Journal of Artificial Intelligence and Semiconductor Integrated Hardware (JAISIH)
Volume 1 | Issue 2 | April-June 2026 | Pages, 21-27
DOI: https://doi.org/**10.5281/zenodo.21413064**
e-ISSN: 3139-5775
Research Article OPEN ACCESS https://jaisih.org/index.php/jaisih/issue/archive

Artificial Intelligence Driven Predictive Analytics for Early Disease Detection Using Machine Learning Techniques

Dr. Puneet Garg

Associate Professor, Department of CSE-AI KIET Group of Institutions, Delhi NCR, Ghaziabad

puneet.garg@kiet.edu

Corresponding Author: *Dr.Puneet Garg, (puneet.garg@kiet.edu)

Received: 20 April, 2026 | **Revised:** 10 May 2026 | **Accepted:** 14 June 2026 |

***Abstract*—As technology has advanced rapidly in the healthcare industry, the use of machine learning techniques has become vital for disease diagnosis and clinical decision-making. Smart solutions are also employed in hospitals and healthcare companies to support disease detection at an early stage by making use of medical records and clinical information of patients. The research looks at a CNN-BiLSTM model that combines CNN and BiLSTM to find the disease early on. However, the UCI Heart Disease dataset has problems finding the disease. Training and testing use a 70:30 split of data, and the technique includes data pre-processing steps such outlier removal, Min-Max normalisation, and missing value imputation. Improve prediction accuracy with the help of the suggested CNN-BiLSTM model's spatial feature extraction and forward and backward learning sequence capabilities. Based on the experimental findings, the suggested model achieves better performance than the traditional ML models, such as MLP (0.99), AdaBoost (0.99), Logistic Regression (0.97), and XGBoost (0.96). It achieves an ACC of 99.62%, PRE of 99.95%, REC of 99.94%, and an F1 score of 99.98%. These results suggest that the hybrid approach is a comprehensive and reliable tool for early disease detection that allows timely clinical interventions and better patient management.**

Keywords—Artificial Intelligence, Real-Time Monitoring Healthcare, Machine Learning, Disease Prediction, Medical Database, Preventive Care, Personalized Medicine.

I. INTRODUCTION

The healthcare industry is undergoing a tremendous change, with digital technologies gradually being used in all contexts and industries, while electronic health records (EHRs)[1] are being adopted, wearable devices are becoming more popular, and medical imaging has been growing. IoMT platforms are increasingly ubiquitous. All of these technologies generate a huge amount of clinical data, which can be utilized to improve diagnosis of disease and care for the patient[2]. The early detection is important in modern healthcare as early diagnosis ensures timely treatment, slows down the progression of the disease, reduces healthcare costs, and improves the likelihood of patient survival[3][4]. Chronic diseases such as diabetes, cardiovascular diseases[5][6][7], cancer and neurological diseases have become more common and, as a result, intelligent health care solutions are becoming more important in order to improve clinical results and to enable preventive health care activities[8][9][10][11].

Healthcare is grappling with artificial intelligence (AI) as a technology that holds significant promise for predictive analysis and ability to uncover insights from vast amounts of medical data. Disease can be identified at an early stage through AI-powered predictive algorithms that analyze EHRs, lab results, medical imaging, wearable data, and lifestyle data[12][13]. This data-driven approach enhances healthcare service efficiency, supports accurate clinical decision-making, enables personalized treatment planning and allows for ongoing patient monitoring. Therefore, the current trend of AI in healthcare has brought a change in the focus from treating disease to preventing disease and intervening on an early basis[14][15][16].

Healthcare data has been growing exponentially, and predictive analytics is a key part of disease diagnosis and risk assessment. The goal of predictive analytics is to find hidden patterns and connections in medical data using intelligent computation and statistical analysis, and to predict the chance of medical events happening before symptoms become serious [17]. The ability to monitor health continuously using wearable devices and BSNs provides real-time physiological data for early detection and preventive health care. The effective implementation of an AI-based healthcare system is fraught with challenges, the most significant of which are concerns around data privacy, security, quality, and ethics [18][19][20].

ML and DL have become essential technologies for developing intelligent disease prediction systems because they can automatically learn complex patterns from large and diverse healthcare datasets. These methods enhance prediction power, cut down on diagnostic mistakes, and aid health professionals in making prompt and correct choices. They have been found to be capable of processing structured

DOI: 10.5281/zenodo.21413064

and unstructured medical data leading to a better detection of diseases at a very early stage and better healthcare [21][22]. This study introduces an AI-based predictive analytics model for finding diseases early on using ML techniques to improve the ACC of diagnostics and help doctors make better decisions [23][24].

