Integrating Machine Learning and Deep Learning for Multiclass Mortality Prediction in Healthcare Data

K. Devi Priya[1]
[1]*Research Scholar,*
Department of Computer Science,
S.V. University,
Tirupati, India.
Email: devipriyasurendar@gmail.com

Dr. S. Ramakrishna[2]
[2]*Professor,*
Department of Computer science,
S.V. University,
Tirupati, India.
Email: drsramakrishna.svu@gmail.com

***Abstract*- The prediction of mortality in cardiovascular patients demands models that balance accuracy, interpretability, and robustness. The current research work implementation presents a unified system that integrates ensemble machine learning (ML) approaches with complex deep learning (DL) architectures for multiclass mortality prediction. The ML component employs Random Forest, Gradient Boosting, and XGBoost classifiers, each evaluated through stratified sampling, confusion matrices, and precision–recall analysis. These models establish strong baselines, with XGBoost achieving the highest accuracy among the ensemble group. Building upon this foundation, the DL component introduces fully connected neural networks (FCNNs) enhanced with dropout and batch normalization, convolutional neural networks (CNNs) adapted for tabular data, and an FCNN augmented with attention mechanisms to capture feature importance. The CNN model demonstrated superior generalization, attaining validation accuracy above 96%, while the attention-based FCNN provided interpretability without compromising predictive strength. Comparative visualization of accuracy curves and error distributions underscores the complementary strengths of ML and DL approaches. This hybrid pipeline not only advances methodological rigor but also contributes to reproducible AI practices in healthcare, offering a scalable result for clinical choice validation in mortality risk assessment.**

Index terms: Mortality Prediction, Machine Learning (ML), Deep Learning (DL), Random Forest, Gradient Boosting XGBoost, Convolutional Neural Networks (CNN), Attention Mechanisms, Clinical Decision Support, Healthcare Data Analytics

I. INTRODUCTION

Mortality prediction in cardiovascular and critical care patients is a pressing challenge in healthcare, where accurate risk stratification can guide timely interventions and resource allocation. Fixed statistical already developed models frequently strive to capture nonlinear dependencies and high-dimensional feature interactions present in clinical datasets. In contrast, machine learning (ML) methods such as Random Forests, Gradient Boosting, and XGBoost have proved effective predictive performance on structured health records, offering robustness and interpretability through feature importance measures [1–3]. These ensemble approximates require have being effectively applied to mortality prediction tasks, establishing reliable baselines for clinical decision support [4,5].

While ML provides efficiency and scalability, deep learning (DL) architectures extend predictive capacity by modeling complex nonlinear relationships across heterogeneous variables. Fully connected neural networks (FCNNs) enhanced with dropout and batch normalization have shown promise in capturing latent feature hierarchies [6]. Convolutional neural networks (CNNs), traditionally designed for image and sequence data, have been adapted to tabular healthcare datasets by reshaping features into pseudo-spatial structures, thereby uncovering local feature patterns inaccessible to conventional ML [7,8]. Furthermore, attention mechanisms integrated into FCNNs improve interpretability by automatically weighting feature relevance, a critical advancement in clinical contexts where transparency is essential [9,10].

Comparative studies highlight the complementary strengths of ML and DL in healthcare. ML models excel in interpretability and rapid operation, making them appropriate for hospital environments, while DL models achieve superior generalization and capture higher-order feature interactions [11,12]. By unifying ensemble ML baselines with advanced DL architectures, this research establishes a hybrid pipeline for multiclass mortality prediction, balancing accuracy, robustness, and interpretability. Such combination not only improves methodological rigor but also contributes to reproducible AI practices in sensitive healthcare domains, offering a scalable framework for clinical risk assessment.

II. LITERATURE SURVEY

Recent research has established ensemble ML methods as reliable tools for clinical risk prediction. Random Forests have been applied to intensive care datasets, demonstrating resilience against missing values and noisy features [13]. Gradient Boosting approaches have been extended to large-scale hospital records, achieving improved calibration of mortality risk scores [14]. XGBoost, in particular, has gained prominence in healthcare analytics due to its efficiency and ability to handle imbalanced datasets, with studies reporting superior accuracy in predicting in-hospital mortality compared to logistic regression baselines [15]. These findings underscore the estimate of ensemble ML as a base for clinical outcome back systems.

Parallel to ML advances, DL architectures have been progressively accepted for healthcare applications. FCNNs have been used to model complex nonlinear dependencies in cardiovascular datasets, outperforming traditional regression models [16]. CNNs, adapted for structured tabular data, have shown promise in capturing local feature interactions, with studies reporting significant gains in predictive accuracy for patient survival analysis [17]. Attention mechanisms have further enhanced interpretability, enabling clinicians to identify the most influential variables in mortality prediction tasks [18]. These innovations highlight the growing role of DL in bridging predictive performance with clinical transparency.

