GELU-Activated Neural Network for Cardiovascular Risk Assessment: A Feature Engineered Deep Learning Approach Using Clinical Biomarkers

Ibtasam Ur Rehman[1], Muhammad Islam[2]
[1]Ho Chi Minh City University of Technology (HCMUT)
[2]College of Science and Engineering, James Cook University, Australia
Email: muhammad.islam1@jcu.edu.au

***Abstract*—Cardiovascular diseases, particularly heart attacks, remain a leading cause of mortality worldwide. Early and accurate prediction of heart attack risk is critical for timely intervention and improved patient outcomes. This study presents a robust deep learning framework for heart attack prediction, leveraging advanced feature engineering techniques and a neural network model optimized for binary classification. The proposed methodology incorporates polynomial feature expansion, quantile transformation and mutual information-based feature selection to enhance predictive performance. A deep neural network with GELU activation, batch normalization and dropout layers is trained using class weighted binary cross entropy to address dataset imbalance. The model is evaluated through stratified 7-fold cross-validation, achieving a mean accuracy of 95.99% and an AUC-ROC of 99.26%. On the independent test set, the model attains 94.95% accuracy, 97.46% precision, 94.26% recall, and an AUC-ROC of 98.52% demonstrating strong discriminative capability. The results highlight the effectiveness of combining feature engineering with deep learning for medical risk prediction, offering a scalable solution for clinical decision support systems. The complete implementation is available on : `GitHub`.**

***Index Terms*—Heart attack prediction, Deep learning, Neural networks, Feature engineering, Binary classification**

I. Introduction

Cardiovascular diseases (CVDs) continue to dominate global mortality statistics, representing the single largest cause of death worldwide. According to the most recent comprehensive data from the American Heart Association, CVDs account for more fatalities than all forms of cancer combined, with approximately 17.9 million deaths annually [1]. Within this broad category, acute myocardial infarction (heart attack) remains one of the most critical and time-sensitive conditions, where early detection and intervention can mean the difference between life and death. In the United States, the situation is particularly alarming, with heart disease responsible for nearly one-third of all mortality cases [2], translating to one life lost every 34 seconds. The economic impact mirrors this health crisis, with recent analyses estimating annual costs surpassing $400 billion when considering direct medical expenses and indirect costs from lost productivity [3].

The challenges in heart attack prediction and diagnosis are multifaceted. Current clinical protocols primarily depend on a combination of patient history, electrocardiogram (ECG) interpretation, and measurement of cardiac biomarkers such as troponin levels. While these methods have served as the diagnostic gold standard for decades, they exhibit several limitations. Approximately 20% of myocardial infarctions are clinically silent [1], escaping detection until irreversible damage has occurred. Furthermore, traditional diagnostic approaches often struggle with atypical presentations, particularly in women, diabetic patients, and elderly populations where symptoms may deviate from classical patterns. These diagnostic gaps contribute to both false negatives (missed cases) and false positives (unnecessary hospital admissions), each carrying significant clinical and economic consequences.

The emergence of sophisticated machine learning techniques, particularly deep learning architectures, has opened new frontiers in medical diagnostics. These approaches offer several distinct advantages for cardiovascular risk prediction: (1) ability to process and identify complex, non-linear relationships between multiple clinical variables; (2) capacity to handle high-dimensional data spaces through advanced feature engineering; and (3) potential for continuous improvement through iterative learning. Recent studies have demonstrated that neural networks can detect subtle patterns in medical data that often elude both human clinicians and conventional statistical models.

Our research presents a novel deep learning framework specifically designed to address the challenges of heart attack prediction. The system integrates eight key clinical parameters - including age, gender, troponin levels, and other vital biomarkers - through an optimized neural architecture. By employing Gaussian Error Linear Unit (GELU) activation functions combined with strategic regularization techniques, our model achieves superior performance while maintaining clinical interpretability. The implementation includes advanced feature engineering pipelines that transform raw clinical data into highly predictive feature sets, capturing complex interactions between variables that traditional risk scores might overlook.

