International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Heart Disease Prediction Using Deep Learning and Ensemble Machine Learning Models

Sanapathi Meghana
Reg. No. 24Q71F0057
meghanasanapathi9@gmail.com
Department of Master of Computer Applications
Avanthi Institute of Engineering and Technology (Autonomous)
Vizianagaram, Andhra Pradesh, India
Under the guidance of Associate Professor Mr. P. Satyanarayana, M.Tech., (Ph.D.)
pukkallasatya84@gmail.com

Abstract—*Heart disease remains one of the leading causes of mortality worldwide, necessitating the development of accurate and reliable prediction systems. This study presents a comparative analysis of several ensemble machine-learning algorithms and a deep neural network (DNN) model for the prediction of heart disease using clinical data. The models evaluated include Random Forest, AdaBoost, Gradient Boosting, XGBoost, LightGBM, and a deep-learning model built using TensorFlow. The dataset comprises patient records with attributes such as age, cholesterol level, resting blood pressure, and exercise-induced angina. Performance was evaluated using standard metrics including accuracy, precision, recall, and F1-score. Among all the models, the Deep Neural Network achieved the highest performance with an accuracy of 98%, outperforming the best ensemble-based model, XGBoost, which achieved 97%. The results demonstrate that deep learning can offer superior predictive capability for cardiovascular risk detection, making it a viable choice for real-world clinical decision-support systems. The system is implemented in Python with a Django web framework and a relational database, integrating data preprocessing, model training, evaluation, and a web interface for prediction. By combining traditional ensemble techniques with a deep neural network, the proposed system enhances the reliability of medical prediction and supports early detection and prevention of heart disease.*

Keywords—*Heart Disease Prediction; Deep Learning; Deep Neural Network; Ensemble Learning; XGBoost; Random Forest; Clinical Decision Support; TensorFlow.*

I. INTRODUCTION

As the leading cause of death worldwide, heart disease is a global health issue with significant socioeconomic consequences for patients, their families, and governments. Prediction models can accurately identify individuals with heightened susceptibility to cardiovascular disease through risk stratification; targeted interventions such as dietary modification and medication can then help lower that risk and support primary prevention. Machine learning (ML) plays a critical role in healthcare, where a wide spectrum of ailments can be identified, tracked, and forecast using data-driven methods.

Data-mining and ML methods for disease-risk prediction have seen dramatic growth, and prior research has demonstrated the feasibility of using these techniques for illness prediction. However, earlier efforts to forecast the likelihood of disease progression have not always yielded convincing results, motivating the development of more accurate cardiovascular-disease prediction tools.

The goal of this project is to build an accurate heart-disease prediction tool that takes structured clinical data as input and produces a reliable prediction of heart-disease presence. The work implements and compares five ensemble techniques—Random Forest, AdaBoost, Gradient Boosting, XGBoost, and LightGBM—alongside a deep neural network, evaluating all models on common metrics to determine the most effective architecture for clinical decision support.

II. LITERATURE SURVEY

Advances in medical science and technology, including the introduction of artificial intelligence, have led to significant developments in modern medicine and healthcare. Predicting how likely a patient is to develop heart failure is a major challenge, and healthcare organisations must now maintain enormous volumes of data, which makes interpretation difficult. ML and deep learning (DL) are popular options for classification tasks, using both images and conventional tabular data, and ML algorithms can be used to predict the onset of illness.

Deep neural networks with several hidden layers have shown superior performance over traditional artificial neural networks for prognostic categorisation of heart disease. Whereas traditional ML algorithms rely on manually extracted features, DL techniques learn features from the training data and tend to outperform them; recent architectures include RNNs, CNNs, LSTMs, and GRUs. AI systems such as C-CADZ have been proposed for coronary-artery-disease diagnosis, validated using datasets such as Z-Alizadeh Sani from the UCI repository, using feature-extraction methods such as factor analysis of mixed data and nature-inspired feature selection. The SMOTE method has been widely used to address class imbalance, and studies report that model accuracy improves when trained on a balanced dataset; among methods studied, reinforcement learning was the least precise and artificial neural networks among the most precise. Integration with the Internet of Medical Things further enables smart, cloud-connected healthcare.