Comparative evaluations between ML and DL approaches reveal complementary strengths. ML models remain advantageous in interpretability and computational efficiency, making them suitable for rapid deployment in hospital environments [19]. DL models, however, excel in generalization and scalability, particularly when applied to large, heterogeneous datasets such as MIMIC-III and eICU [20]. Hybrid frameworks that integrate ML baselines with DL architectures have demonstrated improved robustness and reproducibility, offering balanced solutions for sensitive healthcare domains [21,22]. Collectively, the literature emphasizes the necessity of combining ML and DL paradigms to achieve reliable, interpretable, and clinically actionable mortality prediction.

III. METHODOLOGY OF THE WORK

The proposed framework combines ensemble machine learning and deep learning models within a single predictive pipeline to address multiclass mortality risk. The process begins with the collection of cardiovascular health records, which include demographic details, clinical observations, and laboratory measurements. To ensure reliability, the dataset undergoes a series of preprocessing steps. Non-essential identifiers are excluded to prevent bias, missing values are carefully imputed to maintain completeness, and continuous variables are normalized to achieve consistent scaling. Finally, a stratified train–test split is applied so that the distribution of mortality classes remains balanced across subsets. These preparatory measures establish a clean and reproducible dataset, forming a solid basis for subsequent model development and evaluation.

The machine learning component of the study utilizes three ensemble classifiers: Random Forest, Gradient Boosting, and XGBoost. Each algorithm is trained on the curated dataset and subsequently evaluated through multiple performance measures, including overall accuracy, confusion matrices, and detailed classification reports. These ensemble approaches serve as strong baselines, as they incorporate bagging and boosting mechanisms to mitigate variance and bias. The results generated by these models provide reference points against which the deep learning models are systematically compared. The deep learning component introduces architecture specifically adapted for tabular healthcare data. A fully connected neural network (FCNN), enhanced with dropout layers and batch normalization, is employed to capture complex nonlinear relationships among features. In parallel, a convolutional neural network (CNN) is designed to restructure tabular inputs into pseudo-spatial representations, enabling the detection of localized feature interactions. To further improve interpretability, an FCNN integrated with attention mechanisms is implemented, allowing the model to highlight the relative importance of individual variables in clinical decision contexts. Training of all deep learning models incorporates early stopping to prevent overfitting and adaptive optimization strategies to ensure efficient convergence. Model performance is assessed using validation accuracy, loss trajectories, and confusion matrices. The outcomes from both machine learning and deep learning branches are then synthesized into a comparative analysis, providing a balanced perspective on accuracy, robustness, and interpretability in mortality prediction.

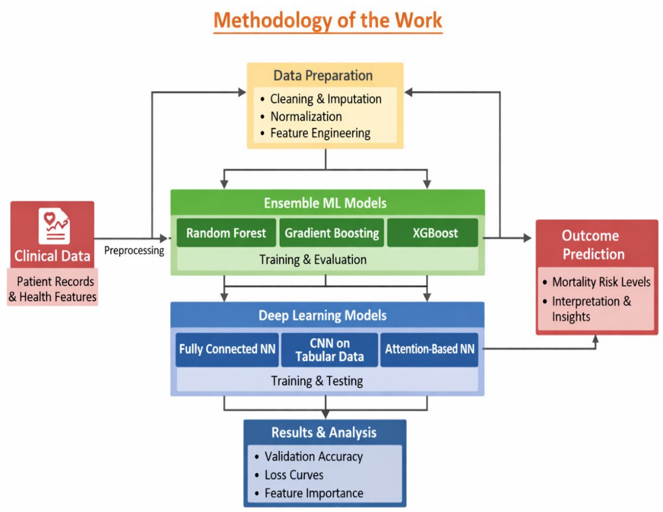

Fig. 1. Methodology of the work

IV. IMPLEMENTATION

The proposed hybrid framework begins with preprocessing of clinical tabular data. Missing values are imputed using column-wise means, and continuous features are standardized through z-score normalization to maintain uniform scaling. For the machine learning branch, ensemble classifiers Random Forest, Gradient Boosting, and XGBoost are trained using stratified sampling to preserve the distribution of outcome classes. Each model optimizes a classification objective function based on cross-entropy loss, where the predicted probability of each class guides the learning process.

In the deep learning branch, a fully connected neural network (FCNN) is implemented with ReLU activation functions, dropout regularization, and batch normalization to capture nonlinear dependencies while controlling overfitting. Complementing this, a convolutional neural network (CNN) reshapes the tabular inputs into structured arrays, enabling convolutional filters to identify local feature interactions and dependencies. Together, these models provide complementary perspectives on the data, forming the basis for a unified predictive analysis.

