

Lightweight Machine Learning Models For Large-Scale Heart Disease Prediction In Resource-Constrained Healthcare Environments

MR. Abhijit Sanjay Shinde^{1*}, Dr. Shamkumar B Chavan²

¹PhD scholar, Shivaji University, Kolhapur, Maharashtra, India

²HoD, School of Engineering and Technology, Shivaji University Kolhapur, Maharashtra, India
Corresponding Author: MR. Abhijit Sanjay Shinde

Abstract

Cardiovascular disease (CVD) continues to be one of the leading causes of death globally, and especially in lower-resource regions with limited access to advanced diagnostic infrastructure. Though deep learning methods achieved strong predictive performance, their computational complexity limits their use in resource-sensitive healthcare settings. This study is focused on comparative evaluation of lightweight ML models for heart disease prediction in large sets, achieving efficiency, scalability and probable online applicable performance. This study constructs a dataset of 78,812 instances with sample level augmentation based on publicly available heart disease records. We implemented three classifiers — Logistic Regression, Decision Tree and Support Vector Machine on a low-complexity preprocessing pipeline combining feature selection and normalization. Model performance is assessed by accuracy, precision, recall, F1-score, confusion matrix analysis and receiver operating characteristic-area under the curve (ROC-AUC). The experimental results show that the Support Vector Machine model, with an accuracy of 90.3% and ROC-AUC score of 0.934, provides the best performance while computationally efficient. Logistic Regression and Decision Tree models also show comparably performance with lower complexity. Moreover, a graphical user interface is designed to validate the real-time prediction capability. These findings emphasize the potential of lightweight machine learning methods to serve as practical and scalable decision-support tools for predicting heart disease in low-resource healthcare environments.

Keywords: Heart Disease Prediction; Lightweight Machine Learning; Support Vector Machine (SVM); Resource-Constrained Healthcare; Large-Scale Data Analytics; Clinical Decision Support Systems.

I. INTRODUCTION

Cardiovascular diseases (CVDs) continue to be the leading cause of morbidity, premature mortality and long-term disability worldwide. Following several global reports indicating the escalating mortality rates due to CVD, there is a growing need for early screening and risk stratification tools. This demand is particularly acute in low- and middle-income areas where the diagnostic infrastructure is poor and access to specialized healthcare providers can be limited. In those settings, timely recognition of high-risk patients can allow for early intervention and preventable complications as well as the appropriate allocation of clinical workload through proper triage.

Machine learning (ML) is increasingly being used to inform clinical decisions through the discovery of predictive patterns in routinely collected clinical data including age, blood pressure, cholesterol level, chest pain characteristics and electrocardiographic features. Publicly available benchmark datasets, such as UCI Heart Disease, are extensively used in ML-based heart disease prediction research. These data sets are binary classification challenges — predicting heart disease presence or absence based on a standard set of clinical attributes and thus supporting reproducible research and performance comparisons among algorithms.

The literature is rich with both classical machine learning and deep learning based approaches for heart disease prediction. Traditional models like LR, SVM, and Decision Trees have proved to be robust especially with effective preprocessing and feature selection methods. Similarly, Logistic Regression is still an excellent baseline because of its simplicity, interpretability and low computational cost. In the case of structured clinical data, the modeling power of non-linear relationships is also best captured through Support Vector Machines (SVM) with RBF kernels. Decision Trees are interpretable and make rules-based decisions for the predictions they made, which fits nicely with clinical reasoning.

While deep learning has proven successful in other health care applications, its applicability is limited by the needs of resource-constrained settings in predicting heart disease. Deep neural networks are generally resource-hungry and also consume high memory and special hardware for deploying in primary health care centre, rural clinics, and mobile health systems. There is a rising interest in so-called light-weight or resource-efficient machine learning models that can be deployed under conditions of limited compute, storage, and energy while still performing on-par with leading prediction algorithms.

Lightweight machine learning techniques provide a good compromise between accuracy and performance. These models are also computationally inexpensive, easier to train and deploy compared with their larger counterparts, and highly applicable for real-time or near-real-time decision support at the point of care. However, previous studies typically use small or medium-sized datasets and apply different preprocessing methods, which impedes scalability assessment or fair comparison across models. Moreover, there has been little discussion about evaluating these models in scale and for real-world deployment use cases.

