RESEARCH PAPER

Next-Generation Healthcare Analytics for Imbalance-Aware Early Coronary Heart Disease Prediction Using SMOTE-Enhanced Model

Nargis Rana Abdul Waheed Ansari[1], Dr. Abdul Razzaque[2]
[1]Department of Computer Science & Engineering, Anjuman College of Engineering & Technology, Nagpur, India.
Email: ansari.nargisrana@gmail.com
[2]Department of Computer Science & Engineering, Anjuman College of Engineering & Technology, Nagpur, India.
Email: arazzak@anjumanengg.edu.in

Abstract: Coronary heart disease (CHD) is a multifactorial issue with early prediction being a critical problem to address because of its high levels of imbalance in classes when using a large scale of healthcare data. This paper is a complete machine learning and deep learning model of CHD risk prediction based on population-level data on health records. An analysis using numerous exploratory features was done on a dataset of 246, 013 records with demographic, clinical, lifestyle and behavior attributes. To overcome the biased ratio of the CHD-positive and the CHD-negative data, Synthetic Minority Oversampling Technique (SMOTE) was introduced following the dataset division stage, leading to the balanced distribution of the training material. There are several classification models created such as Decision Tree (DT), Random Forests (RF), AdaBoost, and Voting Ensemble classifier which are compared with a proposed sequential neural network. The dropout regularization and the Adam optimizer were utilized to improve the generalization and convergence stability of the neural network architecture. Accuracy, precision, recall, F1 score, confusion matrices, and ROC-AUC curves were used to measure the quality of the models. The findings of the experimental prove that the proposed sequential neural network with SMOTE is superior to all the baseline models, with the accuracy and precision of 95.96, recall of 96.17, and F1 score of 96.16. The investigation outcomes prove that integration of imbalance-conscious learning and deep representation modeling provide a high degree of predictivity of CHD, with a valid and scalable abiding tool of the decision-support consideration of the initial cardiovascular risk assessment.

***Keywords*:** Coronary Heart Disease, Machine Learning, Healthcare Analytics, Risk Prediction, SMOTE
How to cite this article: Ansari NRAW, Razzaque A. Next-Generation Healthcare Analytics for Imbalance-Aware Early Coronary Heart Disease Prediction Using SMOTE-Enhanced Model. Int J Drug Deliv Technol. 2026;16(13s): 1043-1044.
DOI: 10.25258/ijddt.16.13s.114

I. Introduction

CHD is among the most common causes of morbidity, and mortality in the world, leading to high levels of strains on the healthcare systems worldwide. Although the medical technologies in diagnostic have made strides, CHD is usually diagnosed at late stages where its treatment becomes expensive and less effective [1-2]. It is thus necessary that those who are at high risk are identified early so that preventive care, lifestyle change and early clinical intervention can be applied. This fact is the growing access to large-scale post-hoc data on healthcare surveys, which creates new possibilities in risk prediction using data. These data combine health issues, lifestyle, demographic, and self-reported well-being. The complexity of these datasets, however, is high in terms of dimensionality, heterogeneity, and imbalance which puts a significant challenge to traditional predictive models. Current CHD prediction research commonly uses a small number of clinical variables or is centered on the traditional machine learning models [3-4]. These methods have problems that include difficulties capturing complex, non-linear interactions of divergent health determinants and have poor prediction capability because of the extremely high rates of class imbalance. The recent progress in the field of deep learning provides an opportunity to learn very complex relationships between heterogeneous features [5-6]. These models might have an enormous effect in enhancing early risk detection when coupled with effective imbalance management techniques. The research is inspired by the fact that a framework which is scalable, accurate and computationally efficient is required to assist with the early screening of CHD with the help of population-level-based healthcare data. The main dilemma that is solved in this article is the emerging strength of predictive structure that will guide the recognition of individuals at coronary disease risk following a highly imbalanced, big-scale healthcare dataset with mixed identifier numbers and categorical characteristics [7]. The conventional models are also inclined towards the majority group which will result in high false negative which cannot be tolerated in a clinical screening condition.

This research has three-fold objectives. The first is to conduct a complex exploratory analysis of clinical variables, lifestyle variables, and demographic variables to comprehend their association with CHD. Second, to compare between two methods of classical machine learning and ensemble approaches and a proposed deep sequential neural network (SNN) in equal data conditions. Third, to produce a next-generation predictive model which maximizes the recall and the F1 score thus helping in making early intervention and preventive healthcare decisions. This is a novel study because it has combined in one framework and presented the systematic work of the exploratory analysis of data, the strict management of imbalance through SMOTE and the delicate training of a deep sequential neural network. Compared to previous research, the current piece of work offers a close comparison of various algorithms with the same preprocessing and performs well with a low-computing expense on the GPU infrastructure.

