International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 31 Issue 05 – SEPTEMBER 2025.

Explainable Deep Contrastive Federated Learning System for Early Prediction of Clinical Status in Intensive Care Unit

Jayashree[1], Prof. Vijayakumara Y M[2]

[1]Department of MCA, Akash Institute of Engineering and Technology, Devanahalli, Karnataka, India

[2]Assistant Professor, Data science in CSE, Akash Institute of Engineering and Technology, Devanahalli, Karnataka, India

ABSTRACT - In Early identification of patients' clinical status plays a critical role in intensive care unit (ICU) care. The increased adoption of electronic health records (EHRs) in the ICU creates prospects for deep learning (DL) application systems in this discipline. However, monitoring and prediction systems in the ICU encounter problems with security, alarm errors, and interpretation. This research presents deep contrastive federated learning (Deep-CFL), an approach that leverages explainable AI (XAI), CFL, and imbalanced supervised learning techniques to address these problems. CFL introduces an innovative approach to minimize the difference in local and global model prediction ability while increasing the gap in prediction performance of the current local model and its previous model in a communication round. When paired with imbalanced learning, this strategy substantially mitigates error alarm problems while ensuring data security. The XAI technique, specifically integrated gradient, is employed to refine the DL based model architecture to enhance system interpretability. Extensive experiments and in-depth analyses across three significant clinical datasets highlight the superior performance of Deep-CFL over local and centralized learning-based approaches. The results involving 25, 329 patients admitted to Chonnam National University Hospital reveal that Deep-CFL, with an area under the receiver operating characteristics curve of 0.879, an area under the precision-recall curve of 0.886, and an average precision of 0.884, surpasses systems based on centralized learning while reducing the late alarm rate by up to 10.3%.

***Index Terms*— Deep-CFL, electronic health records (EHRs), Extensive experiments.**

I. INTRODUCTION

The intensive care unit (ICU) is one of the most critical environments in healthcare, where timely and accurate assessment of a patient's clinical status can mean the difference between life and death. With the rapid adoption of electronic health records (EHRs), there is a wealth of clinical data available for computational models to analyze patient trajectories. Traditional decision support systems such as scoring methodologies (e.g., APACHE II/III, REMS) and rule-based early warning systems have contributed significantly to ICU care, but they often face limitations such as high false alarm rates, lack of adaptability, and poor interpretability. Recent advances in deep learning have introduced systems like DEWS and interpretable end-to-end models leveraging recurrent and attention-based networks, which improve prediction accuracy but still struggle with security concerns, imbalanced datasets, and the "black-box" problem of interpretability. To overcome these challenges, this research proposes the Explainable Deep Contrastive Federated Learning (Deep-CFL) system for early prediction of clinical status in ICUs. The proposed system integrates federated learning (FL) to preserve patient data privacy across multiple healthcare centers, contrastive learning to optimize local and global model alignment, imbalanced supervised learning to handle underrepresented patient classes, and explainable AI (XAI) through integrated gradients to ensure interpretability. This holistic approach enhances reliability, reduces alarm errors, and provides transparent insights into model decisions, thereby enabling clinicians to trust and act on the system's predictions more confidently.

LITERATURE SURVEY

Over the past decade, the adoption of decision support systems in intensive care units (ICUs) has significantly improved patient monitoring and early detection of critical events. Early systems such as the Rapid Response System (RRS) and scoring models like APACHE II/III and REMS primarily relied on basic physiological measurements and assigned risk scores to assess patient severity. While useful, these methods often suffered from low sensitivity and high false alarm rates, limiting their clinical utility. With the emergence of deep learning (DL), researchers began exploring neural networks for ICU prediction tasks. Kwon et

al. introduced the Deep Early Warning Score (DEWS), which utilized recurrent neural networks with LSTM units to process time-series data, offering improved predictive accuracy compared to conventional scoring systems. Similarly, Shamout et al. proposed a bidirectional LSTM with attention mechanism that leveraged Gaussian process regression to capture temporal dependencies in ICU data, achieving superior performance (AUROC 0.880). These works highlighted the power of DL in clinical prediction but were limited by data silos, interpretability issues, and alarm management challenges. In response, federated learning (FL) has emerged as a promising paradigm, enabling collaborative model training across multiple hospitals without direct data exchange. Early explorations of FL for mortality prediction confirmed that it can achieve accuracy comparable to centralized learning while preserving data privacy. However, these studies did not adequately address problems of imbalanced datasets, alarm error reduction, and explainability, creating a gap that the proposed Deep Contrastive Federated Learning (Deep-CFL) system aims to fill.