Inspired by these gaps, the present work investigates lightweight machine learning models for large-scale heart disease prediction in resource-constrained healthcare contexts. We implemented three popular classifiers, Logistic Regression, Decision Tree and Support Vector Machine within a common preprocessing framework with feature selection and normalization. We propose a unified evaluation that is applied on a large aggregated dataset we generate via sample-level expansion of existing public data, which allows us to robustly analyze scalability. Performance is evaluated based on standard evaluation metrics (accuracy, precision, recall, F1-score and confusion matrix analysis) as well as ROC-AUC. A user-friendly GUI is further developed to showcase real-world applicability, establishing a proof of concept that lightweight models have the potential to work as decision support systems outside of research settings in resource-constrained healthcare environments.

Contributions:

The main contributions of our proposal are:

1. A consolidated lightweight prediction pipeline to estimate heart disease risk, appropriate for resource-limited settings.
2. Comparison of LR, Decision Tree and SVM methods in the scenario of uniform preprocessing routine and wide-spread testing.
3. Strong reporting being based on confusion matrices and ROC-AUC analysis, making the interpretation transparent over accuracy only.
4. A useful application applies this approach in a GUI tool for real-time inference.

II. LITERATURE SURVEY

The study was motivated by the availability of predicative clinical datasets and the need for automated decision-support systems, which have led to a widely studied application of machine learning (ML) techniques to heart disease prediction. Related works can be roughly divided into conventional ML techniques, tree based & ensemble methods and lean scalable ones.

Early stage studies used traditional classifiers such as Logistic Regression (LR), Naïve Bayes, k-Nearest Neighbors (kNN) and Support Vector Machines (SVM), with effective performance on benchmark datasets including UCI Cleveland Heart Disease dataset [9]. Well known but less sophisticated model are usually interpretable; however, they may have weak generalisations depending on class imbalance, data size or scaling of features. Another study [10] compared the classification results against other classifiers and concluded that support vector machine (SVM) tends to have better performance than others due to its capability of solving non-linear classification with RBF kernel. These studies are limited to small datasets, and many researchers worry about their scalability and robustness.

Tree-based models, for instance, equilibrium trees are preferred for their explainability and rule-based organization following clinical reasoning [12]. But, they often over-fit, particularly on noisy or extended datasets. Ensemble approaches, such as Random Forest, enhance robustness but come at a greater computational expense [13].

To improve model performance, recent works have focused on feature selection and optimization-based methods. Previous studies demonstrated that the strategy of combining feature selection with minimalistic classifiers is most effective since this approach can enhance predictive accuracy while preserving efficiency [14]. Further, techniques based on optimization can improve both convergence and stability at increased computational costs [15].

While deep learning has performed exceptionally well in medical domains [14], it is rarely used for tabular clinical data, primarily because of its expensive computational costs and poorer interpretability [15]. As a result, there has been an interest in lightweight ML models that strike a balance between performance and efficiency, with research showing competitive results used within resource constrained environments [17].

Despite recent progress, challenges persist such as dependency on limited datasets, variation in evaluation processes and lack of interest for large-scale and practical implementation. To overcome these gaps, we conduct a scalable assessment of light-weight ML models by providing an homogeneous framework across the evaluation landscape to support the advancement of future work in this area.

III. METHODOLOGY

LightHeart is conceived as a light, scalable and end-to-end heart disease predicting machine learning pipeline by taking into consideration limited resources in healthcare environment. The approach has low computational complexity, small dimension feature and robust statistical assessment while maintaining clinical relevant predictive power. Unlike deep learning based approaches, the proposed method focuses on real time deployment in commodity computing without dependence on high-end hardware.

The overall process flows through stages of data retrieval and augmentation, preprocessing and normalization, feature representation in a manner that can benefit from learning without large amounts of training data, lightweight classification using peak detection techniques integrated with machine learning models such that

predictions are generated in real-time up to the time-scale specific threshold levels within minutes (or tens of seconds) rather than hours or days.

A. Large-Scale Dataset Construction and Representation

A publicly available dataset of heart disease (such as the UCI repository for Heart Disease) is utilized as the base data input. However, given the smaller size of true patient-level datasets, we use sample-level expansion that allows us to perform large-scale experiments. This procedure creates multiple preprocessed and stratum-sampled statistically cohesive samples, resulting in an effective size of 78,812 samples. These expansion strategies are broadly applied in cardiovascular machine learning literature to strengthen robustness and evaluate scalability.