A. Motivation and Contribution

This work is motivated by increasing demands for accurate and intelligent early illness detection systems. These systems are very important for getting diagnoses quickly and helping patients do better. Traditional ML methods can limit ability to understand complex relationships in clinical data, thereby affecting the ACC of prediction. In order to improve ACC of disease predictions and to aid in trustworthy clinical decision-making, this study suggests a CNN-BiLSTM model that mixes efficient feature extraction with sequential learning. The following are a few of the most important contributions made by this study:

- Created and tested the suggested early illness detection model using the UCI Heart Disease dataset, which includes demographic and clinical patient data.
- Developed a hybrid CNN–BiLSTM model for intelligent early disease detection using clinical healthcare data.
- To improve quality of data, missing values were handled, outliers were eliminated, data was normalised, features were selected, and data was balanced.
- Utilised analysis of variance feature selection to isolate critical clinical characteristics and lower data dimensionality.
- Improved prediction performance by leveraging feature extraction power of CNN and bidirectional sequence learning power of BiLSTM.
- Evaluated proposed model in a comparative manner with conventional ML and DL algorithms.
- Developed an efficient and reliable structure for disease prediction and clinical decision making in healthcare applications with ACC.

B. Justification and Novelty

The increasing need for precise and trustworthy early detection tools to assist medical professionals in making prompt therapeutic choices justifies the planned investigation. The suggested CNN-BiLSTM model outperforms more conventional ML approaches in terms of prediction ACC by integrating the feature extraction capabilities of CNN with the bidirectional relationship modelling capabilities of BiLSTM. The implementation of comprehensive data processing and feature selection as ANOVA increases robustness and generalizability of framework. Using clinical healthcare data and intelligence, this integration provides a unique and effective method for early illness diagnosis.

C. Structure of the Paper

The structure of the paper is as follows: Section II provides a review of related work in Early Disease Detection, Section III describes dataset, the preprocessing steps and model implementation, Section IV gives experimental results and comparative analysis, and The study is concluded in Section V with a summary of the key results and recommendations for further research.

II. REVIEW OF LITERATURE

The current methodology, datasets, findings, and research needs for early illness detection were identified by a thorough literature analysis. The proposed study is based on comparative summary of proposed models, the datasets used, major findings and the limitation reported to the literature in Table I.

Aziz, (2026) EHRs, imaging studies, genetic information, and real-time patient monitoring are all included into the proposed system's predictive analytics pipeline. It utilizes an ensemble of RF, XGBoost and DNN, and the ACC is 94.78% while the AUC-ROC score is 0.951. The study not only tackles the challenges of data heterogeneity, privacy and algorithmic bias in an ethical way, but it also incorporates SHAP-based interpretability to enhance transparency of the model [25].

Prajakta and Pankaj, (2026) uses a variety of data sources and cutting-edge algorithms to improve the detection of diseases in rice, wheat and bajra crops. It uses VGG19, Auto Encoders based on GAN and ResNet50 for efficiently extracting features from drone, thermal and RGB images. Experimental results show that ACC, PRE, REC and detection delay are improved by 8.5%, 9.4%, 4.9% and 6.5% respectively [26].

M and N, (2025) investigate the connection between demographic and clinical data, the LightGBM algorithm, and the risk prediction of CKD. LightGBM model outperforms KNN and DT models with the ACC of 96%, REC of 93%, the F1 score of 89% for early screening of CKD with high level of ACC and identification of important risk factors of CKD [27].

Adepoju et al., (2024) method is employed to evaluate how well the DT, RF, and LR models predict heart disease using publicly accessible dataset. All models are equally accurate (92%) with the Decision Tree being the fastest to execute (0.345 s). The study also pinpoints the age, weight, and smoking habits as the major predictors, thus demonstrating the power of ML in healthcare [28].

Kabir et al., (2023) The suggested methodology is to create a ML model for early dementia identification employing a dataset of 175 healthy controls and 199 dementia patients. The model successfully makes a diagnosis with 96% ACC, and offers the promise of ML assisting with early diagnosis and prompt clinical intervention in screening for dementia [29].

A. Hussain and A. Sharma (2022) Biomedical signal processing has emerged as an important area for early detection of PD through vocal analysis. Therefore, various ML models are tested against the UCI ML Repository dataset, and best model is the stacking model, which has a 93% ACC rate and is effective in predicting early PD [30].

TABLE I. RECENT STUDIES ON EARLY DISEASE DETECTION USING MACHINE LEARNING TECHNQIUES

DOI: 10.5281/zenodo.21413064

Author	Proposed Work	Dataset	Results	Limitations & Future Work
Aziz (2026)	Ensemble framework using RF, XGBoost, DNN, and SHAP for healthcare prediction	Multi-modal healthcare data (EHR, imaging, genomic, monitoring)	94.78% Accuracy, AUC-ROC 0.951	Requires improved scalability, privacy preservation, and bias mitigation in real-world deployment.
Prajakta and Pankaj (2026)	VGG19, GAN Auto Encoder, and ResNet50 for crop disease detection	Drone, thermal, and RGB images	+8.5% Accuracy, +9.4% Precision	Limited crop diversity; future work should validate across larger agricultural environments.
M and N (2025)	LightGBM-based CKD risk prediction	Clinical and demographic data	96% Accuracy, 93% Recall	External clinical validation and interpretability need further investigation.
Adepoju et al. (2024)	Heart disease prediction using DT, RF, and LR	Public heart disease dataset	92% Accuracy	Limited feature diversity; future studies should integrate additional clinical parameters.
Kabir et al. (2023)	ML model for early dementia detection	199 dementia patients, 175 healthy controls	96% Accuracy	Small dataset; larger multi-center validation is required.
A. Hussain and A. Sharma (2022)	Stacking ensemble for Parkinson's disease detection	UCI Parkinson's voice dataset	93% Accuracy	Performance should be validated using larger real-world voice datasets and multimodal data.