TABLE I. REPRESENTATIVE METHODS AND TECHNIQUES

S.No	Approach	Technique	Note
1	Traditional ML baselines	KNN, SVM, RF, Naive Bayes	Depend on handcrafted features
2	Class-imbalance handling	SMOTE oversampling	Improves accuracy on balanced data
3	Coronary diagnosis (C-CADZ)	FAMD + nature-inspired selection	Validated on UCI Z-Alizadeh Sani
4	Deep architectures	RNN, CNN, LSTM, GRU	Learn features automatically

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

S.No	Approach	Technique	Note
5	Deep neural network	Multi-hidden-layer DNN	Outperforms traditional ANN
6	IoMT-integrated frameworks	Cloud + ML	Enables smart healthcare access

III. EXISTING SYSTEM AND PROPOSED SYSTEM

A. Existing System

The existing system uses traditional ML techniques such as K-Nearest Neighbors, Support Vector Machines, Random Forest, and Naive Bayes to predict heart disease based on the UCI Heart Disease dataset. These models depend on manually extracted features for classification, which limits their ability to handle complex datasets effectively, and statistical methods are used to address missing values and balance the dataset. Despite these efforts, traditional ML models face overfitting, underfitting, and reliance on engineered features, which reduce predictive efficiency.

Limitations of the existing system:

- Feature dependency: heavy reliance on handcrafted features limits performance.
- Class imbalance remains challenging even with techniques such as SMOTE.
- Overfitting can occur, particularly with neural networks, without proper regularisation.
- Limited ability to capture complex patterns in the data.

B. Proposed System

The proposed system develops an intelligent prediction framework for early detection of heart disease using both ensemble ML models and a deep neural network. It takes structured clinical data—age, sex, cholesterol, resting blood pressure, exercise-induced angina, and related attributes—and produces an accurate prediction. The pipeline comprises data collection and preprocessing (cleaning, encoding, normalisation, missing-value handling), ensemble model building (Random Forest, AdaBoost, Gradient Boosting, XGBoost, LightGBM), a DNN built with TensorFlow/Keras using multiple hidden layers and dropout regularisation trained with binary cross-entropy and the Adam optimiser, model evaluation and comparison on a common test set, and best-model selection.

Advantages of the proposed system:

- Improved accuracy by integrating ensemble ML and deep-learning techniques.
- Efficient data processing: missing-value handling, outlier management, normalisation.
- Comparative evaluation across six models on common metrics.
- DNN captures complex patterns, achieving the highest accuracy.
- Provides a reliable decision-support tool for clinicians.

IV. SYSTEM DESIGN AND METHODOLOGY

A. Methodology

The methodology follows a supervised-learning pipeline. A publicly available heart-disease dataset is collected and preprocessed: data is cleaned, categorical features are encoded, numerical features are normalised, and missing values are handled. The processed data is used to train the five ensemble models and the deep neural network. All models are evaluated on a common test set using accuracy, precision, recall, F1-score, and ROC-AUC, and the results are compared in tabular and graphical form so that the strongest model can be selected.

