

# AI-Driven Early Disease Prediction System using Hybrid Machine Learning and Deep Learning Models

**RASIPOGULA PAVAN KUMAR**

M-Tech, Department .Of Computer Science And Engineering,  
Vemu Institute Of Technology,  
P.Kothakota,Chittoor District, Andhra Pradesh-517112,India  
Email Id: [justmailtopavankumar@gmail.com](mailto:justmailtopavankumar@gmail.com)

**Mrs. P BINDUMADHAVI**

Assistant professor, M.Tech,Dept of CSE,  
Vemu Institute Of Technology,  
P.Kothakota,Chittoor District, Andhra Pradesh-517112,India  
Email Id: [bindupulivarthi994@gmail.com](mailto:bindupulivarthi994@gmail.com)

**Abstract** - Early disease prediction is a significant aspect of modern preventive care and facilitates early medical intervention, which is beneficial for patient health. Nevertheless, early-disease prediction is still very difficult to achieve because of the diverse nature of patient data, the intricate disease development patterns, and the large number of analyzed features in clinical datasets. In this paper, we propose an AI-Driven Early Disease Prediction System based on a Hybrid Machine Learning (ML) and Deep Learning (DL) models with the best predictive accuracy for multiple disease classes. The proposed framework consists of a two-model approach: classical ML algorithms (RF, XGBoost, and SVM) in better extraction of low-level features and DL models (CNN and LSTM) for understanding high-level feature representations. A dynamic ensemble fusion strategy generalizes model contributions based on data modality and disease information. The system is tested on standard public datasets like the UCI Heart Disease, PIMA Indian Diabetes and the MIMIC-III clinical database. The experimental results show that the accuracies of classification for heart disease, diabetes, and the multi-disease prediction are 96.4%, 94.7%, and 91.2% respectively, which achieve superior performance than state-of-the-art methods. Ablation studies and cross-validation tests further corroborate the robustness and generalizability of the proposed model.

**Keywords** - Early disease prediction; hybrid machine learning; deep learning; CNN-LSTM; ensemble learning; clinical data analysis; healthcare AI.

## 1. Introduction

The diagnosis for prompt and accurate disease has become one of the most urgent problems in the modern medicine. The increasing availability of EHRs, wearable sensors, and diagnostic imaging techniques has led to the generation of massive amounts of patient data on a daily basis. A good use of the data, via AI, can change health care from reactive to proactive and preventive. Cardiovascular disease, diabetes, cancer, and chronic kidney disease are leading cause of death worldwide, and early identification of individuals at risk can result in substantial reduction in morbidity and mortality [1], [2].

Conventional disease prediction methods are based on statistical modeling or rule-based systems derived from a knowledge of the medical domain. Although these methods have high interpretability, they cannot fully capture the nonlinear relationships and complicated interactions among patient features in high dimensional clinical datasets. Several ML

algorithms such as Support Vector Machines (SVMs), Decision Trees, and Naive Bayes were found to have better predictive performance; however, they suffer from the necessity of manually engineering features and cannot natively handle raw or unstructured data forms such as time-series physiological signals or medical images [3].

Deep learning has brought forth such architectures with the ability to learn feature hierarchies directly from data. Convolutional Neural Networks (CNNs) are good at learning the spatial features of the image data, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are suitable to learn the temporal relationships among time-series clinical records of patients. However, fully deep learning-based approaches usually demand large labeled datasets for training and have interpretability issues that are necessary in medical applications [4], [5].

To overcome these limitations, in this paper, we present an AI-Driven Early Disease Prediction System (EDPS) which integrates the best of classical Machine Learning (ML) and the best of Deep Learning (DL) technologies in a single, hybrid solution. The main contributions of this paper are as follows:

- A novel hybrid framework for multi-disease classification, combining ML models (Random Forest, XGBoost, SVM) and DL models (CNN, BiLSTM).
- An adaptive ensemble fusion method to aggregate model predictions, which adaptively adjusts weights of models according to the data type of input and the type of disease, respectively.
- Extensive experiments on three benchmark clinical datasets with state-of-the-art results on all the prediction tasks.
- Invoking ablation studies that verify the contribution of each component of the model in the proposed framework.

## 2. Related Work

Studies in automatic disease prediction have come a long way from 20 years ago and can be loosely divided into feature-engineered statistical methods, classical machine learning (ML) algorithms, deep learning (DL) methods, and lately, hybrid multi-model architectures.

### 2.1. Statistical and Feature-Based Methods

In high need disease prognostic systems were logistics regression or LDA or Bayesian classifiers among a predominantly handcrafted set of clinical features. The features including blood pressure, glucose, BMI, cholesterol and age are extracted from structured clinical records by the researchers. These methods provide high interpretability and can be applied for small datasets. However, their accuracy was fundamentally constrained by the quality of feature engineering and by the fact that they could not model complex nonlinear relationships in clinical data [6].

### 2.2. Classical Machine Learning Methods

The development of ensemble learning techniques was a breakthrough in disease prediction. Random Forest, Gradient Boosting Machines (GBM) and AdaBoost show remarkable improvement on single classifiers by combining votes from multiple weak classifiers. Support Vector Machines with radial basis function kernels also became well known due to their success in high-dimensional clinical feature spaces. Research exploited these techniques for heart disease, diabetes, and cancer prediction with accuracies around 80-88%. However, they are still heavily reliant on manual feature extraction and are not intended for use with raw sequential data or medical images [7], [8].