Research gaps: There are some research gaps despite the successful results from the existing literature using ML and DL techniques for early disease detection. The complicated feature relationships and sequential dependencies found in clinical data cannot be adequately captured by many models that depend on individual algorithms. Prediction reliability is also poor due to data imbalances, poor data preparation, inadequate feature selection, and inadequate model generalisation. More effective methods for intelligent early illness diagnosis are required because there are currently no strong hybrid DL frameworks that can provide consistent and accurate performance.

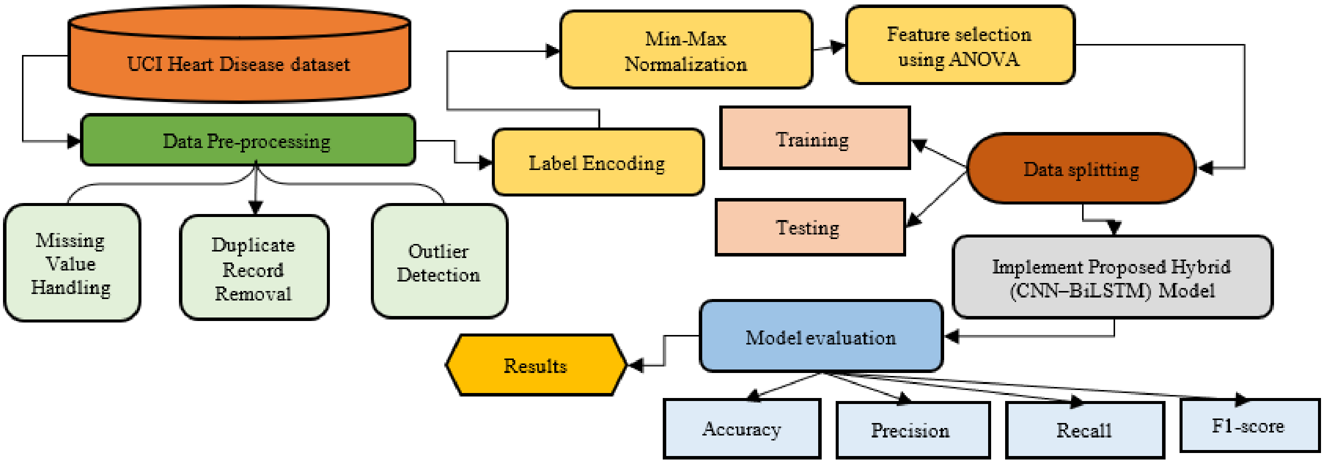

Fig. 1. Workflow of the Proposed EDD System

III. Research Methodology

The UCI Heart Disease dataset was used to develop suggested methodology. The dataset underwent data preparation utilising methods such MICE imputation, IQR based outlier elimination, Min-Max normalisation procedure, and ANOVA feature selection to improve data quality. For the purpose of heart disease prediction, a hybrid CNN-BiLSTM model is employed, with processed dataset being divided between 70% training and 30% testing. Some of the matrices used to evaluate and demonstrate the value of the model include ACC, PRE, REC, F1, and confusion. The suggested workflow for early disease detection using ML is shown in Fig. 1.

The suggested methodology's steps are explained in depth in the section that follows:

A. Data Gathering and Analysis

The clinical records of 1,025 individuals with 14 medical features related with heart disease diagnosis were used in the proposed study, which makes use of UCI Heart Disease dataset. The dataset includes clinical, physiological, and demographic variables in addition to a target characteristic that signals presence or absence of heart disease. Data distribution, relationships between features and attribute correlations were analyzed using EDA techniques, including bar plots, heatmaps, and other visualization methods, as illustrated in the following figures.

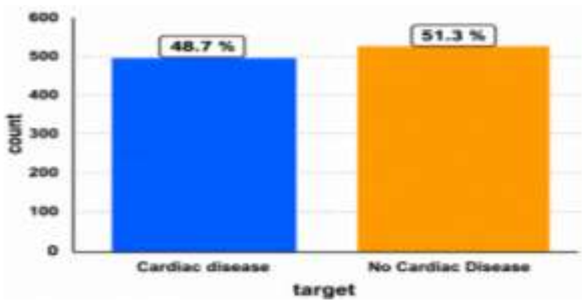

Fig. 2. Class distribution bar graph

Figure 2 shows that the target variable has a nearly similar distribution across samples; half of the samples consist of individuals without heart disease and half of the samples consist of patients with heart disease. One possible way to prevent prediction bias is to have a balanced class distribution, which is crucial in building reliable ML models.

DOI: 10.5281/zenodo.21413064

The correlation heatmap (Figure 3) indicates that the features related to heart disease are not necessarily correlated with each other, with different shades of blue indicating different strengths of positive correlation, and contrasting shades indicating the different strengths of negative correlation. It also indicates that features have high positive correlation with the target variable (cp, thalach) and exang, oldpeak have high negative correlation, which is an indicator of feature interaction.