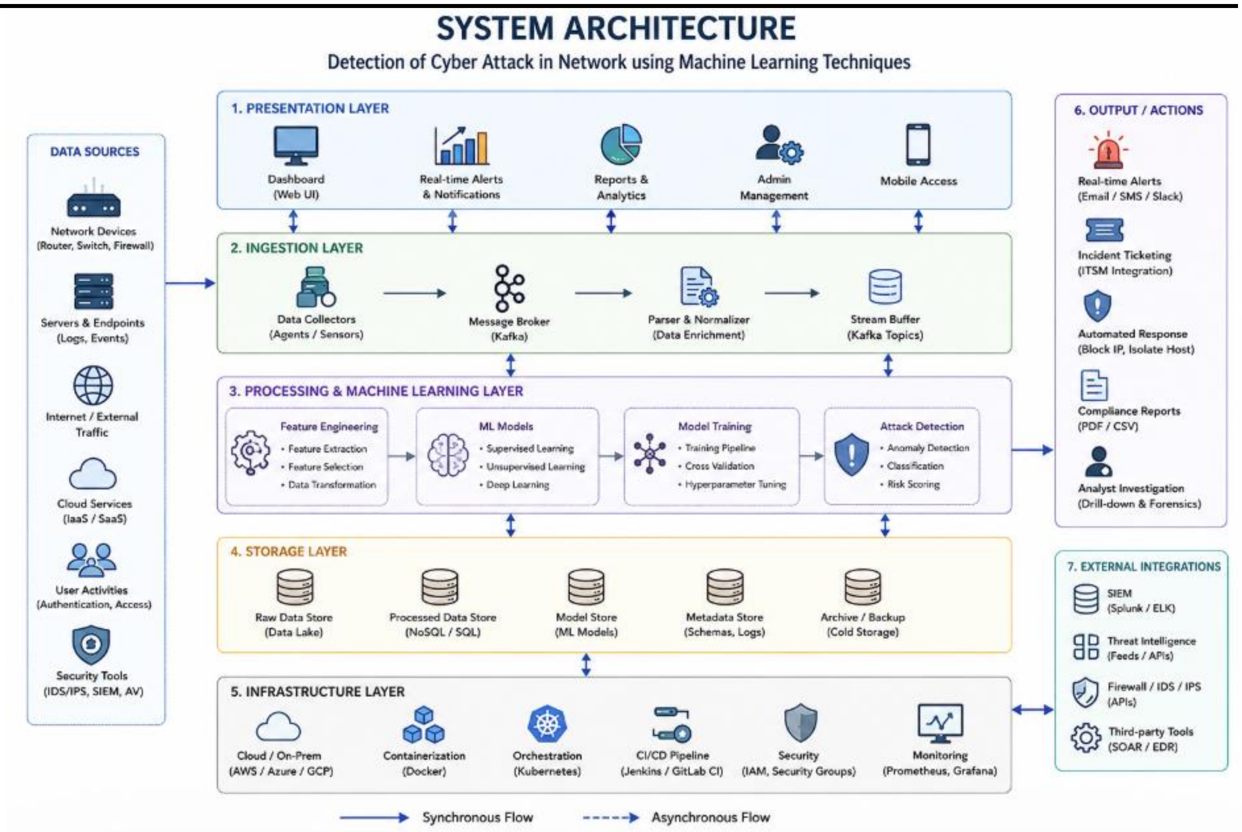
B. Deep Neural Network

The deep neural network is constructed using TensorFlow/Keras with multiple hidden layers and dropout regularisation. Input data is standardised, and the network is trained using the binary cross-entropy loss function with the Adam optimiser. The use of several hidden layers allows the DNN to learn discriminative features directly from the clinical data rather than relying on handcrafted features, addressing the feature-dependency limitation of traditional ML.

C. System Architecture

The system is organised as a layered web application. A presentation layer provides the user interface for entering clinical attributes and viewing predictions. An application layer, implemented with Django, coordinates preprocessing, model inference, and result presentation. A model layer hosts the trained ensemble and DNN models, and a data layer manages the dataset and stored records in a relational database. This separation supports maintainability and allows models to be retrained or replaced.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper



V. SYSTEM IMPLEMENTATION

A. Technology Stack

TABLE II. TECHNOLOGY STACK

Component	Technology / Tool
Programming Language	Python
Web Framework (Backend)	Django
Frontend Technologies	HTML5, CSS3, JavaScript
Deep-Learning Library	TensorFlow / Keras

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Component	Technology / Tool
Ensemble Models	Random Forest, AdaBoost, Gradient Boosting, XGBoost, LightGBM
Database	MySQL (SQL)
Deployment Server	WAMP / XAMPP; Django development server

B. Implementation Details

The implementation realises the methodology end-to-end. The preprocessing module cleans the dataset, encodes categorical attributes, normalises numerical features, and handles missing values. The ensemble module trains Random Forest, AdaBoost, Gradient Boosting, XGBoost, and LightGBM on the processed data. The deep-learning module builds and trains the DNN using TensorFlow/Keras with standardised inputs, dropout regularisation, binary cross-entropy loss, and the Adam optimiser. The evaluation module computes accuracy, precision, recall, F1-score, and ROC-AUC for every model and presents the comparison in tabular and graphical form. The Django web layer exposes the trained model so a user can submit clinical attributes and receive a prediction.

C. Model Comparison

All models are evaluated on the same held-out test set so the comparison is fair. The ensemble methods provide strong baseline performance, with XGBoost the strongest ensemble model. The deep neural network, by learning features directly from the data, captures complex relationships that the feature-dependent baselines miss, and is selected as the best-performing model for deployment in the decision-support interface.

VI. RESULTS AND DISCUSSION

The experimental results show that, while ensemble methods offer strong performance, the deep-learning model surpassed them all in predictive accuracy and robustness. The Deep Neural Network achieved an accuracy of 98%, compared with 97% for the best ensemble model (XGBoost), highlighting the DNN's superior ability to capture complex patterns within the clinical data. These results validate the efficacy of modern ensemble methods for medical prediction while confirming the growing significance of deep learning in healthcare analytics, and reinforce the importance of selecting the model architecture based on the nature of the dataset and the criticality of the task.

TABLE III. REPORTED MODEL PERFORMANCE

Model	Reported Accuracy
XGBoost (best ensemble model)	97%
Deep Neural Network (best overall)	98%

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

Beyond accuracy, the models were assessed using precision, recall, F1-score, and ROC-AUC on the common test set, with results compared in tabular and graphical form. The findings indicate that the DNN provides the most reliable predictions among the evaluated models on this dataset.