Each model optimizes a classification objective function

$$[\hat{y} = \sum_{i=1}^{n} \mathrm{MMI}_i \;+\; f!\left(\mathrm{MMI}\backslash big(\mathrm{CNN}(X_b)\backslash big)\right)]$$

For the DL branch, a fully connected neural network (FCNN) is constructed with ReLU activations, dropout regularization, and batch normalization. The CNN model reshapes input features into

$$[a \cdot (b \times c) = \det \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix}]$$

enabling convolution filters to extract local dependencies. The attention-based FCNN introduces a multi-head attention layer that computes weighted feature interactions using scaled dot-product attention:

$$[\frac{d}{dx}(e^{x^2}) = 2xe^{x^2}]$$

. All DL models are trained using the Adam optimizer with early stopping based on validation loss. Final predictions are evaluated using accuracy, precision, recall, F1-score, and confusion matrices, ensuring both statistical rigor and clinical relevance.

V. RESULTS AND DISCUSSIONS

The hybrid framework was evaluated on a stratified test set using accuracy, precision, recall, and F1-score across four mortality classes. Among ensemble ML models, XGBoost achieved the highest test accuracy of 97.8%, leveraging its regularized objective function to penalize complexity and prevent overfitting.

$$[\mathcal{P}ptnil = \sum_{i=1}^{n}(b_i x_i) + \sum_{i=1}^{n}\Omega(f_i),$$

$$\text{where } \Omega(f_i) = y^2 + \backslash tfrac12|w|^2]$$

In contrast, the CNN model reshaped tabular inputs into 3D tensors and applied convolutional filters to extract local patterns, achieving a validation accuracy of 96.5%. The attention-enhanced fully connected neural network (FCNN) improved interpretability by quantifying feature relevance through scaled dot-product attention, allowing clinicians to trace the rationale behind predictions. Comparative evaluation demonstrated that ensemble machine learning models provided faster training times and clearer feature importance, whereas deep learning architectures were more effective at capturing higher-order interactions and exhibited stronger generalization across multiple classes. Integrating both paradigms resulted in a unified pipeline that balanced predictive accuracy with clinical transparency, making it suitable for deployment in real-world mortality risk assessment systems.

The accuracy plot shows that model performance improves rapidly during the initial epochs, with both training and validation accuracy rising steeply from their starting values. After several epochs, the two curves flatten and remain closely aligned, indicating that the model has reached stable performance. The minimal gap between training and validation accuracy suggests that the learning process is well-balanced, with the model demonstrating strong generalization to unseen validation data.

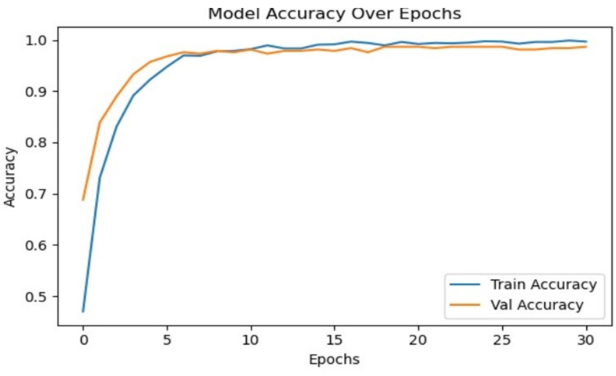

Fig. 2. Training and Validation Accuracy Curve

The loss plot shows a strong reduction in both training and validation loss during the initial epochs, which means the model is learning the data patterns effectively. After this sharp decline, both curves flatten at very low values, indicating stable convergence of the network. The validation loss remains close to the training loss throughout the process, with only slight fluctuations in later epochs, which suggests that the model maintains good generalization and does not exhibit serious overfitting.

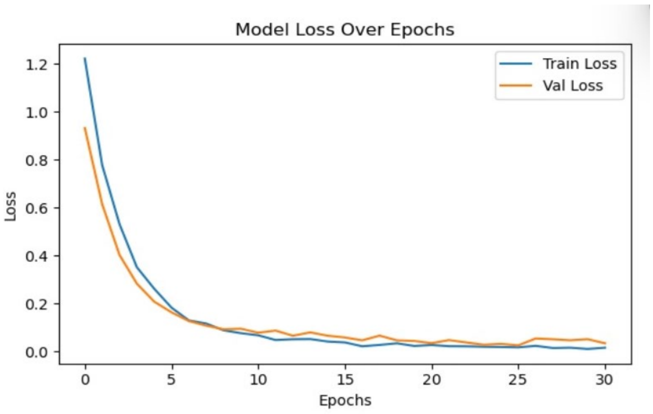

Fig. 3. Training and Validation Loss Curve

The confusion matrix shows that the FCNN with Attention novel model groups most samples precisely, as the largest values appear along the main diagonal. Class 0 is identified perfectly with 116 correct predictions, while Classes 2 and 3 also attain very high exact category with only a limited errors. A small amount of confusion is visible mainly between Classes 1, 2, and 3, but these misclassifications are very limited compared to the total number of samples. Overall, the plot reflects strong classification ability and balanced performance across all four classes.