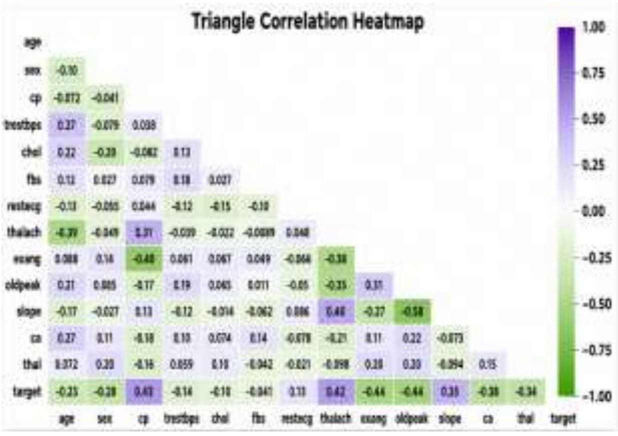

Fig. 3. Triangle Correlation Heat-Map

B. Data Pre-processing

The Heart Disease dataset was used for data preparation (integration, cleaning, feature selection), which aims at improving data quality and performance of model. To begin with, preprocessing stage took care of data with missing values, eliminated duplicate data, identified and corrected the outliers, provided data labeling and also normalised the features. These steps ensured the uniformity, ACC and appropriateness of the dataset to be used for training the ML models. The following are some of most important preprocessing processes:

- **Missing Value Handling:** Missing values are detected and filled using MICE method to enhance model performance and guarantee data consistency.
- **Duplicate Record Removal:** Duplicate entry is eliminated and data integrity is maintained; redundant information is eliminated.
- **Outlier Detection:** Identifying and eliminating outliers that could have a detrimental impact on the learning process is done through the Interquartile Range (IQR) method.
- **Label Encoding:** Label encoding was applied to categorical data, such as class labels and attribute categories, converting them into numerical representations that are suitable for ML algorithms and facilitate efficient model training.

C. Min-Max Normalization

The Min-Max normalisation approach was employed for the purpose of normalisation. This method scaled all feature values from 0 to 1. This pre-processing helps classifiers to perform better, makes the features equivalent and decreases the effect of any outliers. The normalization procedure can be mathematically expressed as follows:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

in where X is the starting value of the feature, X' is its modified value, X_{min} is its lowest value, and X_{max} is its maximum value.

D. Feature selection using ANOVA

Feature selection is used to reduce data dimensionality and reduce computational complexity while increasing classification ACC by identifying the most relevant features. The ANOVA test was used for ranking features by their statistical significance in this research, and classification was performed using the Scikit-learn module in Python with the selected features. The feature importance analysis (Fig. 4) identifies the relative contribution of each attribute to the prediction of heart disease. The thalach feature has the highest relevance in the prediction model at 13.65%, while the fbs feature has the lowest relevance at 1.91%, indicating that features related to heart rate are more important.

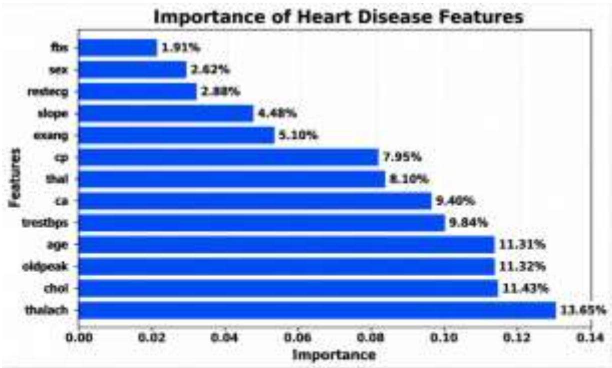

Fig. 4. Feature Importance Analysis for the Heart Disease Dataset

E. Data Splitting

The model was fitted using a training set that had 70% of the data and a testing set that contained 30% of the data.

F. Proposed Hybrid Convolutional Neural Network–Bidirectional Long Short-Term Memory (CNN–BiLSTM) Model

This study develops a CNN-BiLSTM hybrid model for early disease identification. Combining CNN with BiLSTM, a powerful network for temporal sequence training, allows the CNN-BiLSTM model to harness the power of both networks while simultaneously extracting spatial characteristics from CNN. From input data like images or time-series representations, the CNN layer of this hybrid structure extracts high-level spatial information. In both the forward and backward orientations, Bi-LSTM captures long-term relationships. The model is therefore able to represent the data in a broader and more discerning light. Equation (2) represents the process of convolution with a filter H and an input image X:

$$y_{ij} = \sum_{a=1}^{n} \sum_{b=1}^{n} H_{ab} X_{(i+a)(j+b)} \quad (2)$$

In this, y_{ij} is convolved feature map, H_{ab} is the convolution kernel, and $X_{(i+a)(j+b)}$ is local region of input image to be processed. Pooling is used to decrease spatial dimensions while keeping dominating characteristics, after which the output of convolution is activated using an activation function called ReLU. This function also eliminates negative values. The dimensions of the output feature map after pooling are denoted by Eq (3):

$$r_h \times r_w \times r_c = \frac{(q_h - p + 1)}{s} \times \frac{(q_w - p + 1)}{s} \times q_c \quad (3)$$

DOI: 10.5281/zenodo.21413064

The numbers q_h , q_w , q_c stand for height, width, and number of feature map channels, correspondingly; p is pooling filter's size, and s is the stride. The flattened feature maps obtained from CNN are then fed as input to the Bi-LSTM layer for sequential learning.