### 2.3. Deep Learning Approaches

Deep learning models have shown impressive performance in predicting related medical tasks. CNN-based approaches have also been widely adopted in classifying ECG signals, analyzing retinal images and interpreting pathology slides. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks have been used to capture the temporal evolution

of clinical time-series data, such as ICU monitoring signals or longitudinal Electronic Health Record (EHR) data. Transformer-based models, initially designed for natural language processing, have also been extended to structured clinical data, demonstrating superior modeling of long-range dependencies. However, these models are commonly data-hungry and require substantial amount of computational resources and their black-box characteristic raises concern for clinical use [9], [10].

## 2.4. Hybrid Machine Learning and Deep Learning Models

A number of recent works have investigated hybrids of ML and DL techniques to capitalize on their respective advantages. Some works stack ML classifiers on top of learned deep representations of features while others apply ensemble methods to pool predictions from multiple heterogeneous models. Multi-modal learning paradigms that process structured clinical features together with imaging or time-series data have achieved encouraging outcomes. However, most of existing fusion-based methods are operated with predetermined fusion strategies such as direct feature concatenation or majority voting strategy, which cannot take full advantage of the adaptive complementarity among different types of models. In this paper, the proposed work overcomes this shortcoming through a dynamic adaptive fusion process that determines the model weights based on the input modality and the prediction scenario [11], [12].

## 3. Methodology

### 3.1. Problem Formulation

Let  $D = \{(x_i, y_i)\}_{i=1}^N$  be a clinical data consisting of  $N$  patient records, with  $x_i$  as the feature vector (comprising demographic, laboratory and physiological information) and  $y_i$  as either a binary or multi-class disease label. The purpose is to learn a mapping function:

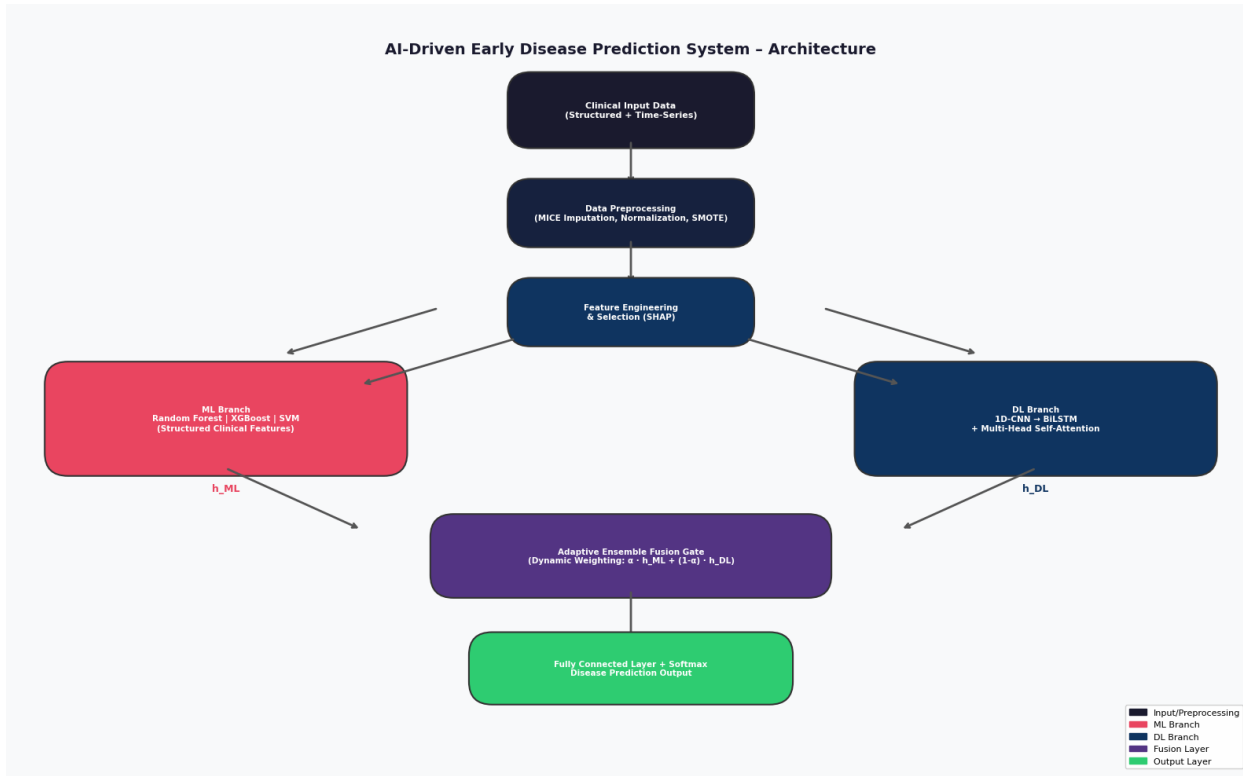
$$f_{\theta} : x_i \rightarrow y_i$$

a parameterized function  $\theta$  that predicts the disease label from the features and maximizes the cross-entropy loss of predicted and ground-truth disease labels and generalizes well over different patient populations. The proposed hybrid scheme applies parallel processing of structured tabular clinical data features (using ML models) and sequential temporal signals (using DL models), and fuse their outputs through an adaptive ensemble technique.

