An Explainable Hybrid CNN–LSTM Framework with Federated Learning for Privacy-Preserving Healthcare Analytics

Dr. Yalakala Dinesh Kumar [1], Dr. Lowlesh Nandkishor Yadav[2]

[1] Post-Doctoral Researcher, [2] Supervisor

[1] Lincoln University College; [2] Suryodaya college of Engineering and Technology

Email ID: pdf.ydineshkumar@lincoln.edu.my ; lowlesh.yadav@gmail.com

Abstract: AI is transforming healthcare in the digital age, providing robust resources for medical diagnosis, clinical forecasting, and patient care monitoring. One of the challenges to trust is that many Deep Learning (DL) models are "black boxes" and do not provide transparency of how the decision was made. This opacity will make it difficult to depend on automation outputs. What complicates matters even more is the fact that healthcare data is highly diverse, drawn from a multitude of sources including medical scans, digital health records, and time-stamped clinical notes, making it difficult to combine and analyze. Sensitive patient data also poses privacy barriers to institutions to easily collaborate. This study presents a multi-modal DL approach that is explainable and privacy-preserving for intelligent healthcare analysis, addressing these inter-related challenges. The methodology utilizes Convolutional Neural Networks (CNNs) to pick out spatial patterns, Long Short-Term Memory (LSTM) networks to process sequences of clinical data, and Transformer-based attention mechanisms to intelligently fuse these different types of information. We have incorporated tools such as Explainable AI (XAI), Grad-CAM, SHAP, and LIME to bring a visual and feature-based explanation to each prediction, opening the door to the decision-making process. Further, in a federated learning environment, the model can be trained at several institutions without any raw patient data leaving the home server. In fact, the experimental results demonstrates that the framework achieves the accuracy of 98.32%, precision of 98.29%, recall of 98.43%, and F1 score of 98.36%, which are all better than the conventional deep learning techniques. In conclusion, this design enhances the interpretability, reliability, privacy, and overall generalizability, opening the door for AI systems that are reliable and scalable towards clinical application.

Keywords—: CNN, LSTM, XAI, CAM, SHAP, LIME.

Introduction

AI is revolutionizing healthcare with its ability to diagnose diseases more accurately, monitor patients continuously, interpret medical images, and provide predictive clinical support. In particular, DL, which can normally extract intricate patterns from the massive amounts of data in the medical field, has shown impressive results [1]. In various medical disciplines, such as radiology, cardiology, neurology, and pathology, the development of models including CNNs, RNNs and LSTM networks which has widely enhanced the accuracy and speed of disease identification and classification. The combination of cutting-edge computing and healthcare delivery has developed a new opportunity for intensifying the patient care and minimizing diagnostic errors [2]. However, implementing these models in an actual hospital or clinic in a reliable manner is a difficult hurdle to overcome. One of the issues is the 'black box' effect of many DL algorithms. Most of the models fails to provide reasons for predictions that are clear and understandable. Clinicians need to grasp the

reasoning behind an AI's recommendation, particularly when it comes to a critical diagnosis. If it's not interpretable, there's a natural hesitation to adopt the technology [3]. The other significant pain is the volume of healthcare data, especially the variety. Data comes in from different sources like imaging systems, electronic health records; lab reports, wearable technologies and patient histories, and they all have their own structure or format. This is a very complex and multi topic data that, in many cases, traditional ML methods fail to make sense of and find meaningful connections [4]. This requires more intelligent multimodal deep learning systems that incorporate spatial, temporal and contextual information and enhance predictive accuracy and clinical insight. In addition to interpretability and data fusion the generalization capability of a model is a major test [5]. A system that is trained using a small limited amount of data from one hospital might not work well for patients from another hospital or in a different clinical environment. Data imbalances, demographic differences, changes in data distribution, and so on can compromise the robustness of a model. The creation of generalizable and scalable AI frameworks is thus a priority. But there's also privacy and security [6]. Patient data is very personal and it is unethical and not legal to casually pool the data from various institutions and use it for centralized training. Centralization on one server makes it very vulnerable to data leaks and unauthorized access. This has brought the field one step closer to a new method of training models: federated learning, which entails training models locally at each institution while retaining the data safely there while collaborating on model construction [7].