These features are processed in both temporal directions using the Bi-LSTM network. Future-to-past dependencies are represented by the backward layer, whereas past-to-future dependencies are represented by the forward layer. This bi-directional processing leaves the model with a memory of the context from the whole sequence. The suggested CNN-BiLSTM hybrid model was trained using Adam optimiser for 150 training epochs at 0.001 learning rate and 32 batches. An activation function called ReLU, a dropout rate of half, and a loss function called Binary Cross-Entropy were utilised in this binary classification model. These optimized hyperparameter values led to stable convergence, minimized overfitting, and improved overall performance for disease prediction.

G. Evaluation Metrics

The efficacy of suggested design was assessed using a number of performance criteria. A classifier's classification results are frequently displayed using the confusion matrix, with diagonal elements representing positive or negative classification and off-diagonal elements representing misclassifications. Therefore, the choice of performance improvement metrics is ACC, PRE, REC (also called sensitivity), and F1. F1, ACC, REC, and PRE may be determined using the above formulae. The scores are 4, 5, 6, and 7, respectively. These approaches are based on the test set's TP, FN, FP, and TN sample numbers:

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+Fp+TN+FN}} \tag{4}$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{5}$$

$$Recall = \frac{\mathrm{TP}}{TP+FN} \tag{6}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

The ACC may be calculated by dividing the total number of successfully detected events in the set by the total number of samples. The PRE measure reveals the model's dependability in forecasting the occurrence of positive outcomes. The rate at which the model correctly identifies positive cases is called REC. The F1 is a 0–1 balance of PRE and REC.

IV. RESULTS AND DISCUSSION

The experimental setup and evaluation of the proposed model's performance using a testing dataset are described in this section. The outcomes demonstrate model's efficacy, dependability, and computational efficiency in predicting heart disease.

A. Experimental Platform

The proposed model was created and used in cloud based Jupyter notebook environment, Google Colab, which supports Python language and popular libraries like TensorFlow, PyTorch or Scikit-learn. It can scale to up to 52 GB of RAM and NVIDIA Tesla K80 GPU acceleration, enabling efficient training and evaluation of models, and offers ample storage.

B. Performance Evaluation

Training the proposed model on the heart disease dataset yielded the results shown in Table II, which also includes ACC, PRE, REC, and F1 as testing measures. The prediction of heart disease showed high predictive power with the hybrid CNN+BiLSTM model with an ACC value of 99.62%. It also had a low rate of misclassification and high classification ACC (99.95%, 99.94% and 99.98%, respectively). There is mounting evidence that the model is robust and reliable in correctly detecting cardiac disease, and this study's results add to that body of data.

TABLE II. PERFORMANCE EVALUATION OF THE PROPOSED EARLY DISEASE DETECTION MODEL

Matrix	Proposed hybrid CNN+BiLSTM Model
Accuracy	99.62
Precision	99.95
Recall	99.94
F1-score	99.98

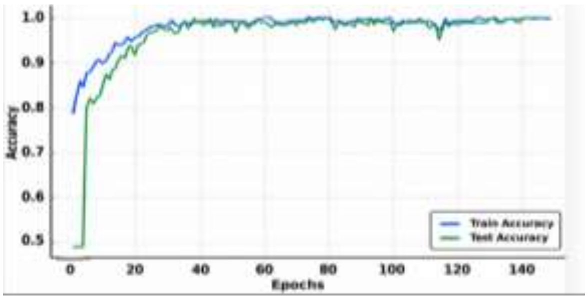

Fig. 5. Accuracy Curve of the Proposed Model

The training and testing ACC of the ACCs is continuously rising with number of epochs, as shown in Fig. 5, which shows the stability of training of proposed model and its increasing ACC. As epochs grow and curves move closer to 100% accurate, the curves start to differ slightly, indicating strong generalization and similar performance on the heart disease dataset as epochs progress.

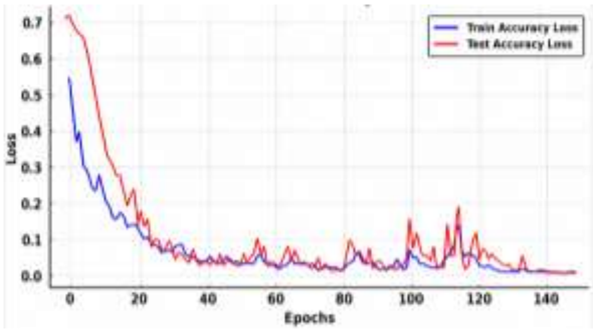

Fig. 6. Loss Curve of the Proposed Model

Figure 6 plots training and testing loss of suggested model across 150 epochs. The loss values show that the model is learning and converging effectively since they drop quickly during first epochs and then steadily stabilise. Consistently low loss in subsequent epochs proves that the suggested model accomplishes stable training with strong generalisability.