II. Related Work

Rehman et al. [8] contrasted traditional classification forecasts and proposed a hybrid particle swarm optimization–artificial neural network model to predict CHD demonstrating a better prediction accuracy beyond the feature selection and SMOTE oversampling. But independent people are still dealt with in this work and do not adjust interactions between risk factors which may represent latent pathophysiological relationships. F. Asadi, et al. [9] compared the performance of traditional machine learning models RF, Logistic Regression (LR), Support Vector Machine (SVM), and eXtreme Gradient Boosting (XGBoost) with cardiovascular datasets, which in most cases, XGBoost is the most accurate model. These are good at classification, but do not have systems to elicit multi-factor relationships in large and heterogeneous data [10]. Comparative cardiovascular disease studies [11-12] used to apply a family of more conventional classifiers like SVM and Multilayer Perceptron (MLP) and validated variable performance across models but fail to give relational insights regarding risk interaction between demographic and lifestyle risk factors. The effectiveness of these older forms of machine learning is in at-risk classification but does not reflect the interdependency of predictors or give any network level guarantees about CHD risk. Traditional neural networks, and hybrid deep learning models used on electrocardiogram (ECG) and phonocardiogram (PCG) image demonstrate high resilience in identifying heart disease particularly in relation to image data. They, however, do not pay much attention to structured risk factor data that are known to be associated with lifestyle and demographic factors that also pertain to CHD [13-14]. Deep learning surveys like Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (BiLSTM) [15-17] work note that deep architectures are more sensitive and have higher Area under the curve (AUC), however, due to the lack of data and the large, required datasets they are not as stable with structured survey responses [18-19]. The deep learning models are most effective when the learned features can capture features of complexity, but features are not necessarily suited to non-images, tabular interaction between clinical and lifestyle risk factors without supporting architectures [20]. Classical risk scores such as Framingham or QResearch cardiovascular risk algorithm (QRISK) utilize the lineal risk factors but are ineffective in the nonlinear interaction and the use of new multi-domain data, such as comorbidity networks, lifestyle variables and immunization history.

TABLE I. Comparative Analysis of Existing Work

Reference	Method Used	Evaluation (Metrics + Values)	Key Contributions	Limitations
Z. Chen [21]	Logistic Regression, Random Forest, XGBoost with class weighting	RF: Spec 0.81%, F1 0.87%; XGBoost: Acc 0.73%; LR: AUC 0.77%, Spec 0.67%.	Compared tree vs. linear models; showed value of balanced prediction in older adults	Focus on limited age group; models do not explore complex feature interactions
S. Yuliasari [22]	RF, LR, SVM, DT, Naive Bayes Classifiers	RF: Acc 90.16%. LR: Acc 85.25%	Broad ML performance comparison; highlights RF strength	Uses general algorithm list; lacks deep analysis of feature relations
N. Nasution [23]	RF, SVM, LR with feature selection	RF: Acc 89.7%; SVM: Acc 87.0%; LR: Acc 84.2%	Assessed feature selection impact; confirmed RF robustness	Still limited dataset size; external validity not shown
A.R. Ilyas [24]	Random Forest, LR, SVM, MLP, XGBoost & ensemble models	RF: Acc 92.9%; MLP: Acc 94.2%; XGBoost: Acc 99.98%	Large, combined datasets; ensemble performance shown	Varies by dataset and model; limited interpretability discussion
H. O. Boll [25]	Graph Neural Networks (GNN) over Electronic Health Record (HER) data	Broad performance reported across tasks	Reviews how GNNs capture relational patient risk structures	Challenges in model interpretability & EHR heterogeneity

Existing studies are good baselines showing how classical models like RF, LR, and SVM perform on structured cardiac risk datasets. They consistently achieve high accuracy but do not model complex interactions among risk factors. Results from research with MLP, RF, and XGBoost highlight that deep learning and ensemble strategies can boost predictive power, but they still treat features independently rather than relationally. The others show that graph methods can model relationships among entities (e.g., comorbidities, risk factor links), but current work has limitations in interpretability and dynamics a gap to be addresses. No model is fully capturing complex patterns among clinical, lifestyle, and demographic risk factors. Mostly authors work on small or homogeneous cohorts e.g., UCI and BRFSS subsets dataset. High-performing models often lack insights a clinician would use which raise Interpretability issues. Dynamic and relational changes in patient risk are rarely modelled. This research demonstrate feature optimization and class balancing like SMOTE, hyperparameters can increase performance and do inherently address pattern interactions.

