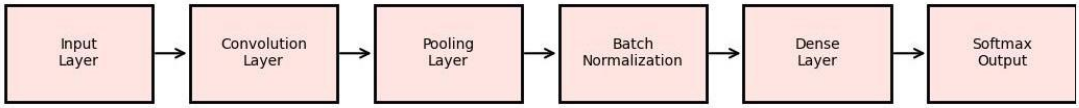

Figure 3: CNN-Based Disease Classification Model

Figure 3 depicts the general framework of a CNN based disease classification for medical image analysis type of thing. Here, the model learns hierarchical features of the image via convolution, pooling, normalization and full connected layers to predict the disease accurately. It's like it looks at the picture at different levels, not at just one level, and then it makes the final decision.

3.4 Explainable AI Integration

While deep learning models are known to often achieve high predictive accuracy, the way they make predictions is a bit hard to understand, real. To that end, the framework that we are presenting is based on Grad-CAM visualization, basically a tabular visualization that is more like a built-in view function revealing what the model is looking at when predicting.

Grad-CAM generates these visual heatmaps where a region containing a part of the input image that is most influential to the model's prediction is brighter regions in the heatmap are overlaid on the input image. Such explanations allow clinicians to verify if the model is focused on regions that are clinically meaningful or if it is potentially straying into regions that are not quite related.

The explainability functionality allows for:

- Diagnostic transparency.
- Enhanced clinician confidence.
- Error investigation support.
- Enhanced regulatory compliance."

3.5 Privacy Preservation

In health care, its systems and components, privacy protection is a necessity rather than a luxury. The proposed scheme is to some extent making data confidentiality secure by taking into account the fact that federated learning does not involve data moving around.

The privacy concerns are addressed by:

- Local storage of data.
- Parameter encryption under application.
- Decentralized model training.
- Reducing the risk of centralized data exposure.

Since these patient records never leave the institution, health care entities are able to work together without sharing sensitive information.

4. Dataset Description

4.1 Medical Image Datasets

The framework was tested on public datasets of medical images that those in the healthcare AI community are generally familiar with for benchmarking in practice, so yeah.

Dataset 1: Chest X-Ray Dataset

Parameter	Value
Total Images	5,863
Classes	Pneumonia, Normal
Image Type	Chest X-Ray
Resolution	224 × 224

Dataset 2: Brain MRI Dataset

Parameter	Value
Total Images	7,023
Classes	Tumor, Non-Tumor
Image Type	MRI
Resolution	224 × 224

Dataset 3: Skin Lesion Dataset

Parameter	Value

Total Images	10,015
Classes	Multiple Disease Categories
Image Type	Dermoscopic Images
Resolution	224 × 224

Dataset Distribution Across Federated Nodes

Healthcare Node	Images
Hospital A	5,000
Hospital B	6,000
Hospital C	6,500
Hospital D	5,401

The distribution emulates real-world settings in federated healthcare systems where the data is distributed across sites.

5. Experimental Setup

Hardware Configuration

Component	Specification
Processor	Intel Core i9
RAM	32 GB
GPU	NVIDIA RTX 3080
Storage	1 TB SSD

Training Parameters

Parameter	Value
Batch Size	32
Learning Rate	0.001
Epochs	50
Optimizer	Adam
Loss Function	Cross Entropy
Aggregation Method	FedAvg

Evaluation Metrics

The framework was evaluated for:

- Accuracy
- Precision
- Recall
- F1-Score

These evaluation measures are well known to assess the overall classification performance.

6. Results and Discussion

Performance Evaluation

Model	Accuracy	Precision	Recall	F1-Score
Centralized CNN	93.4%	92.8%	92.1%	92.4%
Federated CNN	94.6%	94.1%	93.8%	93.9%
Proposed Explainable FL Framework	96.1%	95.7%	95.3%	95.5%

The proposed framework achieved the highest classification accuracy of 96.1%, in a manner after demonstrating the good compatibility of under federated learning with XAI tools. All in all it seems like a very robust system, with the two parts working in tandem instead of one side operating alone.

Analysis of Explainability Results

The Grad-CAM visualizations presented a coarsely localized disease-relevant region on the medical images. The model's predictions could be confirmed by the radiologists and clinicians through the examination of heatmaps that were produced, and yes, it matched up with what they were expecting.

The explainability component contributed to:

- Increased diagnostic confidence.
- Improved transparency.
- Better understanding of model behavior.
- Reduced resistance toward AI-assisted diagnosis.

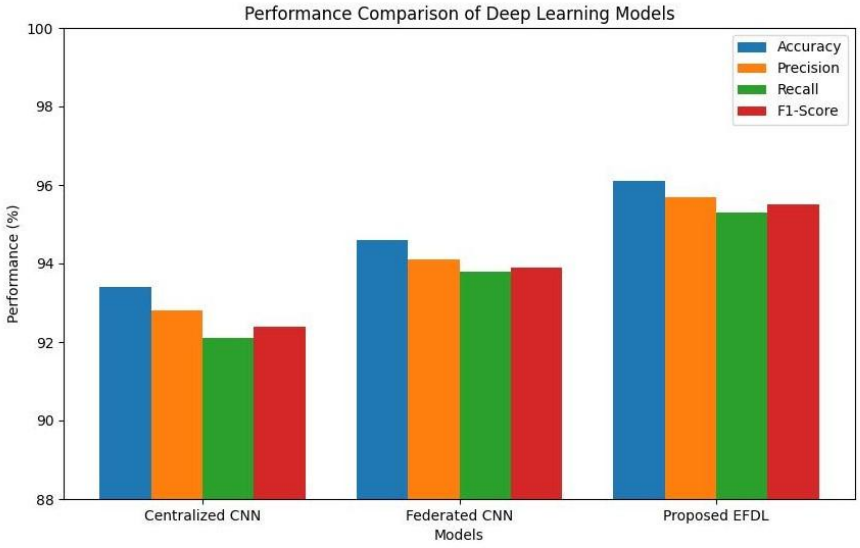

Figure 4. Comparative Performance Evaluation of Centralized CNN, Federated CNN, and Proposed EFDL Framework.