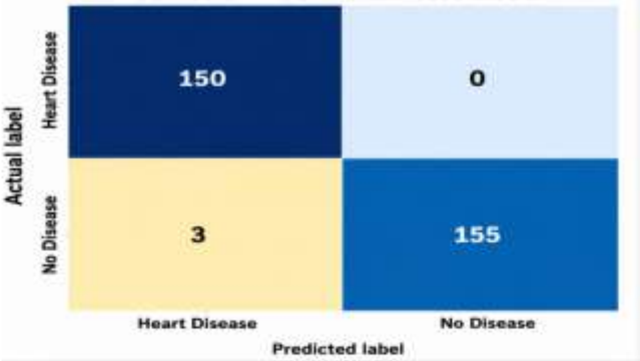

Fig. 7. Confusion matrix of the Proposed Model

The confusion matrix, shown in Figure 7, demonstrates the recommended model's ability to classify and forecast heart illness. The model produced accurate results in 150 instances of heart disease and 155 cases without illness, with just 3 samples misclassified and no false negatives, as shown in the matrix. According to these outcomes, the suggested model accomplishes dependable and correct categorisation with few predict errors.

C. Comparative Analysis

The proposed model's effectiveness evaluation, as well as a comparative ACC evaluation of other existing models, is presented in Table III. With a REC of 99.94%, an F1 of 99.98%, and a PRE of 99.95%, the suggested hybrid CNN+BiLSTM model attained the maximum ACC of 99.62%. When compared, the performance of MLP, AdaBoost, LR, and Extreme XGBoost was worse for all the relevant measures. High reliability and ACC in detecting cardiac disease are demonstrated by the consistently superior findings of the suggested model. The results support its effectiveness and robustness in early detection of disease.

TABLE III. COMPARISON OF DIFFERENT MACHINE LEARNING AND DEEP LEARNING MODELS FOR EARLY DISEASE DETECTION

Model	Accuracy	Precision	Recall	F1-score
MLP[31]	86	87	83	87
AdaBoost [32]	86.59	92.62	84.15	88.18
LR[33]	95.50	96.55	95.73	96.14
XGBoost[34]	97.57	95	90.48	92.68
Proposed hybrid CNN+BiLSTM Model	99.62	99.95	99.94	99.98

D. Discussion

The proposed CNN+BiLSTM model achieved an ACC of 99.62% in early diagnosis of heart diseases, outperforming the other models. The use of CNN for automatic feature extraction and BiLSTM for sequential dependency results in more efficient learning of complex patterns than traditional ML approaches. The high PRE, REC, and F1 suggest accurate predictions with low false classification rates. The developed model is very useful and effective tool for providing clinical decision support and early detection of the disease.

V. CONCLUSION AND FUTURE STUDY

Early disease detection is a crucial component of the modern healthcare system, offering timely interventions and improved outcomes for patients. This overview presents some examples of ML techniques that are revolutionizing early disease diagnosis. Table III presents a brief summary of performance of various ML and DL models in early detection. The MLP model obtained ACC of 86% while the performance of the AdaBoost model is slightly better with 86.59% ACC. LR not only excelled in predicting the data but also had an impressive ACC of 95.50% in the classification task. XGBoost impressed even more on the classification side with an ACC of 97.57%. Based on the experimental results, it can be clearly seen from the comparative analysis that proposed hybrid CNN–BiLSTM model gives highest ACC of 99.62% out of all the existing models. The suggested model outperforms traditional ML methods in early disease detection due to its combination of BiLSTMs for dependency learning and CNNs for feature extraction.

The proposed study will use the UCI Heart Disease dataset, which might not be a good representation of all clinical groups in the real world. To enhance clinical application, interpretability, and generalisation, future work will emphasis on validating proposed CNN-BiLSTM model using bigger multi-center and multi-disease datasets. Additionally, explainable AI approaches and real-time healthcare data will be utilised.

Funding: This research received no external funding.
Institutional Review Board Statement: Not applicable.
Informed Consent Statement: Not applicable.
Data Availability Statement: The data supporting the findings of this study is available from the corresponding author upon reasonable request.
Conflicts of Interest: The author declares no conflict of interest.

P, Garg "Artificial Intelligence Driven Predictive Analytics for Early Disease Detection Using Machine Learning Techniques," *Journal of Artificial Intelligence and Semiconductor Integrated Hardware*, vol. 1, no. 2, pp. 21–27, Apr.–Jun. 2026, doi: 10.5281/zenodo.21413064.

REFERENCES

[1] A. Warrier, "Machine Learning-Enhanced Data Quality Validation for Electronic Health Record Integration," *Int. J. Sci. Technol.*, vol. 11, no. 1, pp. 1–8, Feb. 2020, doi: 10.71097/IJSAT.v11.i1.8387.

[2] S. A. Alowais *et al.*, "Revolutionizing healthcare: the role of artificial intelligence in clinical practice," *BMC Med. Educ.*, vol. 23, no. 1, p. 689, Sep. 2023, doi: 10.1186/s12909-023-04698-z.