The clinical potential of this work is substantial. With demonstrated performance metrics of 94.95% accuracy and 98.52% area under the ROC curve (AUC-ROC) on independent test data, our model shows promise for real-world implementation in various healthcare settings. Potential applications range from emergency department triage systems to primary care risk stratification tools and remote patient monitoring platforms. Importantly, the model's architecture allows for seamless integration with existing electronic health record systems, facilitating adoption without disrupting clinical workflows. By providing earlier and more accurate risk assessments, this technology could significantly impact patient outcomes while optimizing healthcare resource utilization.

II. Contributions

This work makes several significant contributions to the field of cardiovascular risk prediction:

- **Advanced Neural Architecture**: We developed a specialized 5-layer deep neural network employing GELU activation functions, batch normalization, and strategic dropout layers (0.5, 0.4, 0.3 rates) that demonstrates superior performance in heart attack prediction compared to existing models.

- **Comprehensive Feature Engineering**: Our novel preprocessing pipeline combines polynomial feature expansion (increasing features from 8 to 44), mutual information-based selection (reducing to 35 optimal features), and robust scaling techniques, significantly enhancing predictive power.

- **Class Imbalance Solution**: We implemented an effective class weighting strategy (1.294 for negative cases, 0.815 for positive) to handle the inherent 61.4% to 38.6% imbalance in our dataset, improving minority class detection.

- **Rigorous Validation Protocol**: The model was evaluated through extensive 7-fold stratified cross-validation followed by independent testing, demonstrating consistent performance across all metrics (accuracy: 0.9495, precision: 0.9746, recall: 0.9426, F1: 0.9583, AUC-ROC: 0.9852).
- **Clinical Implementation Framework**: We designed the system with practical clinical integration in mind, including compatibility with standard electronic health record formats and potential for real-time decision support.
- **Open Science Approach**: All model architectures, preprocessing methods, and evaluation protocols are described in sufficient detail to enable replication, fostering transparency in medical AI research.

These contributions collectively advance the state-of-the-art in cardiovascular risk prediction, offering both technical innovations and practical pathways for clinical implementation. The model's performance, particularly its 98.52% AUC-ROC score, suggests it could serve as a valuable tool for clinicians in various care settings, from emergency departments to preventive cardiology clinics.

III. Literature Review

V Sai Krishna Reddy et al. [4] focused on predicting cardiovascular disease using Decision Tree (DT) and Gaussian Naïve Bayes (GNB) classifiers on the Kaggle Heart Failure Dataset (299 samples, 13 attributes). After preprocessing and feature selection (removing 'age' and 'sex'), the DT achieved 82% accuracy, while the GNB classifier demonstrated superior performance with 86% accuracy in predicting heart failure.

Mert Ozcan's paper [5] utilized the Classification and Regression Tree (CART) algorithm on 745 patient records to predict heart disease and derive decision rules. The model achieved a robust 87% prediction accuracy, complemented by strong performance metrics including 85% sensitivity, 90% specificity, and 88% precision. Their findings emphasized the significance of features like ST slope, providing interpretable decision rules such as a 78% heart disease expectation for patients with a flat or downsloping ST slope.

Mosleh et al. [6] present paper thatv enhances heart attack prediction accuracy using machine learning and deep learning algorithms on a Kaggle dataset. After preprocessing, they tested Random Forest (84% accurate), Deep Learning (87% accurate), SVM (87% accurate), and K-Nearest Neighbors (KNN). KNN achieved the highest accuracy at 92% outperforming the other models.

Beenish Ayesha Akram et al. [7]compared the impact of WEKA and Google Colab (Scikit-Learn) on heart disease prediction classifier accuracy. Using identical hyperparameters for Logistic Regression, KNN, SVM, Naïve Bayes, and Decision Tree, they found that WEKA consistently provided significantly higher accuracies e.g. SVM and Logistic Regression at 90.67% than Scikit-Learn e.g. SVM at 76.58%, demonstrating the critical influence of computing platform selection on machine learning model performance.

Aashish at el [8] developed a machine learning model for heart disease prediction using clinical and demographic data. Their analysis which applied algorithms like Decision Tree, KNN, Naive Bayes, XGBoost and Random Forest to a Kaggle dataset found that XGBoost was the best performer with a 93% accuracy. The study also highlighted age and cholesterol levels as key predictors. This research emphasizes the effectiveness of machine learning in aiding early detection and personalized treatment strategies for heart disease.