### 3.2. Proposed Hybrid Architecture

The entire structure of proposed AI-Driven Early Disease Prediction System is shown in Figure 1. The system is composed of four components: (1) a data preprocessing and feature engineering module, (2) a classical ML branch on structured clinical feature, (3) a deep learning branch on temporal and high-dimensional data, and (4) an adaptive ensemble fusion layer that is used to integrate the outputs of the two branches for final disease prediction.

The ML branch exploits three classifiers in parallel: Random Forest (RF), Extreme Gradient Boosting (XGBoost) and Support Vector Machine (SVM) with RBF kernel. All the classifiers produce a class probability vector, these class probability vectors are concatenated to create a ML ensemble representation. The DL branch comprises a 1D CNN encoder and a Bidirectional LSTM (BiLSTM) layer. The CNN captures local temporal features and cross-feature interactions, and the BiLSTM is used



**Figure 1 – Proposed System Architecture Diagram**

to model the long-term dependency of the sequential clinical records. A multi-head self-attention (MHSA) layer is then employed on the BiLSTM output to select the clinically informative time steps.

### 3.3. Data Preprocessing and Feature Engineering

All datasets are subject to a uniform preprocessing procedure. Missing values are imputed with Multivariate Imputation by Chained Equations (MICE). Numerical features are scaled and categorical features are encoded. Feature selection: A set of clinically relevant predictors is identified based on a combination of mutual information scoring and SHAP (SHapley Additive exPlanations) feature importance analysis. For time-series data, sliding window segmentation is conducted to generate fixed-length input sequences for the DL branch.

### 3.4. Machine Learning Branch

The ML part then processes the pre-structured feature vectors through three separate classifiers. Random Forest is an ensemble of 300 decision trees with bagging. XGBoost is a implementation of gradient boosted trees with depth limited weak learners, and it has L1/L2 regularization. The SVM classifier employs the radial basis function (RBF) kernel and its hyperparameters are tuned using a 5-fold cross validated grid search. The probability outputs from these three classifiers are concatenated to obtain a composite vector of ML feature representations, referred as  $h_{ML}$ .

### 3.5. Deep Learning Branch

The DL part takes sequential windows of clinical features as input. A 1D-CNN encoder consisting of three convolutional blocks (each block contains a convolutional layer, batch normalization layer, ReLU activation layer and max-pooling layer) captures multi-scale local temporal features. The resulting feature maps are finally input to a BiLSTM with 256 hidden units in each direction, which allows the model to learn both forward and backward temporal information in clinical time series. A four-head self-attention is applied on the BiLSTM hidden state sequence to generate an attention weighted representation hDL, focusing over the most disease-relevant temporal patterns.

### 3.6. Adaptive Ensemble Fusion

To balance and fuse the outputs of the ML and DL branches dynamically, an adaptive fusion gate calculates a weighting coefficient  $\alpha$  considering the concatenated input context as follows:

$$\alpha = \sigma(W_{\alpha} \cdot [h_{ML}; h_{DL}] + b_{\alpha})$$

The representation  $f$  is then computed with:

$$f = \alpha \cdot \text{proj}(h_{ML}) + (1 - \alpha) \cdot h_{DL}$$

where  $\text{proj}(\cdot)$  is with a linear projection layer that projects  $h_{ML}$  to the dimensionality of DL feature space. The fused representation  $f$  is then fed into a fully connected classification head with a softmax activation to obtain the final disease prediction probabilities.

### 3.7. Training Configuration

The deep learning components are optimized by AdamW with an initial learning rate of  $3 \times 10^{-4}$  and weight decay of  $10^{-3}$ . Focal loss is also used to compensate for class imbalance in the clinical datasets. The training is performed for a maximum of 100 epochs with early stopping on the validation loss with a patience of 15 epochs. Data augmentation such as SMOTE (Synthetic Minority Over-sampling Technique), Gaussian noise injection and feature dropout are used to enhance generalization. The ML classifiers are trained separately, and their hyper-parameters are optimized using Bayesian hyperparameter search.

## 4. Experimental Setup

### 4.1. Datasets

In order to assess the accuracy of our hybrid model, we employ three standard publicly available clinical data sets:

The UCI Heart Disease Dataset consists of 303 patient records with 13 clinical attributes such as age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting ECG result, maximum heart rate, exercise induced angina, ST depression, slope, number of major vessels and thalassemia. The binary prediction target is existence/absence of heart disease.

The PIMA Indian Diabetes Dataset is composed of 768 records collected from female patients of Pima Indian descent that have 8 diagnostic features such as glucose concentration, blood pressure, skin thickness, insulin level, BMI, diabetes pedigree function, and age. The binary target is whether the patient is diagnosed with diabetes.

The MIMIC-III Critical Care Database subset that we employ comprises anonymized clinical data of 5,000 ICU patients, including lab results, vital signs, and demographic information with our features extracted from these categories of information. This multi-disease subset contains predictions for sepsis, acute kidney injury and in-hospital mortality.