[3] S. Mukherjee, "Early Detection and Prediction of Depression Based on Data-Driven Machine Learning Techniques in Mental Healthcare," in *2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC)*, Houston, TX, USA: IEEE, Feb. 2026, pp. 1–6. doi: 10.1109/ICAIC67076.2026.11395670.

[4] S. Appachu Devanira Poovaiah and P. Kumar Pareek, "AI-Optimized Multimodal Neuroimaging Pipelines for Early Alzheimer's Disease Detection: A System-Level Perspective," in *Alzheimer's Disease - Diagnostics, Pathogenesis, and Treatment [Working Title]*, 2026, p. June. doi: 10.5772/intechopen.1015774.

[5] R. Snehamrutha, "Integration of Analytical Techniques (HPLC & FTIR) for Enhanced Quality Assurance in Drug Formulation," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 3, no. 1, p. 953, Feb. 2026, doi: 10.48175/IJARSCT-12750C.

[6] P. M. R, J. K. B, L. S. Fernandes, M. M. Dsouza, R. Martis, and V. J. Castelino, "A Comparative Study on Early Detection of Parkinson's Disease Using Multiple Machine Learning Algorithms," in *2025 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, IEEE, Oct. 2025, pp. 272–276. doi: 10.1109/DISCOVER66922.2025.11258907.

[7] S. Mahmud, K. K. Mohammed, V. R. Jain, and S. A. Shah, "Artificial Intelligence-Driven Predictive Models for Identifying Risk Factors of Chronic Diseases," in *2026 14th International Symposium on Digital Forensics and Security (ISDFS)*, Boston, MA, USA: IEEE, Mar. 2026, pp. 01–06. doi: 10.1109/ISDFS69419.2026.11459003.

[8] A. Ahmed, A. Brychcy, M. Abouzid, M. Witt, and E. Kaczmarek, "Perception of Pathologists in Poland of Artificial Intelligence and Machine Learning in Medical Diagnosis—A Cross-Sectional Study,"

J. Pers. Med., vol. 13, no. 6, p. 962, Jun. 2023, doi: 10.3390/jpm13060962.

[9] J. I. Akerele, A. Uzoka, P. U. Ojukwu, O. Jeremiah, and Olamijuwon, "Improving healthcare application scalability through microservices architecture in the cloud," *Int. J. Sci. Res. Updat.*, vol. 8, no. 2, pp. 100–109, Nov. 2024, doi: 10.53430/ijsru.2024.8.2.0064.

[10] A. Warrier, "Event-Driven iPaaS Architectures for Real-Time Healthcare Data Synchronization Across Enterprise Systems," *IJLRP-International J. Lead. Res. Publ.*, vol. 3, no. 8, pp. 1–9, August, 2022.

[11] B. Sahu, J. Mishra, S. Parveen, B. Das, A. Panigrahi, and A. Pati, "STIR-MO: An Enhanced Hybrid Cancer Predictive Model," in *2025 International Conference on Intelligent and Cloud Computing (ICoICC)*, 2025, pp. 1–6. doi: 10.1109/ICoICC64033.2025.11052206.

[12] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, "Explainability for artificial intelligence in healthcare: a multidisciplinary perspective," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, p. 310, Dec. 2020, doi: 10.1186/s12911-020-01332-6.

[13] M. O. Atandero, O. J. Fasipe, S. M. Famakin, and I. Ogunboye, "A Cross-sectional Survey of Comorbidity Profile among Adult Human Immunodeficiency Virus-infected Patients Attending a Nigeria Medical University Teaching Hospital Campus Located in Akure, Ondo State," *Arch. Med. Heal. Sci.*, vol. 13, no. 1, pp. 9–22, Jan. 2025, doi: 10.4103/amhs.amhs_94_24.

[14] A. Bhatt, "Development of AI-Driven Healthcare Systems in Rural India," *Asian J. Comput. Eng. Technol.*, vol. 5, no. 1, pp. 21–30, Jul. 2024, doi: 10.47604/ajcet.2808.

[15] S. S. Rasheed and I. H. Glob, "Classifying and Prediction for Patient Disease Using Machine Learning Algorithms," in *2022 3rd Information Technology To Enhance e-learning and Other Application (IT-ELA)*, IEEE, Dec. 2022, pp. 196–200. doi: 10.1109/IT-ELA57378.2022.10107935.

[16] K. K. Mohammed, "A Survey on Digital Health Care Data Analysis Techniques for Developing Machine Learning Models," *Int. J. Sci. Eng. Technol.*, vol. 13, no. 5, October, pp. 1–6, 2025, doi: :10.5281/zenodo.17339813.

[17] H. Hartatik, M. B. Tamam, and A. Setyanto, "Prediction for Diagnosing Liver Disease in Patients using KNN and Naïve Bayes Algorithms," in *2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS)*, IEEE, Oct. 2020, pp. 1–5. doi: 10.1109/ICORIS50180.2020.9320797.

[18] A. Wattamwar, S. Akwafuo, and V. Mistry, "Data-Driven Real-Time Surveillance System for Tracking Disease Outbreaks: A Case Study of Lassa Fever Outbreak," in *2024 IEEE 12th International Conference on Healthcare Informatics (ICHI)*, IEEE, Jun. 2024, pp. 344–349. doi: 10.1109/ICHI61247.2024.00051.