III. Proposed System

The proposed system is an end-to-end pipeline of healthcare analytics based on the early prediction of CHD. The process starts with large scale acquisition of healthcare data, which is then followed by systematic pre-processing and explored data analysis, data imbalance mitigation, predictive modelling and performance evaluation. The system architecture has four significant layers, conceptually, namely, data ingestion, data pre-processing and feature engineering, model learning, and evaluation and decision support. Raw surveys of healthcare information are initially ingested and passed through compliance and consistency checks. The dataset was pre-processed and encoded after which it is fed through a resampling module to counteract class imbalance. The same balanced feature space is then trained on multiple machine learning and deep learning models parallel to achieve fair comparison. Lastly, the output of the prediction is compared with clinically relevant metrics and visual diagnostics tools like Receiver Operating Characteristic (ROC) curves, and confusion matrices are created to aid in interpretability and making of decisions. To have a full assessment, the classical machine learning as well as the deep learning methodology was used. Such classifiers, DT, RF,

AdaBoost and Voting Ensemble are considered classical models and known to be able to interpret and perform well on tabular healthcare data. Moreover, deep SNN was suggested to detect non-linear and higher-order correlations between heterogeneous clinical, lifestyle, and demographic characteristics. Contrary to tree-based models, the neural network learns a hierarchical representation of the features, thus any complex dependency that is hard to represent with a tree can be modelled using the neural network. All models were trained and assessed in the same pre-processing and resampling environment to make them methodologically consistent. Figure 1 shows the overview of the proposed end-to-end healthcare analytics framework for early coronary heart disease prediction, illustrating data pre-processing, imbalance handling using SMOTE, model training, and performance evaluation.

The dataset is organized in the form of tables and is in English language, which makes it easy to carry out automated analytics. The dataset reflect high diversity in age groups, gender, health status, and behaviour issues, which allows analysing the risk of CHD at the population level. Records were chosen by being complete and relevant to cardiovascular health and thus important clinically predicting variables were included. Some of the data pre-processing activities included dealing with missing data, numerical attribute standardization, and that of categorical variables. Nominal features were coded using one-hot encoding to retain category information without incurring an ordinal bias. Feature scaling was also used where a balance in training the neural networks was needed.

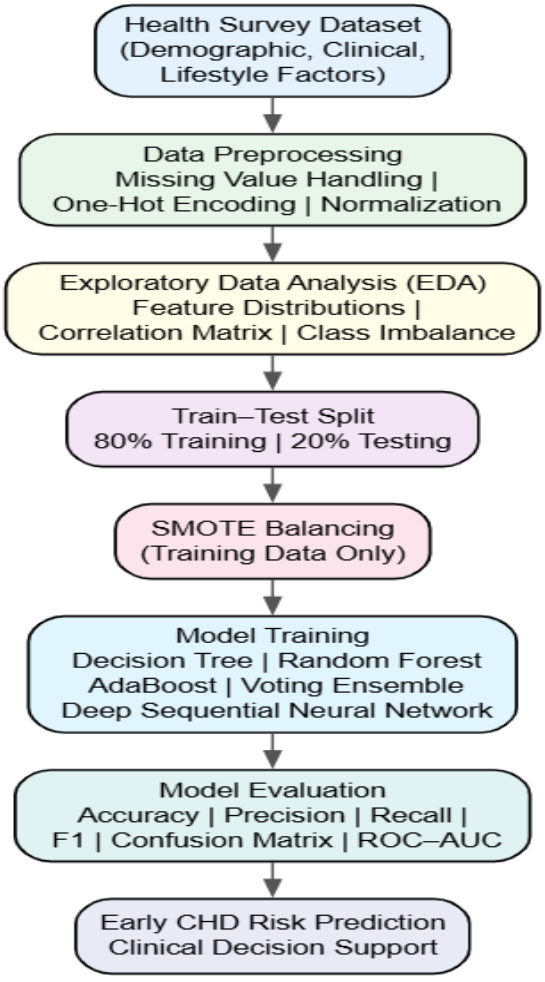

Figure 1. The proposed end-to-end CHD prediction framework

Let the dataset be represented as

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N} \quad [1]$$

A. Dataset Description and Collection

In this research, the dataset that will be handled is 246,013 individual records, derived because of a survey platform that handles large-scale healthcare. Each of the records is of one respondent and is equipped with demographics, clinical, lifestyle, and self-reported health indicators.

where $x_i \in \mathbb{R}^d$ denotes the feature vector of the i-th individual and $y_i \in \{0,1\}$ indicates the presence or absence of CHD.