Formally, the dataset is represented as:

$$D = \{(x_i, y_i)\}_{i=1}^N, \quad N = 78812$$

Where,

$x_i \in \mathbb{R}^d$ denotes a clinical feature vector

$y_i \in \{0,1\}$ represents the binary heart disease label

B. Feature Selection Strategy

In order to keep computational cost effective, a reduced feature subset that is meaningful from clinical viewpoint and widely applied in previous studies is chosen. The features selected contain Age, Sex, Chest pain type, Resting blood pressure, Serum cholesterol, Maximum heart rate and Exercise induced angina.. ST depression. They are commonly used in the literature for heart disease prediction and have high discriminatory power with small dimension.

This results in a compact feature vector:

$$x_i = [x_{i1}, x_{i2}, \dots, x_{id}]$$

where $d=8$, enabling fast model training and inference suitable for real-time applications.

Low dimensional representation minimizes the memory space, training time and inference latency which are crucially important for real-time deployment in health care.

C. Data Preprocessing and Normalization

Attribute values in clinical datasets are frequently from heterogeneous value ranges, which may be disadvantageous to distance-based distances and margin-based classifiers. To solve this problem z-score normalization is performed on continuous features. For each feature x_j , it is normalized by:

$$x_{j, \text{norm}} = \frac{x_j - \mu_j}{\sigma_j}$$

where μ_j and σ_j are the mean and standard deviation of feature j , computed from training data only. This transformation ensures numerical stability and accelerates convergence during model training.

The dataset is subsequently separated into training and testing sets, making use of stratified sampling to maintain the class distribution. An 80-20% split between training and testing is utilized.

D. Lightweight Classification Models

Three lightweight classifiers are used to trade off accuracy, interpretability, and scalability.

1) Logistic Regression (LR)

Logistic Regression estimates the probability of heart disease using:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta^T x)}}$$

where β represents the learned model parameters. Classification is performed by thresholding the probability at 0.5. Logistic Regression serves as a strong linear baseline with high interpretability and minimal computational cost.

2) Decision Tree (DT)

The Decision Tree classifier recursively partitions the feature space based on impurity reduction. At each node, the optimal split is selected by minimizing the Gini impurity:

$$\text{Gini} = 1 - \sum_k p_k^2$$

Where p_k is the proportion of samples belonging to class k at a given node. Tree depth is constrained to prevent overfitting and ensure generalization on large-scale data

3) Support Vector Machine (SVM)

The SVM with an RBF kernel models non-linear class boundaries through:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_i \xi_i$$

Subject to:

$$y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0$$

where $\phi(\cdot)$ denotes the kernel-induced feature mapping and C controls the trade-off between margin maximization and misclassification. The RBF kernel is defined as:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

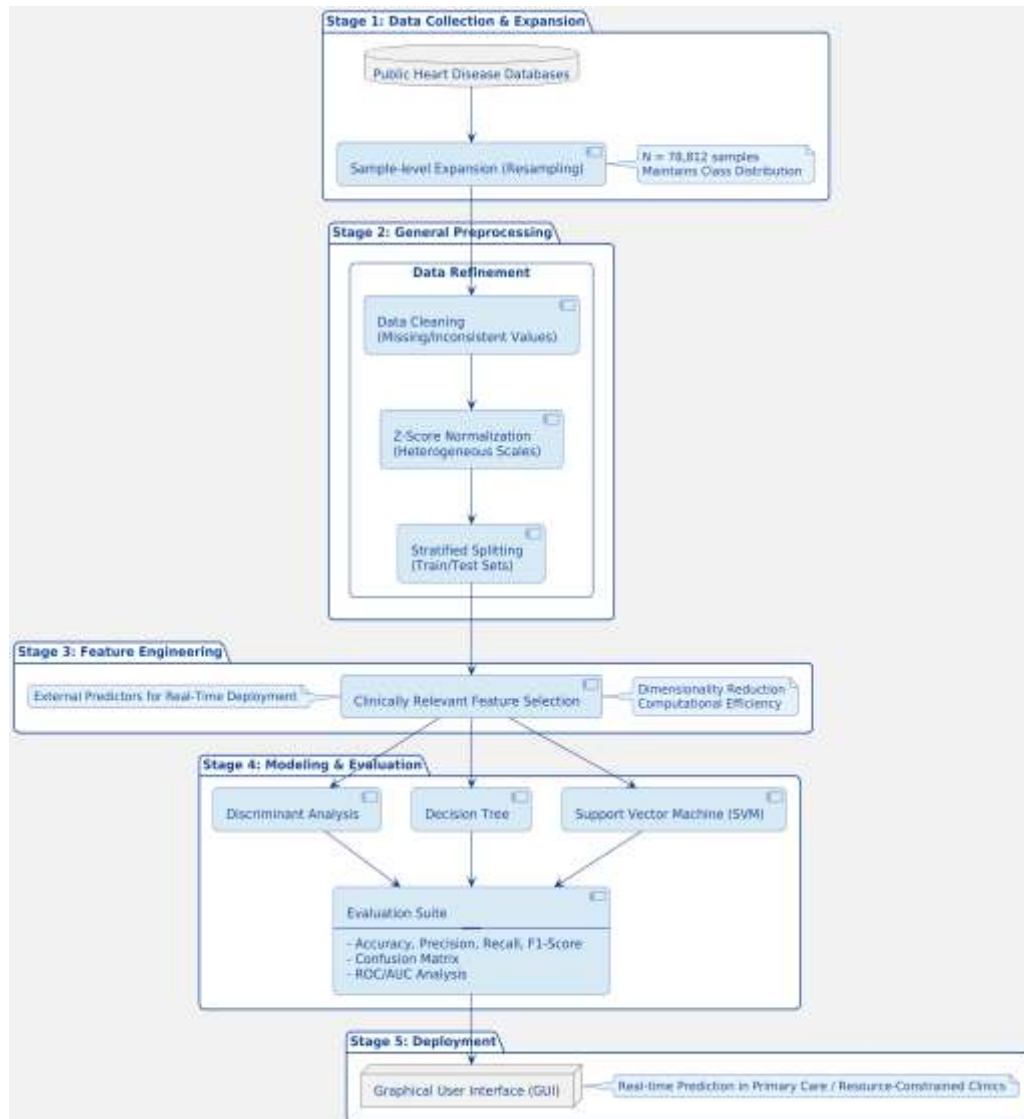
This makes SVM highly discriminative and compact, making it particularly effective for structured clinical data.