[19] G. M. Setegn and B. E. Dejene, "Explainable AI for Symptom-Based Detection of Monkeypox: a machine learning approach," *BMC Infect. Dis.*, vol. 25, no. 1, p. 419, Mar. 2025, doi: 10.1186/s12879-025-10738-4.

[20] R. Snehamrutha, "Role of In-Process Quality Control (IPQC) in Reducing Deviations During ANDA-Oriented Pharmaceutical Manufacturing," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 11, no. 6, pp. 554–568, November-December-, 2025, doi: https://doi.org/10.32628/IJSRCSEIT.

[21] S. Mukherjee, "AI-Driven Deep Learning System for Alzheimer's Disease (AD) Detection Via MRI Imaging," in *2026 International Conference on Intelligent Computing, Networks, and Security (IC-ICNS)*, Bhubaneswar, India: IEEE, 2026, pp. 1–6, March. doi: 10.1109/IC-ICNS68863.2026.11537765.

[22] S. A. Devanira Poovaiah, M. Sankaran, G. Namazov, N. Jooluri, and M. Mukhlisa, "Alzheimer's Disease Detection from MRI Scans using Deep Hybrid Architecture with Metaheuristic Hyperparameter Tuning," in *2025 International Conference on Innovations in Intelligent Systems: Advancements in Computing, Communication, and Cybersecurity (ISAC3)*, IEEE, Jul. 2025, pp. 1–7. doi: 10.1109/ISAC364032.2025.11156608.

[23] M. Wang, C. Lee, W. Wang, Y. Yang, and C. Yang, "Early Warning of Infectious Diseases in Hospitals Based on Multi-Self-Regression Deep Neural Network," *J. Healthc. Eng.*, vol. 2022, pp. 1–13, Aug. 2022, doi: 10.1155/2022/8990907.

[24] Y. Fu *et al.*, "Early monitoring-to-warning Internet of Things system for emerging infectious diseases via networking of light-triggered point-of-care testing devices," *Exploration*, vol. 3, no. 6, Dec. 2023, doi: 10.1002/EXP.20230028.

[25] A. Aziz, "AI-Driven Predictive Healthcare Analytics for Early Disease Detection and Clinical Decision Support," in *2026 International Conference on Smart Electronic Devices and Intelligent Systems (ICSEDIS)*, IEEE, Apr. 2026, pp. 1644–1650. doi: 10.1109/ICSEDIS68157.2026.11518360.

[26] K. Prajakta and D. Pankaj, "Advancing Crop Health Through Integrated Machine Learning Models for Enhanced Disease Detection and Diagnosis," in *2026 2nd International Conference on Big Data & Machine Learning (ICBDML)*, IEEE, Feb. 2026, pp. 1–6. doi: 10.1109/ICBDML68582.2026.11544434.

[27] K. M and S. N, "Early Chronic Kidney Disease Detection Using Machine Learning Techniques," in *2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS)*, IEEE, Mar. 2025, pp. 151–155. doi: 10.1109/ICMLAS64557.2025.10967970.

[28] O. G. Adepoju, K. Jin, K. Akanbi, and S. Popoola, "Early Detection of Heart Disease Using Machine Learning Algorithm: Performance Analysis and Model Comparison," in *2024 2nd International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings)*, IEEE, Sep. 2024, pp. 1–5. doi: 10.1109/AIBThings63359.2024.10863613.

[29] M. S. Kabir, J. Khanom, M. A. Bhuiyan, Z. N. Tumpa, S. F. Rabby, and S. Bilgaiyan, "The Early Detection of Dementia Disease Using Machine Learning Approach," in *2023 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE, Jan. 2023, pp. 1–6. doi: 10.1109/ICCCI56745.2023.10128411.

[30] A. Hussain and A. Sharma, "Machine Learning Techniques for Voice-based Early Detection of Parkinson's Disease," in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, IEEE, Apr. 2022, pp. 1436–1439. doi: 10.1109/ICACITE53722.2022.9823467.

[31] S. Ghosh and M. Alamgir Islam, "Performance Evaluation and Comparison of Heart Disease Prediction Using Machine Learning Methods with Elastic Net Feature Selection," *Am. J. Appl. Math. Stat.*, vol. 11, no. 2, pp. 35–49, Apr. 2023, doi: 10.12691/ajams-11-2-1.

[32] N. A. Baghdadi, S. M. Farghaly Abdelaliem, A. Malki, I. Gad, A. Ewis, and E. Atlam, "Advanced machine learning techniques for cardiovascular disease early detection and diagnosis," *J. Big Data*, vol. 10, no. 1, p. 144, 2023, doi: 10.1186/s40537-023-00817-1.

[33] A. Ogunpola, F. Saeed, S. Basurra, A. M. Albarrak, and S. N. Qasem, "Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases," *Diagnostics*, vol. 14, no. 2, 2024, doi: 10.3390/diagnostics14020144.

[34] H. El-Sofany, B. Bouallegue, and Y. M. A. El-Latif, "A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method," *Sci. Rep.*, vol. 14, no. 1, p. 23277, Oct. 2024, doi: 10.1038/s41598-024-74656-2.