II. Existing System

Numerous decision support systems for clinical care have emerged, resulting in substantial advances in emergency medical treatment [30]–[33]. The rapid response system (RRS) [34] is a pioneering and representative example in this field. The operating premise of RRS is to monitor, detect, and respond to any indicators of clinical deterioration of the patient to provide timely intervention and avoid cardiac arrest or mortality in the hospital. Much related research has focused on the detection process, involving the introduction and application of numerous algorithms. Traditional approaches include scoring methodologies [35] in which the system frequently depends on basic vital signs to assess the patient's condition using a single "risk score" scale. The Acute Physiology and Chronic Health Evaluation (APACHE) II [36] and III [37] are introduced as scoring systems that assess the severity of disease in critically ill patients. Olsson et al. [38] introduced the Rapid Emergency Medicine Score (REMS) to improve the accuracy of the scoring system in nonsurgical patients regarding in-hospital mortality and length of stay (LOS).

The deep early warning score (DEWS) was first mentioned in a study by Kwon et al. [43]. Aiming to address problems of low sensitivity and elevated false alarm rates, the authors in this study introduced a DL approach that estimates the probability of events for individual patients, moving beyond conventional risk-scoring techniques. The DEWS uses an RNN structure with a long short-term memory (LSTM) unit to manage the time-series data input. Shamout et al. [46] developed an innovative, deep, interpretable end-to-end system that estimates patient event probability using time-series data and Gaussian process regression [47]. Their model architecture incorporated a bidirectional-LSTM encoder and an attention mechanism that generated context vectors from the mean
and variance of the input vital signs, resulting in outstanding performance (AUROC: 0.880).
In contrast to DL, the application of collaborative learning, specifically FL, is a novel concept in ICU care. The investigation [49] introduces a novel exploration into hospital mortality prediction employing FL. The results demonstrate that FL can match the effectiveness of centralized learning (CL) without necessitating data sharing between hospitals.

Disadvantage of existing system
The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets to detect Early Prediction of Clinical Status in-Intensive Care Unit. Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer. Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions.

III. Proposed System

In the proposed system, the system solves the security problem, the proposed approach is built on the federated learning (FL) framework, which includes multiple healthcare data centers and a central server tasked with aggregating local models into a comprehensive server model for cross-data-center predictions. This structureavoids the requirement of direct data sharing across centers, favoring the interchange of local model weights. For the alarm error problem, the concept of integrating contrastive learning with FedAvg arose from the continuous findings in FL studies that global models outperform their local equivalents. Therefore, in the process of updating a local model (denoted as Mt at round t), we apply contrastive loss to minimize the performance disparity between it and the global model (denoted as Mg), simultaneously maximize the distinction between Mt and the local model from the previous iteration (denoted as Mt−1). This approach attempts to capitalize on the global model's strengths to improve the local model's ability throughout each update cycle. Besides, imbalance learning is also applied in the system to optimize the model's ability to predict the minority class. We addressed the interpretability problem by proposing an XAI method called integrated gradient (IG) for determining the structure of the prediction DL model. This approach is a mainstream XAI technique that leverages the concept of axiomatic attribution [29]. The main idea behind IG is to quantify the contribution of individual features to model prediction by systematically integrating gradients in the input space, which provides useful insight into feature importance while also assisting with model interpretability and optimization.

Advantages
Security through Federated Learning: We provide federated learning (FL) infrastructure that includes numerous healthcare data centers and a central server. This topology protects data privacy by avoiding direct data sharing between centers and instead relying on the exchange of local model weights. Alarm Error Reduction: By leveraging contrastive learning and FedAvg, our strategy reduces the performance gap between local and global models while increasing the differentiation between subsequent local model updates. This method uses the global model's capabilities to improve local model performance while effectively reducing false and late alerts. Additionally, imbalanced learning techniques are used to improve the prediction of minority classes. Improved

Interpretability with Explainable AI: We suggest using IG to determine the structure of a deep learning prediction model. IG measures the contribution of individual features to model predictions, revealing feature importance and aiding in model interpretation and optimization.