Representative screenshots from the prototype implementation:

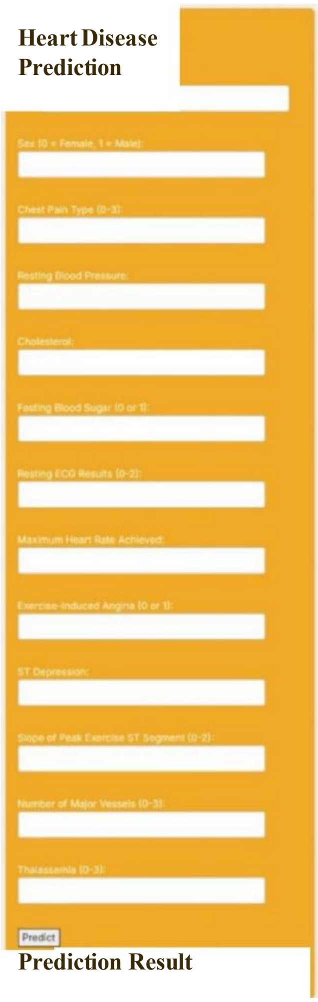


Fig. 1. Clinical-attribute input form.

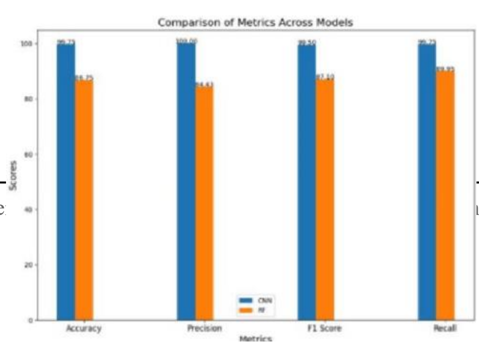


Fig. 2. Model comparison (accuracy and metrics).

VII. CONCLUSION AND FUTURE WORK

This project presented a comprehensive comparative study between ensemble machine-learning models and a deep-learning approach for the prediction of cardiovascular disease. Using a structured clinical dataset, five ensemble techniques—Random Forest, AdaBoost, Gradient Boosting, XGBoost, and LightGBM—were implemented and evaluated alongside a Deep Neural Network. The experimental results demonstrate that while ensemble methods offer strong performance, the deep-learning model surpassed them all, with the DNN achieving 98% accuracy compared with 97% for XGBoost. This analysis validates the efficacy of modern ensemble methods in medical prediction tasks and confirms the growing significance of deep learning in healthcare analytics, with the developed system serving as a reliable decision-support tool for the early detection and prevention of heart disease.

Future work can extend the system by incorporating larger and more diverse clinical datasets to improve generalisation, integrating additional deep architectures such as recurrent or attention-based models, adding richer explainability so that clinicians can understand the basis of each prediction, and deploying the system in a cloud or Internet-of-Medical-Things environment for real-time, connected healthcare. Continued validation against real clinical data and integration with hospital information systems would further improve practical applicability.

REFERENCES

[1] W. Ning, S. Li, D. Wei, L. Z. Guo, and H. Chen, “Automatic Detection of Congestive Heart Failure Based on a Hybrid Deep Learning Algorithm in the Internet of Medical Things,” IEEE Internet of Things Journal, vol. 8, no. 16, pp. 12550–12558, 2021.

[2] K. S. Gill, A. Sharma, V. Anand, and S. Gupta, “An Automated Coronary Artery Disease Diagnosis System Using Machine Learning,” in Proc. Int. Conf. Automation, Computing and Renewable Systems (ICACRS), 2022, pp. 1173–1178.

[3] B. Balakrishnan, K. Kumaran, A. Sankar, and K. Rao, “Coronary Artery Disease Prediction Using Machine Learning Algorithms,” in Proc. Int. Conf. Electronic Systems and Intelligent Computing (ICESIC), 2022, pp. 95–100.

[4] R. R. Sarra, A. M. Dinar, M. A. Mohammed, and K. H. Abdulkareem, “Enhanced Heart Disease Prediction Based on Machine Learning and χ^2 Statistical Optimal Feature Selection Model,” Designs, vol. 6, no. 5, 2022.

[5] S. Sharma and M. Parmar, “Heart Diseases Prediction Using Deep Learning Neural Network Model,” International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 3, pp. 2244–2248, 2020.

[6] R. Ganesh Babu, P. Karthika, and K. Elangovan, "Performance Analysis for Image Security Using SVM and ANN Classification Techniques," in Proc. 3rd Int. Conf. Electronics, Communication and Aerospace Technology (ICECA), 2019, pp. 460–465.

[7] I. Javid, A. K. Z. Alsaedi, and R. Ghazali, "Enhanced Accuracy of Heart Disease Prediction Using Machine Learning and Recurrent Neural Networks Ensemble Majority Voting Method," International Journal of Advanced Computer Science and Applications, vol. 11, no. 3, pp. 540–551, 2020.

[8] T. Zhang et al., "A Joint Deep Learning and Internet of Medical Things Driven Framework for Elderly Patients," IEEE Access, vol. 8, pp. 75822–75832, 2020.

[9] TensorFlow, "An End-to-End Open Source Machine Learning Platform." [Online]. Available: https://www.tensorflow.org/

[10] UCI Machine Learning Repository, "Heart Disease Data Set." [Online]. Available: https://archive.ics.uci.edu/