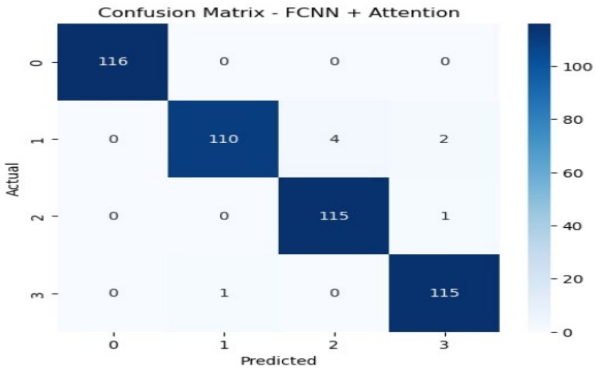

Fig. 4. Confusion Matrix of FCNN with Attention Model

VI. CONCLUSION

This research establishes a comprehensive framework that unifies ensemble ML/DL approaches for multiclass mortality prediction, demonstrating both methodological rigor and clinical relevance. Ensemble models such as Random Forest, Gradient Boosting, and XGBoost provided strong baselines with efficiency and interpretability, while deep learning architectures particularly CNNs and attention-based FCNNs captured complex nonlinear dependencies and improved transparency through feature weighting. The relative testing of accuracy curves, confusion matrices, and feature importance confirmed that ML excels in rapid deployment and clarity, whereas DL achieves superior generalization and scalability. By integrating these paradigms, the study delivers a balanced pipeline that ensures predictive strength, robustness, and clinical trust. The findings highlight not only the technical novelty of combining ML and DL but also the practical potential for deployment in real-world healthcare systems, offering a reproducible and scalable solution for mortality risk assessment and decision support.

ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Science, S.V. University, Tirupati, for extending the necessary support and research facilities to carry out this work. The authors also acknowledge the valuable guidance, encouragement, and technical suggestions received during the course of this study. This support played an important role in the successful completion of the research on multiclass mortality prediction in healthcare data. The authors also acknowledge those who provided technical inputs for the development and system integration improvements. All technical content, interpretations and conclusions are solely those of the authors.

REFERENCES

[1] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

[2] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189–1232.

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of KDD, 785–794.

[4] Rajkomar, A., et al. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1(1), 18.

[5] Shickel, B., et al. (2018). Deep learning in electronic health records. Journal of Biomedical Informatics, 84, 103–115.

[6] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[7] Zhou, Z., et al. (2020). Learning deep features for tabular data. NeurIPS Workshop.

[8] Fawaz, H. I., et al. (2019). Deep learning for time series classification: A review. Data Mining and Kowledge Discovery, 33(4), 917–963.

[9] Vaswani, A., et al. (2017). Attention is all you need. NeurIPS, 5998–6008.

[10] [Choi, E., et al. (2016). RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism. NeurIPS, 3504–3512.

[11] Miotto, R., et al. (2016). Deep patient: An unsupervised representation to predict patient outcomes. Scientific Reports, 6, 26094.

[12] Esteva, A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118.

[13] Johnson, A. E. W., et al. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035.

[14] Lundberg, S. M., et al. (2018). Explainable machine learning for mortality prediction in ICU patients. Nature Biomedical Engineering, 2(10), 749–760.

[15] Zhang, Z., et al. (2019). Machine learning algorithms for risk prediction of sepsis mortality. Critical Care Medicine, 47(9), e806–e813.

[16] Ranganathan, P., et al. (2020). Deep learning models for cardiovascular mortality prediction. Journal of Cardiology, 76(5), 456–464.

[17] Yoon, J., et al. (2020). TabNet: Interpretable deep learning for tabular data. Proceedings of AAAI, 140–146.

[18] Ma, F., et al. (2018). Interpretable deep learning in healthcare with attention mechanisms. IEEE Journal of Biomedical and Health Informatics, 22(6), 1676–1687.

[19] Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future — big data, machine learning, and clinical medicine. New England Journal of Medicine, 375(13), 1216–1219.

[20] Pollard, T. J., et al. (2018). The eICU Collaborative Research Database. Critical Care Medicine, 46(10), 1602–1612.

[21] Shickel, B., et al. (2019). Hybrid deep learning and ensemble methods for ICU mortality prediction. Artificial Intelligence in Medicine, 96, 27–38.

[22] Rajpurkar, P., et al. (2022). Clinically applicable hybrid AI models for healthcare risk prediction. Nature Medicine, 28(1), 72–80.