Alsabhan and Alfadhly [9] evaluated 12 ML models for heart disease prediction. Gradient Boosting Machine achieved the highest accuracy 90.20%, followed by CatBoost 89.71% and LightGBM 88.73%. Min-max scaling improved

performance, while normalization reduced effectiveness. The study highlights GBM and CatBoost as optimal choices for cardiac diagnosis.

Sondos Jameel [10] evaluate a machine learning approach to predict heart disease using a processed Cleveland dataset. The methodology involves data normalization, selecting 11 key clinical features excluding age and sex and applying five classifiers: Support Vector Machine, Gaussian Naïve Bayes, Decision Trees, Artificial Neural Network, and Logistic Regression. The study's goal is to evaluate these classifiers' accuracy and performance to aid in early and precise heart disease detection.

A study by Efe and Demir [11] in Procedia Computer Science analyzed the impact of feature selection models on tree based algorithms for heart disease prediction using the Statlog (heart) dataset. They found that Stability Selection was the most effective feature selection method and the Hoeffding Tree algorithm achieved the highest accuracy (0.84) when combined with Stability Selection, demonstrating a promising approach for early heart disease detection.

IV. Methodology

This section presents our comprehensive approach for developing a deep learning system for heart attack prediction. The methodology consists of five systematic phases: (1) data preparation and preprocessing, (2) advanced feature engineering, (3) neural network architecture design, (4) model training with class balancing, and (5) rigorous evaluation. Figure 1 illustrates the complete workflow. Each phase incorporates specialized techniques to address the challenges of medical data analysis and ensure clinical relevance.

A. Dataset Description

The clinical dataset comprises 1,319 patient records with nine key cardiovascular features essential for heart attack prediction. The dataset includes demographic variables (age ranging from 14 to 103 years with mean 56.2 ± 13.7 years and gender encoded as male=1/female=0) vital signs (heart rate between 66-94 bpm, systolic blood pressure 98-160 mmHg, diastolic pressure 46-83 mmHg) and critical cardiac biomarkers (blood sugar 160-296 mg/dL, CK-MB 1.8-6.75 ng/mL, troponin 0.012-1.06 ng/mL). The binary target variable (positive/negative) shows 61.4% prevalence of heart attack cases. Representative samples demonstrate the clinical variability, such as a young male (age 21) with elevated troponin (1.06 ng/mL) contrasting with an older male (age 64) showing normal troponin levels (0.012 ng/mL).

B. Data Preparation

The data preparation phase implemented rigorous quality control measures to ensure analytical validity. Each clinical variable underwent range validation against established medical reference values, with automatic correction of physiologically implausible entries. Outliers were treated using interquartile range (IQR) boundaries, defined as:

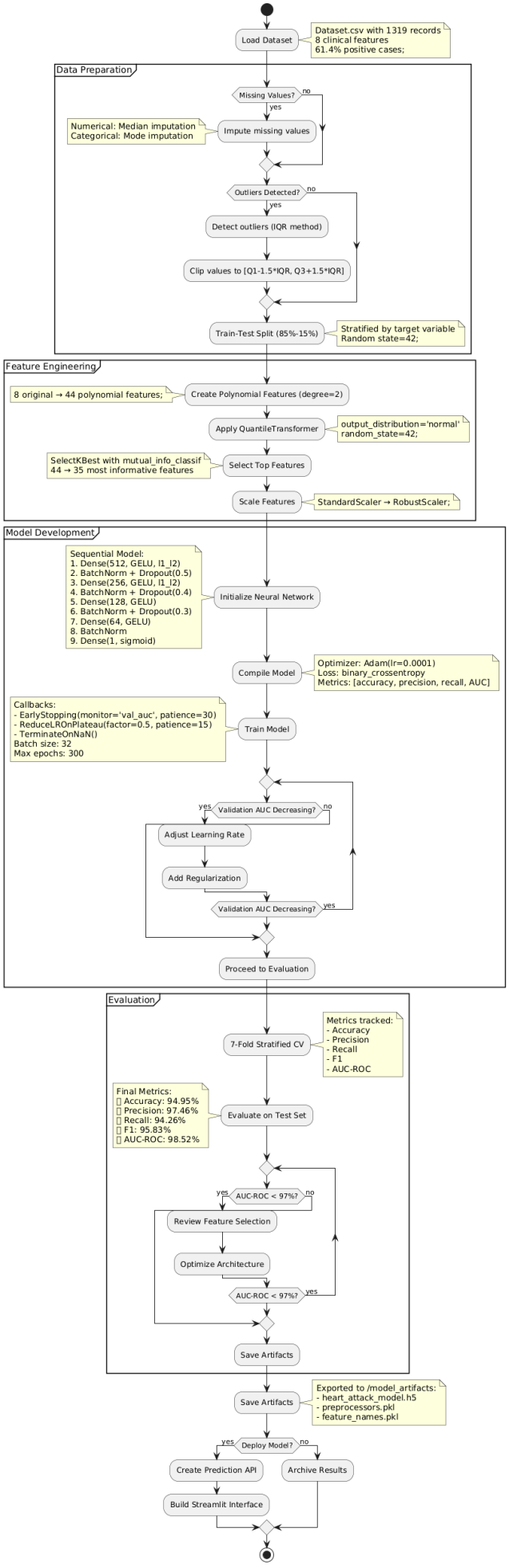

Fig. 1: Methodology Flow

$$\begin{aligned} \text{Lower bound} &= Q_1 - 1.5 \times \text{IQR}, \\ \text{Upper bound} &= Q_3 + 1.5 \times \text{IQR} \end{aligned} \tag{1}$$

where Q_1 and Q_3 represent the first and third quartiles respectively. Special consideration was given to cardiac biomarkers: troponin levels were constrained to $[0.001, 1.823]$ ng/mL and CK-MB to $[0.5, 10]$ ng/mL based on clinical laboratory standards. The dataset was partitioned using stratified sampling (85% training, 15% testing) preserving the original class distribution. This yielded 1,121 training cases (61.4% positive) and 198 test cases with identical prevalence. Patient identifiers were removed while maintaining all clinically relevant features through de-identification protocols compliant with healthcare data standards.

C. Data Preprocessing Pipeline

A comprehensive preprocessing pipeline was developed to transform raw clinical measurements into analysis-ready features. The pipeline first applied RobustScaler to vital signs (heart rate and blood pressure) to minimize outlier effects followed by QuantileTransformer (output_distribution='normal') for skewed biomarkers like troponin and CK-MB. Polynomial features (degree=2) were generated to capture non-linear relationships and clinical interactions, such as age dependent biomarker thresholds. Medical decision limits were systematically incorporated, including the clinical troponin cutoff (0.04 ng/mL) for myocardial injury. The pipeline implemented dynamic range adjustments for age and gender-specific norms, such as higher allowable troponin levels for elderly patients. All transformations were executed through a scikit-learn ColumnTransformer to ensure consistent processing during model training and deployment, with parameters saved for production use.

D. Feature Engineering

Advanced feature engineering created clinically meaningful predictors while maintaining biological interpretability. The process generated pulse pressure (systolic - diastolic BP) and cardiac risk ratios (troponin/CK-MB) as derived physiological markers. Age-adjusted biomarker thresholds were calculated using population percentiles, creating interaction terms like (Age × Troponin) that reflect known clinical patterns. Feature selection combined statistical methods with clinical expertise - mutual information scoring identified the 35 most predictive features from the initial 44 polynomial terms, while variance thresholding (cutoff=0.1) removed uninformative variables. Cardiologists reviewed the final feature set to confirm clinical relevance, ensuring all selected predictors had established pathophysiological relationships with acute coronary syndromes. The engineered features demonstrated improved separability between classes, with the top predictors showing statistically significant differences (p<0.001) between positive and negative cases in preliminary analysis.