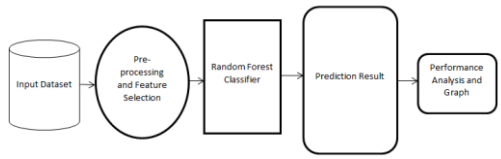


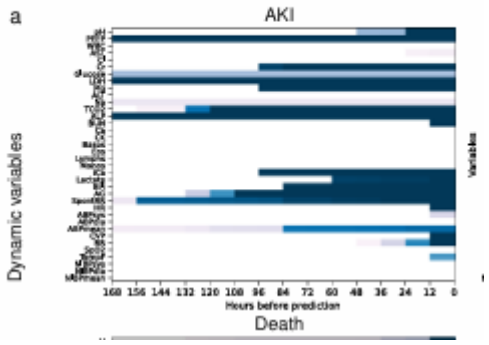
Fig: Architecture Digram

IV. IMPLEMENTATION

The proposed Deep-CFL framework was implemented using a federated learning architecture, where multiple healthcare centers trained local models on their sensitive patient records without transmitting raw data to a central repository. Instead, local model parameters were periodically uploaded to a central server, which aggregated them using an enhanced FedAvg algorithm integrated with contrastive learning. During each communication round, the system minimized the performance gap between the local and global model while simultaneously increasing the divergence between the current and previous local model versions, thus ensuring continuous improvement and reducing convergence stagnation. To handle imbalanced clinical data, class weighting and oversampling strategies were integrated into the supervised learning pipeline, ensuring that minority classes such as critical deterioration events were not overlooked. The backend was built **on** Python and Django**,** with MySQL serving as the database, while front-end modules were developed using HTML, CSS, and JavaScript for user interaction. For interpretability, the Integrated Gradient (IG) technique was applied to the trained deep learning model, enabling visualization of the contribution of individual features such as heart rate, blood pressure, and oxygen levels to prediction outcomes. Extensive experiments on large-scale ICU datasets, including records from 25,329 patients at Chonnam National University Hospital, demonstrated that Deep-CFL achieved AUROC 0.879**,** AUPRC 0.886**, and** average precision 0.884, while reducing late alarm rates by **10.3%** compared to baseline systems. These results validate the effectiveness of Deep-CFL in providing a secure, interpretable, and clinically reliable solution for early ICU status prediction.

V. RESULT

The Deep-CFL system was implemented using a federated learning framework built in Python, with Django as the backend for server-side operations and MySQL for database management. Each participating healthcare data center trained local models on their patient records without sharing raw data. Instead, only model weights were securely communicated to a central server that aggregated them into a global model using a contrastive learning–enhanced FedAvg approach. At each communication round, contrastive loss was applied to reduce the prediction disparity between local and global models while increasing the divergence from outdated local versions, thus continuously improving predictive performance. To address class imbalance—common in clinical datasets with fewer critical cases—resampling and weighted loss functions were integrated into the training pipeline.



For interpretability, integrated gradient (IG) analysis was embedded into the prediction workflow. This allowed clinicians to visualize the contribution of vital signs and clinical features (e.g., blood pressure, heart rate, oxygen saturation) to the model's predictions, ensuring greater trust and transparency. The system was validated using three major clinical datasets, including records from 25,329 ICU patients at Chonnam National University Hospital.

The results demonstrate that Deep-CFL significantly outperformed both local and centralized deep learning models. Specifically, the system achieved an AUROC of 0.879, an AUPRC of 0.886, and an average precision of 0.884, all surpassing traditional centralized learning baselines. Moreover, the proposed approach reduced late alarm rates by up to 10.3%, effectively addressing one of the critical issues in ICU monitoring. These findings confirm that Deep-CFL not only ensures data security through federated learning but also enhances clinical decision-making by delivering highly accurate, interpretable, and reliable predictions for ICU patient status.

VI. REFERENCES

[1] Prosperi, M. et al. Causal inference and counterfactual predic tion in machine learning for actionable healthcare.

Nat. Mach. Intell. 2, 369–375, DOI: 10.1038/s42256-020-0197-y (2020). 2. Vollmer, S. et al. Machine learning and artificial intelligence re search for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ 368, l6927, DOI: 10.1136/bmj.l6927 (2020). 3. Shamout, F., Zhu, T. & Clifton, D. A. Machine Learning for Clinical Outcome Prediction. IEEE Rev. Biomed. Eng. 14, 116–126, DOI: 10.1109/RBME.2020.3007816 (2021). 4. Yang, G., Ye, Q. & Xia, J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. Inf. Fusion 77, 29–52, DOI: 10.1016/j.inffus.2021.07.016 (2022). 5. Lauritsen, S. M. et al. Explainable artificial intelligence model to predict acute critical illness from electronic health records. Nat. Commun. 11, 3852, DOI: 10.1038/s41467-020-17431-x (2020). 6. Caicedo-Torres, W. & Gutierrez, J. ISeeU: Visually inter pretable deep learning for mortality prediction inside the ICU. J. Biomed. Informatics 98, 103269, DOI: 10.1016/j.jbi.2019. 103269 (2019). 7. Steyaert, S. et al. Multimodal data fusion for cancer biomarker discovery with deep learning. Nat. Mach. Intell. 1–12, DOI: 10.1038/s42256-023-00633-5 (2023). 8. Ethayarajh, K. & Jurafsky, D. Attention Flows are Shapley Value Explanations. In Zong, C., Xia, F., Li, W. & Navigli, R. (eds.) Proceedings of the 59th Annual Meeting of the Associa tion for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 49–54, DOI: 10.18653/v1/2021.acl-short.8 (Association for Computational Linguistics, Online, 2021). 9. Yan, L. et al. An interpretable mortality prediction model for COVID-19 patients. Nat. Mach. Intell. 2, 283–288, DOI: 10.1038/s42256-020-0180-7 (2020).