Literature Survey

Integration of AI and deep learning in healthcare have seen massive uptake, as it is able to handle complex medical data and facilitate more intelligent decision making in clinical practice. A variety of intelligent machine learning and deep learning model has been suggested for disease diagnosis, patient monitoring, medical image analysis, and prediction [8]. Despite this, the issues of interpretability, multimodal data integration, model generalization and patient privacy continue to be the most prominent research challenges. This section delves into current research on deep learning, Explainable AI (XAI), multimodal learning, and privacy-enhancing approaches in healthcare. Transformer and attention-based models have recently gained popularity in healthcare analytics due to their superior feature representation capabilities and capacity to learn context. Attention mechanisms allow a model to focus on the most relevant information while making a prediction, improving classification and feature fusion accuracy [9]. Several studies have demonstrated the success of Vision Transformers (ViTs) and multimodal Transformer networks in combining information from images, clinical records and sensors [10]. The need for Explainable AI (XAI) has become a must-have in the ever-evolving landscape of deep learning in healthcare. With the rise of Artificial Intelligence in healthcare, researchers have started to investigate how to make clinical prediction systems transparent and explainable by XAI [11]. Grad-CAM has now been adopted as a benchmark method to understand which parts of a medical image were important for a CNN to classify. SHAP and LIME, meanwhile, gave feature-level explanations which helped to reveal a model's behavior for a single prediction [12]. Research has validated that embedding such XAI techniques into deep learning models increases clinician trust and enables more reliable decisions. But the existing literature focuses solely on interpretability, rather than on the other pressing needs for data privacy and secure, collaborative learning [13]. Thus, the preservation of privacy has become an important and essential research field in itself. The traditional centralized system for deep learning is unacceptable when patient data is sensitive and can be breached into a single server. Federated learning is a decentralized approach that comes into play here. It enables several

healthcare entities to jointly train an AI without sharing any patient data with one another [14]. The recent works demonstrated that federated learning can attain the prediction performance close to the centralized model and significantly improve privacy protection. However, the costs and benefits of the trade-offs are communication overheads, non-uniform data distribution between sites, and possibly slower model convergence [15].

References

1. Raza, Ali. *Secure and privacy-preserving federated learning with explainable artificial intelligence for smart healthcare system*. University of Kent (United Kingdom), 2023.
2. Vanangamoudiar, Maleappane Lawrence, and C. Senthilkumar. "Big Data–Driven Explainable AI And Privacy-Preserving Federated Deep Learning For Predictive Healthcare And Collaborative Clinical Intelligence." *INTERNATIONAL JOURNAL OF ADVANCES IN SIGNAL AND IMAGE SCIENCES* (2026): 2586-2611.
3. Nagamani, G. Muni, and Chanumolu Kiran Kumar. "Design of an improved graph-based model for real-time anomaly detection in healthcare using hybrid CNN-LSTM and federated learning." *Heliyon* 10, no. 24 (2024).
4. Prakash, Vijay, and Arun Kumar Shukla. "An AI-Enabled Privacy-Preserving Federated Learning Framework of Hybrid bi-LSTM-RNN-CNN for Deep Intelligent Intrusion Detection in Fog Computing in the IoT Domain." *Concurrency and Computation: Practice and Experience* 37, no. 27-28 (2025): e70291.
5. Alzamil, Ibrahim. "Federated deep learning for scalable and explainable load forecasting in privacy-conscious smart cities." *IEEE Access* (2025).
6. Manikandan, J., and K. Jayashree. "Enhancing lung nodule classification: a novel CViEBi-CBGWO approach with integrated image preprocessing." Journal of imaging informatics in medicine 37, no. 5 (2024): 2108.
7. Jiang, Shiyi. "Towards Personalized, Privacy-preserving, and Explainable AI for Healthcare." In *Technical Report*. Duke University, 2024.
8. Bukhari, Syed Muhammad Salman, Muhammad Hamza Zafar, Mohamad Abou Houran, Syed Kumayl Raza Moosavi, Majad Mansoor, Muhammad Muaaz, and Filippo Sanfilippo. "Secure and privacy-preserving intrusion detection in wireless sensor networks: Federated learning with SCNN-Bi-LSTM for enhanced reliability." *Ad Hoc Networks* 155 (2024): 103407.
9. Mahir, Shahariar Hossain, Md Tanjum An Tashrif, Aysha Siddika Shathe, Md Ahsan Karim, Anichur Rahman, Dipanjali Kundu, Tanoy Debnath, and Ghulam Muhammad. "PriFL-XAI: Hybrid privacy-preserving federated learning models for monkeypox detection through GAN augmentation and explainable AI." *Biomedical Signal Processing and Control* 112 (2026): 108426.
10. Fuzail, Muhammad Zubair, Irfan ud Din, Shakeel Ahmed, Abdulaziz Alhumam, and Abdul Hannan Khan. "Optimizing sepsis mortality prediction using hybrid federated learning and explainable AI framework." *Scientific Reports* (2026).