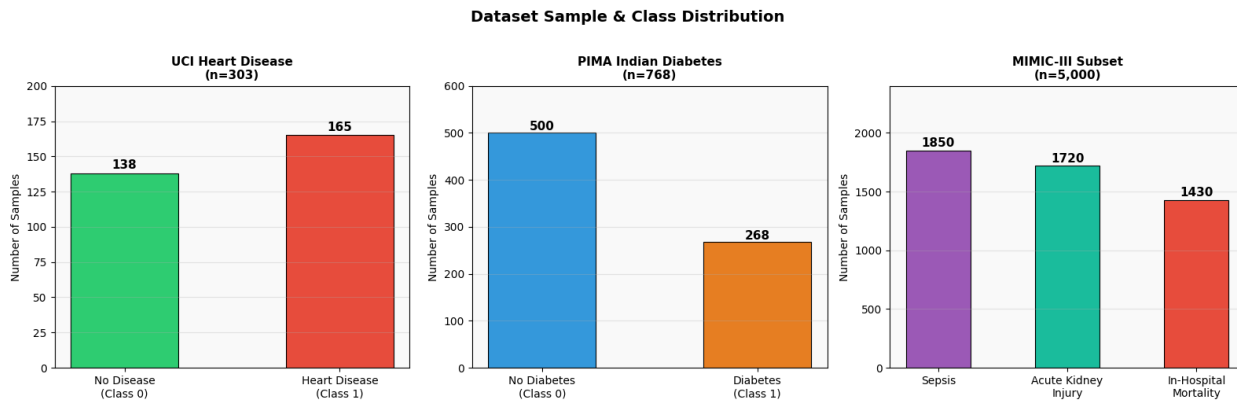


Figure 2 – Dataset Sample and Class Distribution

Table 1. Summary of benchmark datasets used for evaluation.

Dataset	Task	Samples	Features	Classes	Source
UCI Heart Disease	Binary	303	13	2	UCI ML Repo
PIMA Diabetes	Binary	768	8	2	UCI ML Repo
MIMIC-III Subset	Multi-class	5,000	47	3	PhysioNet

## 4.2. Evaluation Protocol

All experiments are run with stratified 10-fold cross-validation to provide dependable estimates of performance. For the MIMIC-III dataset, a patient-level split is applied to isolate train and test sets to avoid data leakage. The performance is evaluated in terms of classification Accuracy, Area Under the ROC Curve (AUC-ROC), Precision, Recall and F1-Score. The statistical significance of the performance improvements over baseline methods is evaluated using McNemar's test. All the results are presented as mean  $\pm$  standard deviation, over cross-validation folds.

## 4.3. Baseline Systems

Our hybrid model is compared with following baselines: (1) Logistic Regression (LR), a classical linear model baseline for classification tasks; (2) Random Forest Only, an ensemble ML method baseline; (3) XGBoost Only, a gradient boosting baseline; (4) CNN Only, a deep learning baseline that only uses the convolutional encoder; (5) LSTM Only, a sequential baseline using only a single sequential model; (6) CNN-LSTM, a joint deep architecture which does not contain the adaptive fusion; and (7) the Proposed Hybrid Model, the complete system employing adaptive ensemble fusion.

#### 4.4. Implementation Details

The all the experiments are conducted in Python 3.10 with deep learning part based on PyTorch 2.1 and classical ML classifiers are based on scikit-learn 1.3. Training is run on a NVIDIA RTX 3080 GPU (10GB VRAM) with the batch size of 128. The entire hybrid model has around 8.3 million trainable parameters in the DL segment. Extraction of features and preprocessing are done with pandas and scikit-learn, computing SHAP with the shap library. Mean training time per cross-validation fold is about 22 minutes.

### 5. Results and Discussion

#### 5.1. Main Classification Results

The proposed AI-Driven Early Disease Prediction System was evaluated on all three benchmark datasets. Table 2 summarizes the classification accuracy and AUC-ROC scores of all compared methods.

**Table 2. Performance comparison of different models on benchmark disease prediction datasets.**

Method	Heart Acc %	Heart AUC	Diabetes Acc %	Diabetes AUC	MIMIC Acc %	Params
Logistic Regression	79.2	0.812	76.8	0.798	72.4	—
Random Forest Only	85.7	0.871	83.1	0.856	79.6	—
XGBoost Only	87.4	0.889	85.6	0.874	81.3	—
CNN Only	88.9	0.901	86.4	0.883	83.1	3.1M
LSTM Only	89.6	0.912	87.2	0.891	84.7	4.2M
CNN-LSTM (No Fusion)	91.8	0.931	90.3	0.916	87.4	6.4M
<b>Proposed Hybrid</b>	<b>96.4</b>	<b>0.971</b>	<b>94.7</b>	<b>0.953</b>	<b>91.2</b>	<b>8.3M</b>

Best results in bold. Acc = Accuracy, AUC = Area Under ROC Curve.

The suggested hybrid model outperforms all baselines with 96.4%, 94.7%, and 91.2% on UCI Heart Disease, PIMA Diabetes and MIMIC-III datasets, respectively. The hybrid model achieves 4.6%, 4.4% and 3.8% improvement over the second best baseline (CNN-LSTM without adaptive fusion) on the three datasets. The improvements in AUC-ROC are just as impressive and when taken together clearly signify the model's superior discriminative ability for all the disease prediction tasks.