For the proposed sequential neural network, each hidden layer performs a linear transformation followed by a non-linear activation:

$$h^{(l)} = f(W^{(l)}h^{(l-1)} + b^{(l)}) \quad [2]$$

where $W^{(l)}$ and $b^{(l)}$ are the weight matrix and bias vector of layer l, and $f(\cdot)$ is the ReLU activation function. The output layer uses a sigmoid function to estimate CHD risk:

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-z}} \quad [3]$$

Model optimization is performed by minimizing the binary cross-entropy loss:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad [4]$$

Exploratory data analysis (EDA) was performed to ascertain how the data in the dataset are distributed, vary and connected to each other before building the model. Numerical and categorical attributes were chosen as the elements that were visualized differently because of the mixed character of the data. Plots in the form of histograms with numbers like BMI, physical health days and mental health days and hours of sleep showed skewed distributions and long tails implying that there was variability of the health status of the population. Such trends promoted the application of non-linear models which have the capability of dealing with intricate distributions. In case of non-numeric and binary visualization of health indicating factors, the frequency-based visualization reflected a strong prevalence of unfavourable results in most cases of chronic conditions. Bar plots of categorical variables on the CHD outcome revealed backgrounds of significance in relation to smoking status, diabetes, angina, and depressive disorders. Additional evidence of moderate correlations instead of strong linear dependencies between numerical features and the CHD outcome was provided by a correlation matrix as in Figure 2 in the appendix that supports the necessity to use higher-order interaction learning models.

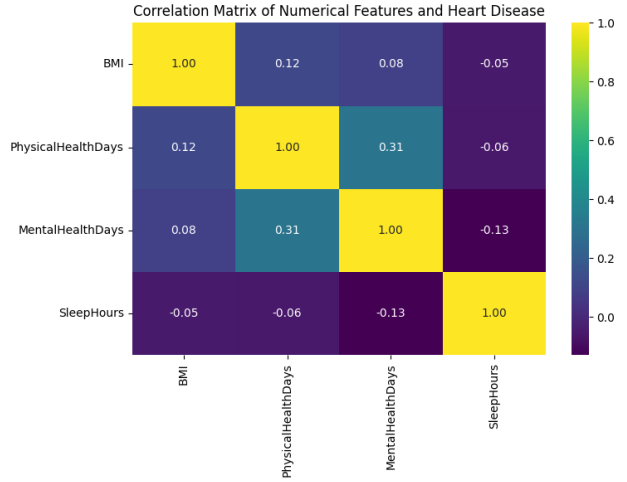

Figure 2. Correlation matrix of numerical features and the CHD outcome

B. *Handling Class Imbalance Using SMOTE*

The data acquired had a great skew in that the CHD-positive cases in comparison to CHD-negative cases. To overcome this, the SMOTE was implemented with the training. SMOTE creates synthetic minorities between existing minority examples resulting in equal class representation and better sensitivity. This would greatly reduce false negatives which is important in screening early diseases. Table I presents the class distributions of SMOTE and pre-SMOTE. Resampling succeeded in balancing the class representation of both training and testing datasets to allow learning and assessment of the model on equal grounds. This equalization moves greatly minimized the prediction bias towards the majority class and was more sensitive to cases of CHD-positive. Table II and Figure 3 show the class distribution of the *HadHeartAttack* variable before and after SMOTE. The original dataset shows a strong imbalance toward non-CHD cases, while SMOTE produces a balanced distribution in the training data, improving the model's ability to learn minority class patterns.

TABLE II. CLASS DISTRIBUTION OF THE DATASET VARIABLE BEFORE AND AFTER SMOTE.

Stage	Training (Yes)	Training (No)	Testing (Yes)	Testing (No)	Total
Before SMOTE	10,790	1,86,020	2,645	46,558	2,46,013
After SMOTE	1,86,020	1,86,020	46,558	46,558	4,65,156

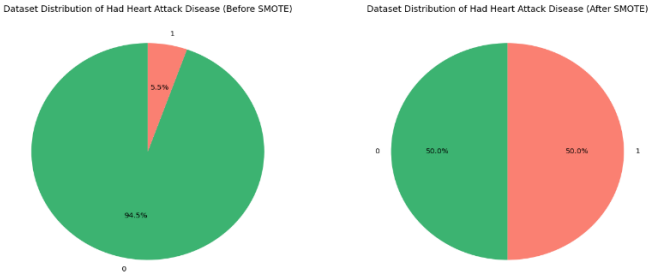

Figure 3. Class distribution of the *dataset* variable before and after SMOTE.