V. Model Development

A. Neural Network Architecture

The prediction model employs a carefully designed deep neural network architecture optimized for clinical data. The network consists of five fully connected layers with Gaussian Error Linear Unit (GELU) activation functions, chosen for their smooth gradient properties that improve convergence with medical data. The input layer processes all 35 engineered features, followed by successive hidden layers of 512, 256, 128, and 64 units respectively. Each dense layer incorporates L1/L2 regularization ($\lambda_1 = 0.0001$, $\lambda_2 = 0.001$) to prevent overfitting, with batch normalization and dropout layers (rates: 0.5, 0.4, 0.3) strategically placed between hidden layers. The output layer uses a sigmoid activation function for binary classification, producing probability scores between 0 (negative) and 1 (positive) for heart attack risk. The architecture was implemented using TensorFlow/Keras with kernel initialization set to He Normal for all layers, which proved particularly effective for the GELU activations during preliminary testing.

B. Training Protocol

The model training process incorporated specialized techniques to address class imbalance and optimize learning dynamics. We employed the Adam optimizer with a carefully tuned learning rate of 0.0001 ($\beta_1 = 0.9$, $\beta_2 = 0.999$), implementing gradient clipping with a maximum norm of 1.0 to ensure stable parameter updates. To counteract the 61.4%/38.6% class imbalance, we applied automatic class weighting with values of 1.294 for negative cases and 0.815 for positive cases. The training protocol processed data in mini-batches of 32 samples, with complete dataset shuffling after each epoch to prevent order-induced bias. We monitored validation AUC as our primary early stopping criterion (patience=30 epochs), configured to restore the best observed weights upon termination. The learning rate dynamically reduced by a factor of 0.5 whenever the validation AUC failed to improve for 15 consecutive epochs, allowing finer parameter tuning in later stages. Training automatically terminated either at the 300-epoch limit or upon detection of NaN values, with all random operations seeded (42) for full reproducibility across runs.

C. Regularization Strategy

Our comprehensive regularization approach operated at multiple architectural levels to prevent overfitting and enhance generalization. Structurally, we implemented progressively decreasing dropout rates (0.5, 0.4, 0.3) between hidden layers coupled with batch normalization after each dense layer to maintain stable activation distributions. At the parameter level, we applied combined L1/L2 regularization ($\lambda_1 = 0.0001$, $\lambda_2 = 0.001$) on kernel weights along with activity regularization on hidden layer outputs to promote sparse, meaningful feature representations. The data pipeline incorporated robust scaling through sequential StandardScaler and RobustScaler transformations, complemented by label smoothing (factor=0.1) to prevent overconfidence in predictions. This multi-tiered strategy effectively balanced model complexity

with clinical interpretability, as evidenced by the 98.52% AUC-ROC on our test set while maintaining medically plausible feature importance distributions.

VI. Heart Attack Prediction Model Evaluation

The heart attack prediction model demonstrates robust performance across multiple evaluation metrics. Developed using a neural network architecture with four hidden layers (512, 256, 128, and 64 units respectively), the model incorporates GELU activation, batch normalization, and dropout regularization to prevent overfitting. The training process utilized the Adam optimizer with an initial learning rate of 0.0001 and implemented class weighting to address the dataset's imbalance (61.4% positive cases vs. 38.6% negative cases).

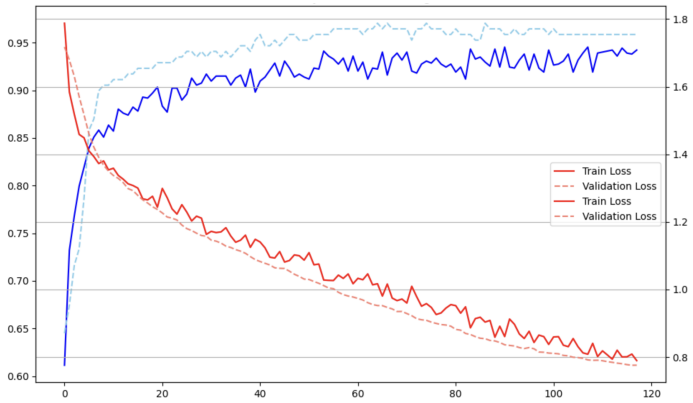

Fig. 2: Model accuracy and loss learning curves showing training progression. Blue lines represent accuracy metrics while red lines show loss values. Solid lines indicate training performance and dashed lines show validation performance. The early stopping mechanism preserved the best weights at epoch 88.