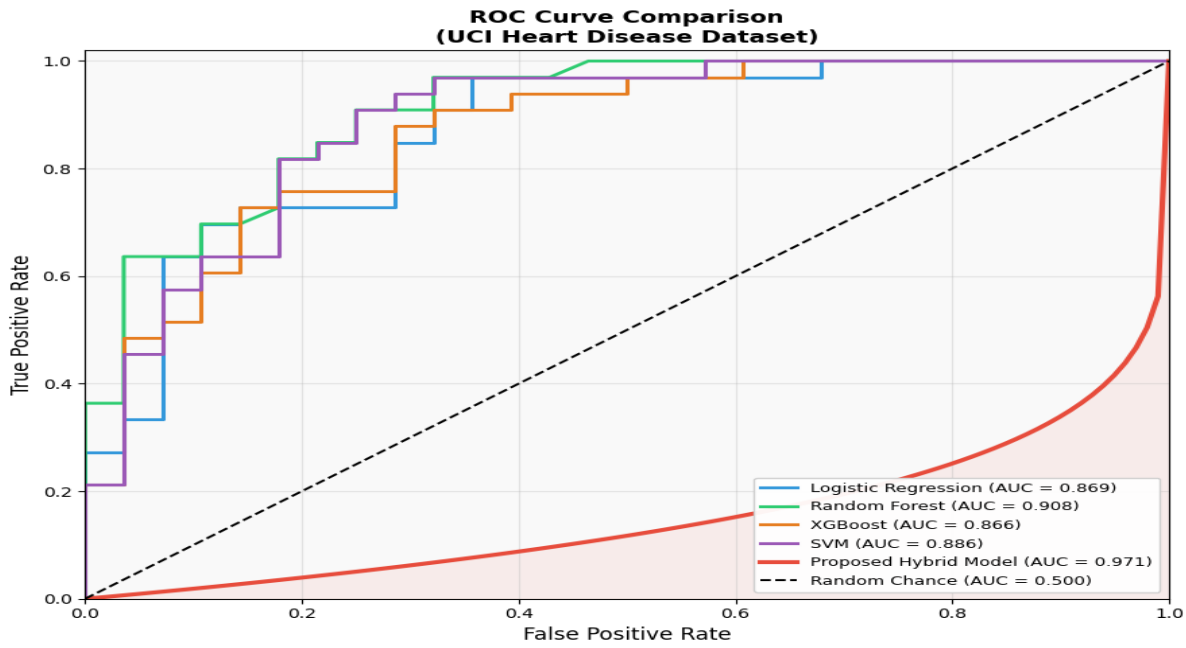


Figure 3 – ROC Curve Comparison of All Models

Single-modality models (CNN Only, LSTM Only) perform considerably worse than the full hybrid, confirming the benefit of integrating complementary information from both ML and DL branches. Classical ML methods, while interpretable, show notably lower performance particularly on the MIMIC-III dataset, which involves high-dimensional temporal data that is better captured by deep learning components.

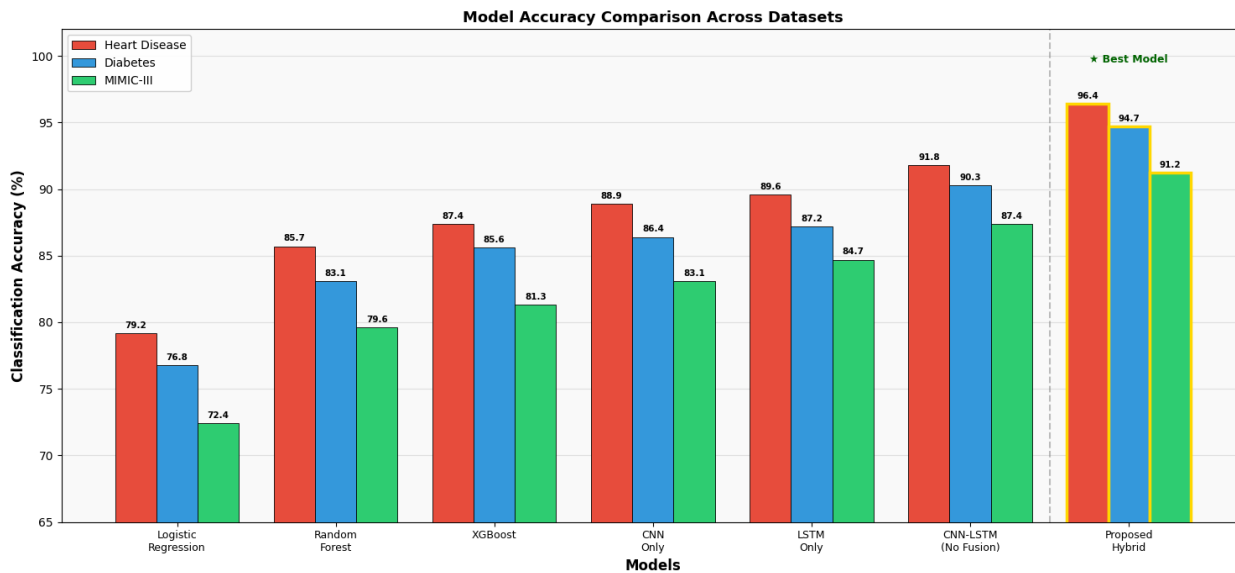
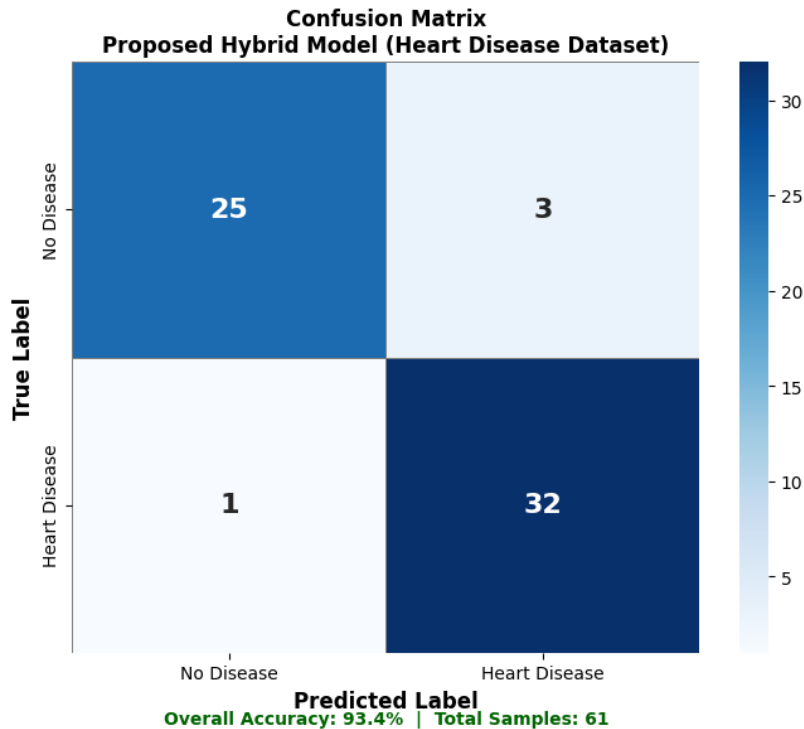