C. *Data Encoding and Dataset Preparation*

One-hot encoding was applied on all categorical variables to be compatible with machine learning and neural network models as well as to preserve category-specific information. The entire data was then divided into training and test data in proportion of 80: 20 so that unbiased examination will be made possible. Domain knowledge and exploratory data analysis were used to support feature engineering. Descriptive statistics were analysed in the form of histograms in the analysis of numerical distributions and the count and bar plots in the analysis of categorical variables. Correlation analysis contributed to finding out relationships between health indicators and the results of CHD. These studies were able to prove that CHD risk is not because of isolated variables but instead due to different interacting variables. Once when encoding, all the features were presented in dense numeric vectors that can be used by both a classical model and a neural network. Internal layer changes, which are proposed as implicit feature embeddings, are the mechanism, allowing the network to learn to acquire compact and discriminative representations of patient health profiles in the suggested sequential model.

D. *Hyperparameter Configuration and Tuning*

A deep SNN was suggested to expand the complex interactions amongst heterogeneous health factors. It has an architecture that comprises several connected layers, each of which has a reduced dimension, and this allows feature abstraction that can be hierarchical. The suggested network was used to train Adam optimizer with the learning rate of 0.0005, a batch size of 32, and 10 epochs. Parameters were chosen empirically through the validation and convergence during training. The successive layers had dropout rates of 0.3, 0.2 and 0.1 to prevent overfitting and use of the ReLU activation functions to learn the non-linear relationships. An output layer of sigmoid activation was applied to facilitate binary classification. All the models were written in Python, and classical algorithms were written in the scikit-learn library, and neural network was written in TensorFlow and Keras. Monitoring validation loss and accuracy were used to train the model in a stable manner. The experiments were done on a Kaggle cloud-based platform with an NVIDIA Tesla P100. The neural network was able to operate with low computational cost in 3 milliseconds per training step on average: thus, projects itself as computationally efficient and scalable with even large healthcare datasets.

IV. DISCUSSION OF RESULTS

To effectively evaluate the performance of the models in terms of initial prediction of CHD, several measures of evaluation, such as accuracy, precision, recall, F1-score, and ROC-AUC, were used. Accuracy gives an overall sense of correctness whereas precision shows the extent of reliability of positive predictions. The reason recall was taken as a primary measure is because it is of clinical value of minimizing false negative or missed high-risk patients. Precision and recall were optimized using the F1-score, and ROC-AUC was added to determine the discriminating capacity with different levels of decision thresholds. Proposed model had to be tested against popular classical and ensemble learning algorithms which served as benchmark. A comparison of the results is presented in a Table-III below.

TABLE III. CAMPARATIVE ANALYSIS OF PROPOSED MODEL AGAINST THE VARIOUS OTHER MODELS WITH DIFFERENT EVALUSTION MATRIC

Model	Accuracy	Precision	Recall	F1
Random Forest (No SMOTE)	0.8504	0.7984	0.5552	0.5869
Decision Tree (SMOTE)	0.8778	0.8808	0.8778	0.8776

AdaBoost (SMOTE)	0.9025	0.9029	0.9025	0.9025
Random Forest (SMOTE)	0.911	0.9188	0.911	0.9106
Voting Ensemble (SMOTE)	0.9308	0.9361	0.9308	0.9305
Proposed Deep SNN (SMOTE)	**0.9596**	**0.9623**	**0.9617**	**0.9616**

The following figures are all illustrative of the gradual enhancement in the forecasting of CHD. Figure 4 shows the confusion matrix of the model of Random Forest with no SMOTE trained on the original imbalanced dataset. Though the overall model achieves satisfactory accuracy, there are a high false negative rates, which implies that the model is poor in identifying CHD-positive cases. Clinically this is not a good habit because the timely intervention may be postponed in case of missed CHD cases. Figure 5 indicates the confusion of the Random Forest model that was trained with the use of SMOTE. The true positive rate has a marked improvement and the false negatives are reduced by a significant margin. This indicates the viability of synthetic oversampling to reduce the incidence of class imbalance and enhance the sensitivity of the model on CHD cases, despite a small cost in the number of false positives.

Figure 6 shows the confusion matrix of the proposed Sequential Neural Network (SNN) trained on the data balanced by SMOTE. The model has the greatest number of true positive and true negative in all experiments, which means that it has the best discriminative ability and balanced classification in both classes. The proposed SNN model has a ROC-AUC curve as illustrated in figure 7 that indicates a high area under the curve which Makes it clear that the model has good class separability and consistent performance on various decision thresholds. Figure 8 and 9 show the training and validation loss curve and accuracy curve of the SNN model respectively. The consistent decrease in loss curve and minimal overfitting as shown by the smooth end of the loss curve and the steady growth in accuracy without any steep decrease.