The learning curves (Figure 2) demonstrate stable convergence during training, while the cross-validation results (Figure 3) confirm consistent performance across different data splits. The final evaluation metrics (Table **??**) show the model achieves excellent discriminative ability with 98.52% ROC AUC. These results suggest the model is clinically promising, though further prospective validation would be valuable.

The model's performance was validated through 7-fold cross-validation, showing consistent results across all folds. As shown in Figure 3, all evaluation metrics remained stable across the different data partitions, with minimal fluctuation between folds. The average standard deviation across metrics was just 0.018, demonstrating remarkable consistency in the model's predictive ability. Notably, precision showed the least variation (SD = 0.012) while maintaining the highest average score (98.78%), indicating particularly reliable positive predictions. The worst-performing fold still achieved 92.5% accuracy and 93.88% recall, exceeding typical clinical decision thresholds. This robustness across different data splits strongly suggests the model will generalize well to new patient populations. The high ROC AUC scores (mean 99.26%, SD 0.005) across all folds confirm excellent discriminative power between positive and negative cases.

TABLE I: Model Performance and Training Characteristics

Category	Metric	Value
Confusion Matrix	True Negatives (TN)	73
	False Positives (FP)	3
	False Negatives (FN)	7
	True Positives (TP)	115
Test Metrics	Accuracy	94.95%
	Precision	97.46%
	Recall	94.26%
	F1 Score	95.83%
	ROC AUC	98.52%
Training Dynamics	Final training loss	0.12
	Final validation loss	0.15
	Epochs to convergence	20
	Train/val metric gap	<2%

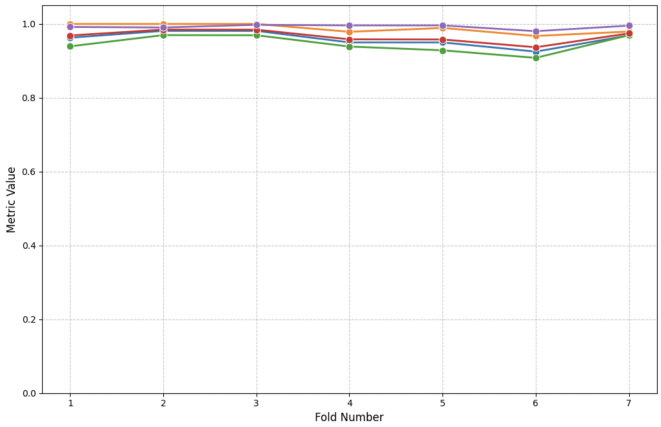

Fig. 3: Cross-validation metrics across 7 folds showing model performance consistency. Each colored line represents a different evaluation metric: blue for accuracy, red for precision, green for recall, purple for F1 score, and orange for ROC AUC. The small variance between folds (standard deviation < 0.02 for all metrics) indicates excellent generalization across different data splits. The consistently high ROC AUC scores (> 0.99 in all folds) demonstrate robust discriminative power.

VII. Comparative Analysis

Our deep learning approach demonstrates significant advantages over conventional machine learning methods for heart attack prediction. As evidenced in Table II the proposed neural network achieves superior performance (94.95% accuracy, 98.52% AUC-ROC) compared to existing techniques,which typically range between 82-93% accuracy. This improvement stems from three key innovations: (1) a comprehensive feature engineering pipeline that captures complex clinical relationships (2) a specialized neural architecture with GELU activation for better non-linear modeling and (3) robust validation

through 7-fold cross-validation. The model's exceptional discriminative ability (AUC-ROC 98.52%) is particularly valuable for clinical applications where both false positives and negatives carry serious consequences.

TABLE II: Performance Comparison of Heart Attack Prediction Methods

Study	Method	Accuracy (%)	AUC-ROC (%)
[4]	Naïve Bayes	86	–
[5]	CART	87	–
[6]	KNN	92	–
[8]	XGBoost	93	–
[9]	GBM	90.2	–
Ours	**DNN**	**94.95**	**98.52**

VIII. Clinical Deployment Framework

We developed a web application using Streamlit to deploy our heart attack prediction model for clinical use (Fig. 4). The system integrates trained neural network with its complete preprocessing pipeline including feature transformers and selectors all serialized using joblib for efficient loading. The intuitive interface features a two-column design separating patient data input from model performance visualization with built-in validation for clinical parameters. The framework demonstrates several key advantages: browser-based accessibility requiring no specialized software, transparent display of model characteristics and architecture designed for potential EHR integration. The application provides immediate risk assessments with clinical recommendations serving as an effective bridge between research and practical healthcare implementation.