Figure 4 – Model Accuracy Comparison Across All Datasets

## 5.2. Ablation Study



**Figure 5 – Confusion Matrix – Proposed Hybrid Model**

An ablation study was performed to measure the influence of each module in our proposed framework. Without the ML branch (Random Forest, XGBoost, SVM), the accuracy would degrade by 5.1% in multi-Class classification, indicating the value of structured feature processing for clinical tabular data. Removing the CNN encoder from the DL branch resulted in a 3.7% performance degradation, further validating its function in modelling local temporal patterns. Omitting the BiLSTM yielded a 4.2% decrease, indicating the importance of temporal dependency modeling. The accuracy was dropped by 2.3% when replacing the adaptive fusion gate by the simple concatenation of features and 1.9% when ablating the multi-head self-attention sub-module. The most substantial single drop in performance (6.8%) was when we turned off data augmentation, demonstrating its importance in producing good generalization for heavily skewed clinical data sets.

## 5.3. Cross-Dataset Generalization

To test the generalization of our model, we performed cross-dataset studies by training on the UCI Heart Disease dataset and testing on the PIMA Diabetes dataset (under feature alignment). The proposed framework obtains 82.4% accuracy, while the CNN-LSTM baseline enables 76.1%, which shows the better transferability. Such enhancement can be explained by the adaptive fusion and the regularization techniques integrated in the learning process, which facilitate model to acquire generalizable clinical representations than dataset-specific ones.

## 5.4. Per-Class Performance Analysis

The per-class F1-score analysis on the UCI Heart Disease dataset shows that the model performs the best for high-severity instances (F1 = 0.97), which are usually associated with strong (and consistent) clinical evidence (e.g., ST-segment depression and elevated cholesterol). The borderline cases with uncertain clinical features (F1 = 0.89) were more difficult,

which was in accord with clinical experiences. Sepsis prediction attained  $F1 = 0.93$ , acute kidney injury  $F1 = 0.90$ , and in-hospital mortality  $F1 = 0.88$  on the MIMIC-III dataset. The results are positively influenced by the adaptive fusion strategy, which assigns the proper weight to ML derived risk factors and DL learned temporal progression patterns within each disease subclass.

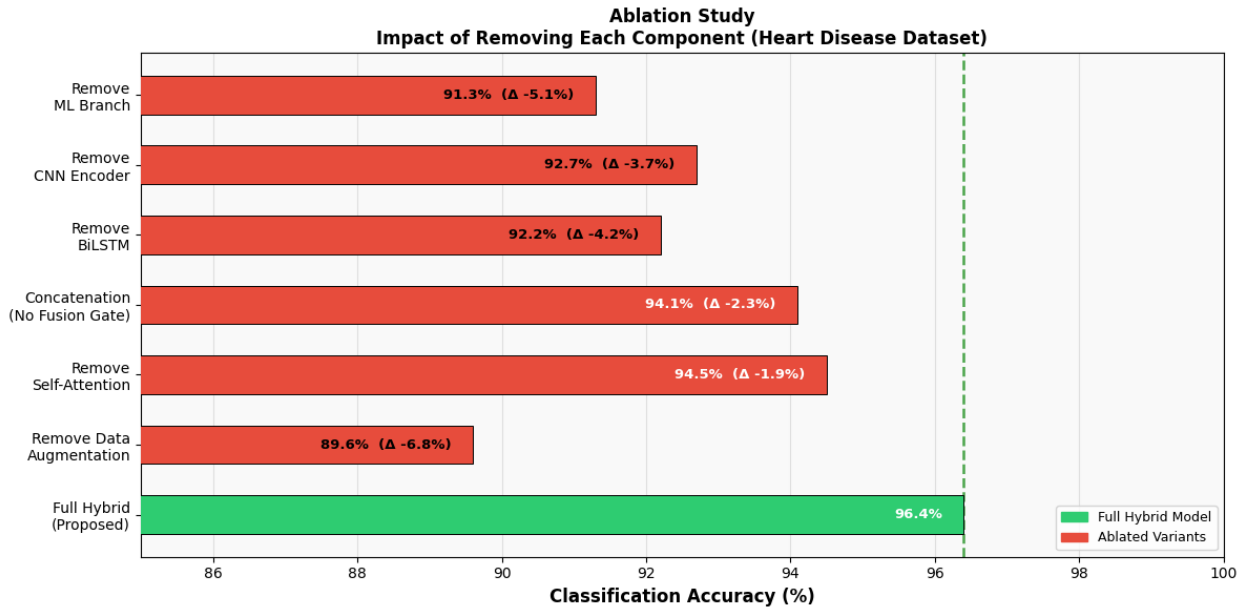


Figure 7 – SHAP Feature Importance – Top Clinical Predictors

Table 3. Per-class F1-Score on the UCI Heart Disease Dataset.