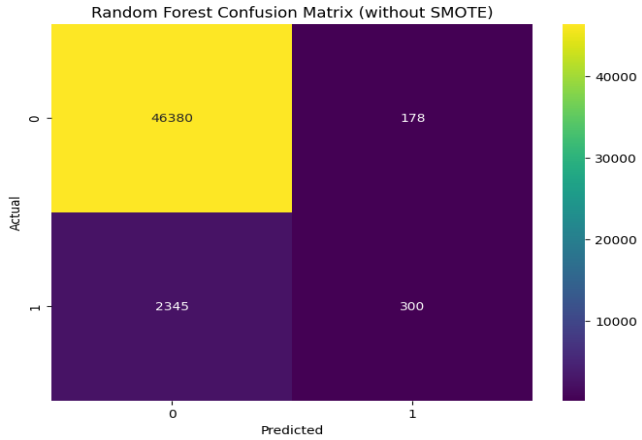

Figure 4. Random Forest Confusion Matrix (without SMOTE)

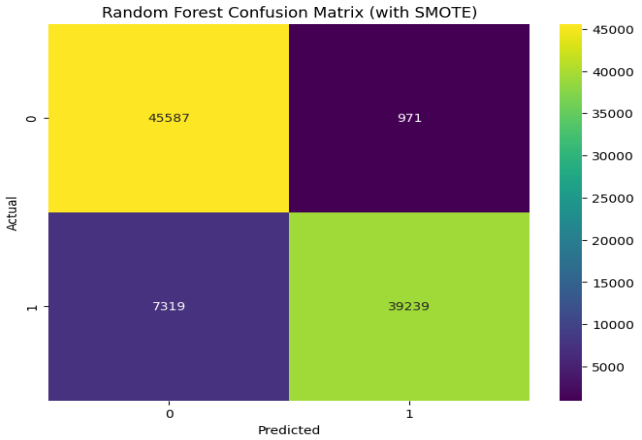

Figure 5. Random Forest Confusion Matrix (with SMOTE)

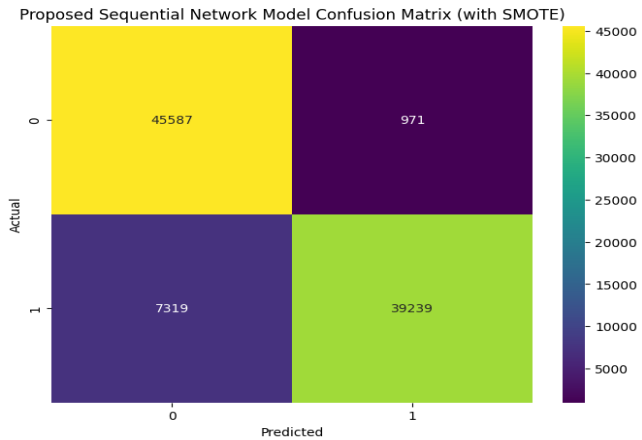

Figure 6. Proposed SNN model Confusion Matrix (with SMOTE)