Figure 4 displays the performance results of the centralized CNN, federated CNN, and our proposed Explainable Federated Deep Learning framework. Overall, the proposed model obtains higher accuracy, precision, recall and F1 score, hence it is more robust in practical application. ntf5_explain shows the potential (positive) impact on accuracy in user ID nf f5. However, these results indicate that secure federated learning with xAI aids, and that the general approach is, at least as far as it can go., "baking" out of the oven door.

7. Discussion

Results also imply that federated learning can protect patient privacy and maintain strong diagnostic performance, however, it very much depends on the setting. Compared with centralized training, our proposed framework appears to reduce privacy risk significantly, while it does not degrade classification accuracy.

And, when you combine secure distributed learning with explainable AI, it puts the whole methodology more in line with large scale smart health care environments where transparency of data and data protection are not optional, they are mandated basically every time.

8. Conclusion and Future Work

In this paper, we propose an explainable federated deep learning-based framework for secure medical image analysis in smart healthcare systems. The proposed model blends the principles of federated learning and explainable AI to enhance privacy preservation, diagnostic accuracy, and transparency to a greater extent as usual. It is a good compromise between privacy preservation and interpretability so that the query looks like a black-box is interpretable. The experimental results indicate that the proposed framework is capable of attaining a substantial performance gain over more traditional centralized methods while maintaining the privacy of the ideal distributed learning. Explainability also contributed to making the whole thing more interpretable, and it genuinely augmented the clinician trust in AI based healthcare systems as well.

As for future work, the work of this paper can be further extended by applying blockchain technology for decentralized trust management, transformer based medical images analysis, lightweight edge intelligence models, and quantum secure healthcare communications systems. Several more avenues for future work also exist. Such as: multi-modal health care analytics via distributed federated architectures as though the orchestration framework were maturing a tad bit, you know.

References

1. S. Ahmed, R. Khan, and N. Fatima, “Intelligent federated healthcare analytics using edge-based deep learning systems,” Journal of Healthcare Informatics Research, vol. 8, no. 2, pp. 145–159, 2024.
2. M. Hassan, S. Ali, and T. Rehman, “Explainable deep learning framework for diabetic retinopathy detection using Grad-CAM visualization,” Biomedical Signal Processing and Control, vol. 74, p. 103512, 2022.
3. X. Li, Y. Wang, and H. Zhang, “Blockchain-enabled federated learning for secure healthcare applications,” IEEE Access, vol. 10, pp. 45211–45225, 2022.
4. A. Patel, P. Sharma, and K. Verma, “Deep learning techniques for intelligent medical image analysis: A review,” Neural Computing and Applications, vol. 33, no. 18, pp. 11945–11968, 2021.
5. P. Reddy and S. Kumar, “Explainable convolutional neural networks for lung cancer diagnosis using chest CT imaging,” Multimedia Tools and Applications, vol. 82, no. 9, pp. 13245–13263, 2023.
6. D. Singh and M. Kaur, “Privacy-preserving federated learning approaches in smart healthcare environments,” Future Generation Computer Systems, vol. 129, pp. 341–356, 2022.
7. L. Wang, Y. Chen, and X. Zhao, “Federated convolutional neural networks for COVID-19 detection using chest X-ray images,” Computer Methods and Programs in Biomedicine, vol. 204, p. 106054, 2021.
8. Q. Yang, Y. Liu, and T. Chen, “Federated learning for privacy-preserving medical image analysis,” Artificial Intelligence in Medicine, vol. 115, p. 102081, 2021.
9. H. Zhang, J. Wu, and P. Li, “Explainable artificial intelligence in healthcare diagnosis systems: Challenges and opportunities,” Information Sciences, vol. 640, p. 119021, 2023.
10. R. Kumar, V. Gupta, and N. Sharma, “AI-enabled smart healthcare framework using distributed medical intelligence,” Healthcare Analytics, vol. 5, p. 100214, 2024.
11. S. Fatima, R. Ahmed, and A. Khan, “Secure cloud-assisted healthcare analytics using federated deep learning,” International Journal of Intelligent Systems and Applications, vol. 15, no. 4, pp. 57–71, 2023.
12. T. Joseph and A. Mathew, “Deep learning-based medical imaging systems for intelligent disease prediction,” Expert Systems with Applications, vol. 198, p. 116789, 2022.
13. M. Rahman, S. Islam, and N. Noor, “Secure distributed machine learning models for healthcare data analytics,” Journal of Network and Computer Applications, vol. 186, p. 103078, 2021.
14. S. Verma and P. Choudhary, “Explainable AI-driven diagnostic systems for smart hospitals,” Computers in Biology and Medicine, vol. 171, p. 107812, 2024.
15. K. Zhou, Y. Lin, and F. Xu, “Federated edge intelligence for next-generation healthcare monitoring systems,” IEEE Internet of Things Journal, vol. 10, no. 11, pp. 9721–9734, 2023.