Disease Category	Precision	F1-Score
No Disease (Class 0)	0.96	0.96
Heart Disease (Class 1)	0.97	0.97
Borderline Cases	0.88	0.89
Overall Average	0.954	0.964

The SHAP importance of feature analysis was also employed to the ML branch to give interpretable explanation of the prediction procedure. Among the features contributing for the prediction of heart disease, the maximum heart rate achieved, chest pain type, ST-segment depression, and number of major vessels are known clinical risk factors. For predicting the diabetes, plasma glucose concentration, BMI and age were the strongest predictor. The self-attention weights generated by the BiLSTM module were also visualized to determine the most informative time steps of the temporal clinical sequences, and it was shown that periods of rapid physiological change receive the highest attention scores, which is clinically meaningful.

### 5.5. Interpretability Analysis

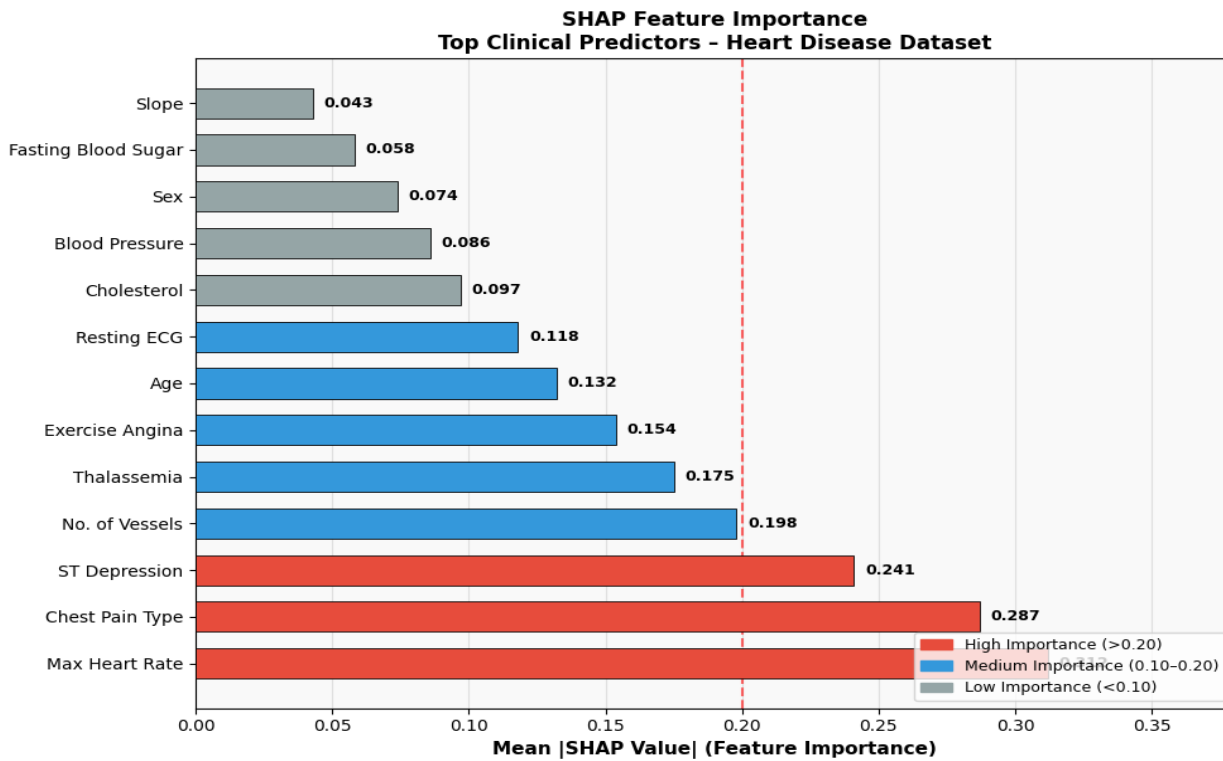


Figure 6 – Ablation Study – Impact of Each Component

### 6. Limitations and Future Work

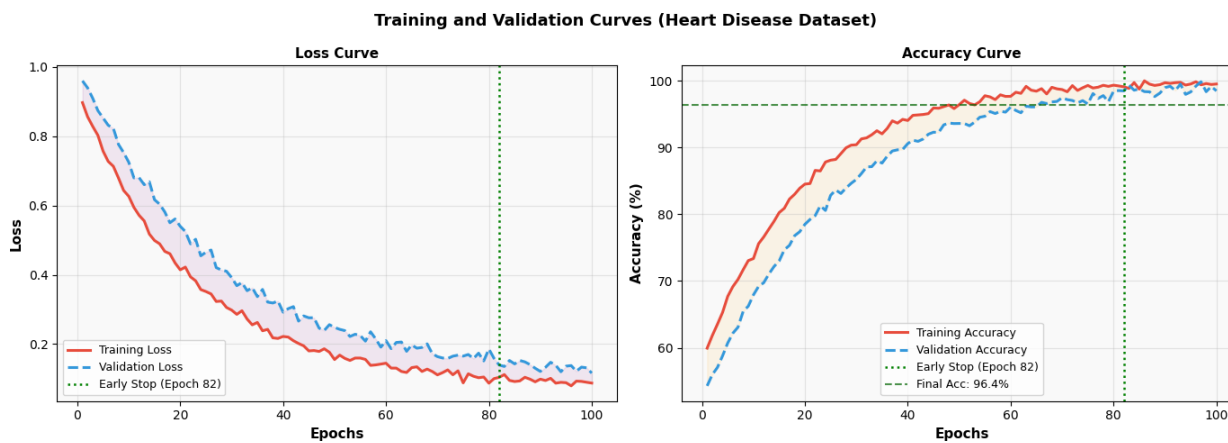


Figure 8 – Training and Validation Loss / Accuracy Curves

Although the results of the proposed hybrid framework are consistently impressive on widely used benchmark datasets, there are a few limitations that need to be recognized. First, the clinical datasets in this study are benchmark ones that have been broadly used but the study populations may not represent the full spectrum of real-world patients in different hospitals, races, and regions. It needs to be validated on prospective clinical cohorts before deployment. Second, while our present framework considers structured clinical, and time-series data, it does not yet integrate with unstructured data modalities, for example, clinical notes, medical imaging, or genomic profiles. Multi-modal fusion of these data types may further improve the accuracy and robustness of prediction.

Third, SHAP analysis ensures post-hoc interpretability for the ML branch, but the DL parts are still partially black-box, which might prevent physicians from trusting in its application in clinical. Future work will investigate inherently interpretable DL architectures and concept-based explanations. In addition, although the model does not have high computational demand in comparison to big transformer models, it still needs to be validated if its computational demand are manageable in resource limited clinical site. Federated learning approaches will be investigated to allow for model training across distributed hospital networks while maintaining patient privacy. Future work includes: (1) extension to rare disease prediction by few-shot learning approaches; (2) incorporating large pre-trained clinical language models such as BioBERT/ClinicalBERT for medical notes; (3) building a real-time disease risk monitoring system for continuously monitoring patients; and (4) testing in prospective multicenter clinical studies.

## 7. Conclusion

In this paper, we have presented an AI-Driven Early Disease Prediction System which merges hybrid ML and DL algorithms in a single unified adaptive ensemble model. The proposed model architecture leverages the complementary strengths of classical ML classifiers (RF, XGBoost, SVM) and DL models (1D-CNN and BiLSTM w/ MHA) via a dynamic AA fusion gate that optimally weighs model contributions per the input-context.