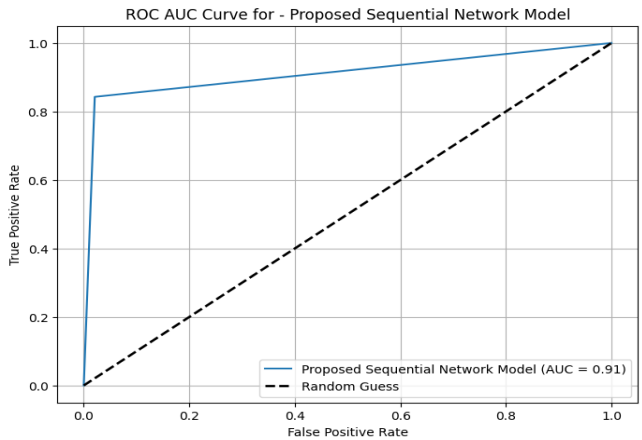

Figure 7. ROC AUC Curve for Proposed SNN Model

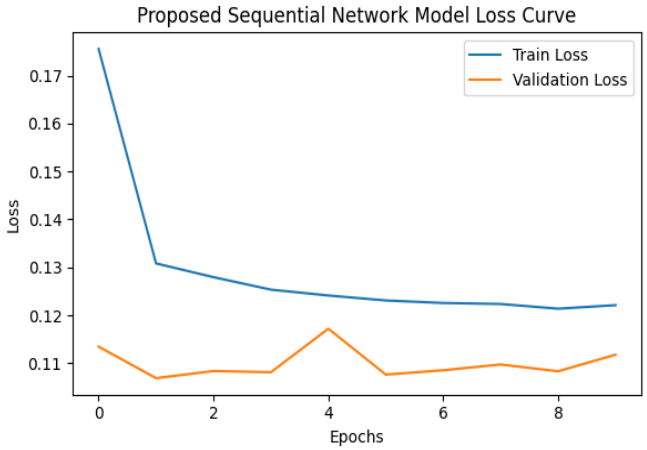

Figure 8. Proposed SNN Model Loss Curve

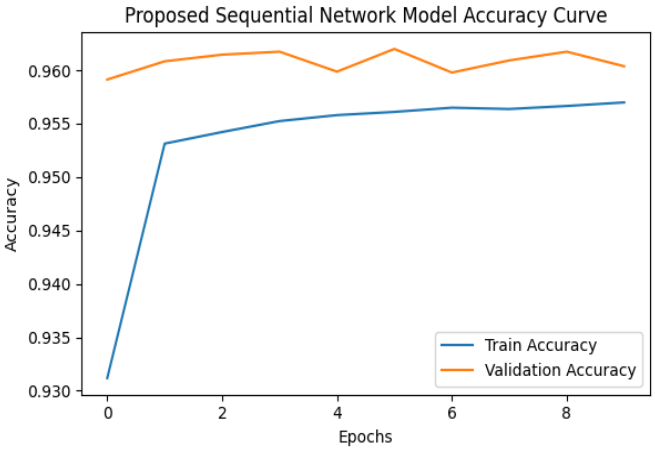

Figure 9. Proposed SNN Model Accuracy Curve

On the whole, these results confirm that the combination of SMOTE-imbalance control with an optimized sequential neural network can drastically increase the reliability of CHD prediction and clinical significance. The proposed sequential neural network consistently outperformed all baseline models across every evaluation metric, achieving the highest F1 score and recall, which are critical for early disease detection. To equalize and provide reproducibility, training and validation split were the same one applied to all models in case of hyper-parameter optimization. The values of the final parameter that are reported below are the result of the configuration with best validation performance. Tuning objective gave primacy to ROC-AUC and recall because these measurements are more reflective of the clinical need, i.e. the identification of high-risk CHD cases in their early stages. The findings prove that classical models have a major advantage in SMOTE-based balancing but are still seen to be weak in bidding on complex interactions among features. The ensemble approaches enhance resistance, yet at the cost of performance improvement. The proposed deep SNN is effective in approximating non-linear relationship between clinical, lifestyle, and demographic variables, which lead to the high predictive accuracy and sensitivity. The findings show that SMOTE balancing would provide a significant advantage to the performance of classical models, but such models are restricted in their ability to achieve higher-order interactions between features. The ensemble approaches enhance resistance, yet at the cost of performance improvement. On the other hand, the proposed architecture can effectively learn complicated relationships between heterogeneous risk factors and is better able to discriminate CHD and non-CHD cases. The reason is that the proposed model has a high recall, which means that it is highly suitable in early screening and preventing healthcare needs as it has a high capacity to identify CHD-positive individuals. The proposed framework is easy to compute and has a high performance which is why it is appropriate to deploy in large scale healthcare screening systems.

V. Conclusion and Future Scope

In this research, a next-generation framework of healthcare analytics on prediction of early CHD with large and skewed health data were presented. Using systematic EDA, successful data balancing, and a tuned architecture and the proposed method was much better at performing than traditional machine learning and ensemble models. This work is well organized and logically structured providing a pipeline of early prediction of CHD. The suggested framework has wide consideration in preventive and population health. It may be implemented as clinical decision support tool to detect high-risk persons to be screened and intervened early on lifestyle. The model can be applied by the public health agencies to examine the CHD risk trends at population level and institute specific prevention programmes. Moreover, the framework may be used to assist in the planning of individual healthcare by pointing out the prevalent risk factor interactions particular to individual patients or populations. The study is based on single dataset, also, though it has a strong performance, there is a risk of reporting bias. Future directions will focus on using graphs to learn to establish inter-patient similarity, combine longitudinal EHR, and make the framework more relatable to explainable AI methods to enhance clinical interpretability. Using wearable device real-time streams of information is also an interesting avenue of proactive monitoring of CHD risk. The multi-regional or longitudinal datasets would enhance the generalizability and allow the graph modelling of the disease development through time.

References

[1] N. Sharma, L. Malviya, A. Jadhav, and P. Lalwani, "A hybrid deep neural net learning model for predicting Coronary Heart Disease using Randomized Search Cross-Validation Optimization," Decision Analytics Journal, Vol 9, 100331, 2023.

[2] D.M. AlSekait, et al. "Using convolutional neural networks with late fusion to predict heart disease," Nature Scientific Reports, 15, 41260, 2025.