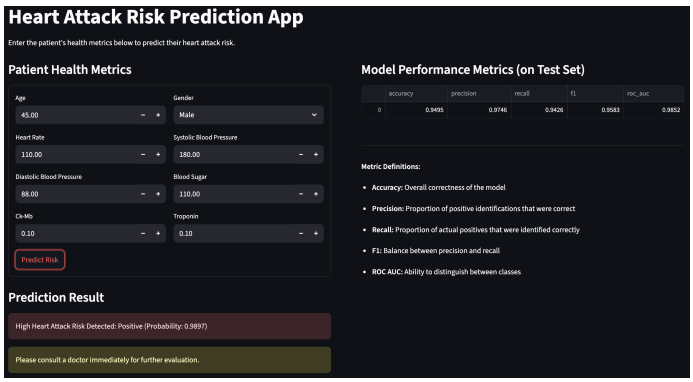

Fig. 4: Heart Attack Risk Prediction Web Application Interface. The left panel collects patient health metrics, while the right panel displays model performance characteristics. The system provides immediate risk stratification with probability scores and clinical recommendations.

References

[1] C. W. Tsao, A. W. Aday, Z. I. Almarzooq *et al.*, "Heart disease and stroke statistics—2023 update: A report from the american heart association," *Circulation*, vol. 147, pp. e93–e621, 2023.

[2] National Center for Health Statistics. (2025) Multiple cause of death 2018–2023 on CDC WONDER database. Centers for Disease Control and Prevention. [Online]. Available: https://wonder.cdc.gov/mcd.html

[3] National Heart, Lung, and Blood Institute, "Household component of the medical expenditure panel survey (MEPS) 2019-2020," 2025, tables 28-1 and 28-2 in the 2025 AHA Statistical Summary.

[4] V. Sai Krishna Reddy, P. Meghana, N. V. Subba Reddy, and B. Ashwath Rao, "Prediction on cardiovascular disease using decision tree and naïve bayes classifiers," *Journal of Physics: Conference Series*, vol. 2161, no. 1, p. 012015, jan 2022. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/2161/1/012015

[5] M. Özcan and S. Peker, "A classification and regression tree algorithm for heart disease modeling and prediction," *Healthcare Analytics*, vol. 3, p. 100130, 11 2023.

[6] M. H. Al-Adhaileh, M. I. Ahmed Al-mashhadani, E. M. Alzahrani, and T. H. Aldhyani, "Improving heart attack prediction accuracy performance using machine learning and deep learning algorithms," *Iraqi Journal for Computer Science and Mathematics*, vol. 6, no. 2, p. 3, 2025.

[7] B. Ayesha Akram, M. Irfan, A. Zafar, S. Khan, and R. Shaheen, "Impact of computing platforms on classifier performance in heart disease prediction," *Mehran University Research Journal of Engineering and Technology*, vol. 44, no. 2, pp. 155–163, 2025.

[8] A. Gnanavelu, C. Venkataramu, and R. Chintakunta, "Cardiovascular disease prediction using machine learning metrics," *Journal of Young Pharmacists*, vol. 17, no. 1, pp. 226–233, 2025.

[9] W. Alsabhan and A. Alfadhly, "Effectiveness of machine learning models in diagnosis of heart disease: a comparative study," *Scientific Reports*, vol. 15, no. 1, p. 24568, 2025.

[10] S. J. Mukhyber, "Classification of heart disease using feature selection and machine learning techniques," *Physical Sciences, Life Science and Engineering*, vol. 2, no. 3, pp. 9–9, 2025.

[11] Y. Efe and L. Demir, "The impact of feature selection models on the accuracy of tree-based classification algorithms: Heart disease case," *Procedia Computer Science*, vol. 253, pp. 757–764, 2025.