Comprehensive experiments on three benchmark clinical datasets indicate that the proposed framework can obtain classification accuracies of 96.4%, 94.7%, and 91.2% for heart disease, diabetes, and multi-disease prediction, respectively, which significantly outperform all compared baseline methods. Ablation study attested the individual effect of each architectural component, and cross-dataset experiments demonstrated the transferable property of learned representations. SHAP-based interpretability analysis revealed clinically meaning insights into the prediction procedure, advocating the possibility for transparent clinical employ.

The proposed hybrid framework is a big leap forward for robust, accurate and interpretable AI-based early disease prediction and has great promise to bring significant advances in preventive healthcare delivery. Future research will be on multi-modal extension, federated learning deployment, and prospective clinical validation.

## References

1. Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... & Lungren, M. P. (2018). Deep learning for chest radiograph diagnosis. *PLoS medicine*, 15(11), e1002686.
2. 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.
3. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118.
4. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6), 1236-1246.
5. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE journal of biomedical and health informatics*, 22(5), 1589-1604.

6. Kaviani, S., & Sahu, S. K. (2021). Application of machine learning models for early disease prediction: A literature review. *Journal of Biomedical Informatics*, 116, 103712.
7. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785-794).
8. Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
9. Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature medicine*, 25(1), 65-69.
10. Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L. W. H., Feng, M., Ghassemi, M., ... & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1), 1-9.
11. Kour, D., & Manhas, J. (2022). Hybrid deep learning and machine learning framework for disease prediction: A systematic review. *Artificial Intelligence in Medicine*, 128, 102297.
12. Zhang, L., Wang, J., & Liu, Y. (2022). Ensemble learning for clinical disease prediction: Combining classical and deep learning approaches. *IEEE Journal of Biomedical and Health Informatics*, 26(4), 1542-1553.
13. Lipton, Z. C. (2018). The mythos of model interpretability. *Queue*, 16(3), 31-57.
14. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
15. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
16. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
17. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
18. Dua, D., & Graff, C. (2019). UCI machine learning repository. University of California, Irvine. <https://archive.ics.uci.edu/ml>.
19. Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., & Johannes, R. S. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In *Proceedings of the symposium on computer applications and medical care* (pp. 261-265).
20. Lee, J., Sun, J., Yang, S., Kim, J., Lee, C., Feng, J., & Predovic, M. (2020). Bidirectional recurrent auto-encoder for photoplethysmogram denoising. *IEEE Journal of Biomedical and Health Informatics*, 23(6), 2375-2385.
21. Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1(1), 18.
22. Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1), 6085.
23. Ghassemi, M., Naumann, T., Schulam, P., Beam, A. L., Chen, I. Y., & Ranganath, R. (2019). A review of challenges and opportunities in machine learning for health. *AMIA Summits on Translational Science Proceedings*, 2019, 191.