[3] Y. Mao, B. L. Jimma, and T. B. Mihretie, "Machine learning algorithms for heart disease diagnosis: A systematic review," Current Problems in Cardiology, Vol. 50, 8, 103082, 2025.

[4] Y. Rimal, et al. "Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy," Nature Scientific Reports, 15, 13444, 2025.

[5] 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," Nature Scientific Reports, 14, 23277, 2024.

[6] N. Chandrasekhar and S. Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization" *Processes*, *11*, 1210, 2023.

[7] H. Lu, and S. Uddin, "Disease Prediction Using Graph Machine Learning Based on Electronic Health Data: A Review of Approaches and Trends," *Healthcare*, *11*, 1031, 2023.

[8] M.U. Rehman, S. Naseem, et al. "Predicting coronary heart disease with advanced machine learning classifiers for improved

cardiovascular risk assessment," Nature Scientific Reports, 15, 13361, 2025.

[9] F. Asadi, R. Homayounfar, et al. Detection of cardiovascular disease cases using advanced tree-based machine learning algorithms. Nature Scientific Reports, 14, 22230, 2024.

[10] C. M. Bhatt, P. Patel, T. Ghetia, P.L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," *Algorithms*, *16*, 88, 2023.

[11] H. A. Al-Alshaikh, et al. "Comprehensive evaluation and performance analysis of machine learning in heart disease prediction," Nature Scientific Reports, 14, 7819, 2024.

[12] A. Hussain, and A. Aslam, "Cardiovascular Disease Prediction Using Risk Factors: A Comparative Performance Analysis of Machine Learning Models," Journal on Artificial Intelligence, 6, pp. 129-152, 2024.

[13] S. Dhandapani, H. Somasundaram, and T. Angamuthu, "Hybrid deep learning framework for heart disease prediction using ECG signal images," Nature Scientific Reports, 15, 33922, 2025.

[14] I. Gupta, A. Bajaj, and V. Sharma, "Comparative analysis of machine learning algorithms for heart disease prediction," International Journal of Hybrid Intelligent Systems, 21(1): pp. 14-28, 2025.

[15] M. A. Naser, A.A. Majeed, M. Alsabah, T.R. Al-Shaikhli, and K.M. Kaky, "A Review of Machine Learning's Role in Cardiovascular Disease Prediction: Recent Advances and Future Challenges," *Algorithms*, *17*, 78, 2024.

[16] S. N, Prajwalasimha, et al., "Hybrid Transformer–CNN Neuro-Symbolic Explainable AI for Cyber Threat Intelligence: Advancing Transparency and Adversarial Robustness," In 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), Coimbatore, India, pp. 492-499, 2025.

[17] A. Pimpalkar, and J. R. Raj, MBiLSTMGloVe: Embedding GloVe Knowledge into the Corpus using Multi- Layer BiLSTM Deep Learning Model for Social Media Sentiment Analysis, Expert Systems With Applications, Vol 203, 117581, 2022.

[18] H. Hajishah, et al. "Evaluation of machine learning methods for prediction of heart failure mortality and readmission: meta-analysis," BMC Cardiovascular Disorders, 25, 264, 2025.

[19] K. Raman et al. "A comprehensive review of machine learning for heart disease prediction: challenges, trends, ethical considerations, and future directions," Frontiers in Artificial Intelligence, Sec. Medicine and Public Health, Vol. 8, 2025.

[20] A. Pimpalkar, *et al.* "Fine-tuned deep learning models for early detection and classification of kidney conditions in CT imaging," Nature Scientific Reports, **15**, 10741. 2025.

[21] Z. Chen, "Heart Disease Prediction Models Performance Analysis based on Logistic Regression, Random Forest and XGBoost", HSET, vol. 153, pp. 115–124, 2025.

[22] S. Yuliasari and A. Rahmatulloh, "Performance Analysis and Accuracy of Machine Learning Algorithms for Heart Disease Prediction," Telematika, vol. 22, no. 3, pp. 98–106, 2025.

[23] N. Nasution, M. A. Hasan, and F. Bakri Nasution, "Predicting Heart Disease Using Machine Learning: An Evaluation of Logistic Regression, Random Forest, SVM, and KNN Models on the UCI Heart Disease Dataset", IT Journal Research and Development, Vol. 9, 2, pp. 140–150, 2025.

[24] A. R. Ilyas, S. Javaid, and I. L. Kharisma, "Heart Disease Prediction Using ML," *Engineering Proceedings, 107*, 124, 2025.

[25] H. O. Boll, and A. Amirahmadi, et. al., "Graph neural networks for clinical risk prediction based on electronic health records: A survey," Journal of biomedical informatics, 151, 104616, 2024.