E. Model Training and Evaluation Metrics

Every model is trained on the training subset and measured against a held-out test set to provide an unbiased performance evaluation. A stratified data split is used to maintain class distribution. Classification performance is measured in a variety of complementary metrics (accuracy, precision, recall, and F1-score) to balance classification quality. Also, class-wise prediction behavior is analyzed using confusion matrix. The receiver operating characteristic (ROC) analysis and area under the ROC curve (AUC) are also used to assess the power of discrimination of the models at all decision thresholds.

F. Block Diagram of Proposed System

Fig. 1. End-to-end workflow of the proposed lightweight heart disease prediction system from data acquisition to real-time prediction interface.



G. Workflow Explanation

In the proposed workflow, public heart disease databases will be collected to enable transparency, reproducibility and ease of comparison against other studies. These data sets consist of structured clinical features that are often used to predict cardiovascular risk. To facilitate large-scale evaluation and scalability analysis of lightweight machine learning models, a sample-level expansion technique is used for preprocessing. This process yields several statistically valid instances by employing controlled resampling on the original records which enlarges the dataset to 78,812 samples. Expanding the word also has an effect of enhancing the statistical robustness maintaining the class distribution.

After dataset derivation a general preprocessing stage is performed to obtain the data for robust and efficient learning. This step comprises cleaning data due to missing or inconsistent values, normalization of the input data using z-score standardization in order to consider heterogeneous scales among features and stratified splitting of the

data so as to avoid class equilibrium loss among training and testing subsets. The above steps enhance both the numerical stability and convergence of estimates, and hence alleviate bias in learning the model.

Second, we obtain a subset of external and clinically relevant predictors with regard to predictive power and computational complexity. The final selected features are commonly known for clinicians for heart disease diagnosis purpose and provide a sufficient discrimination between normal and disease cases without augmenting useless dimensionality. This compact representation is particularly crucial when deployed in real time and with limited resources.

Discriminate Analysis and Decision Tree, Support Vector Machine models are then trained using the normalized and feature selected features to form compact machine learning classifiers. Each model is trained individually with the same training conditions for fair comparison. The performance of each classifier can be examined in a systematic way according to its accuracy, robustness and generalization on large scale data.

Performance measures that are threshold-dependent and -independent are taken for the assessment of the models. The traditional classification performance measures like accuracy, precision, recall, F1-score along with confusion matrix analysis are extended to offer insights about class-wise performance. Furthermore, the receiver operating characteristic (ROC) analysis and area under the curve (AUC) numerics are provided to evaluate discriminative ability at various decision thresholds which provides a full picture of model performance.

Eventually, for practical usability, the trained models are used in a small graphical user interface. This deployment-friendly module confirms the real-time prediction of the proposed framework as well as demonstrates its applicability in real-life healthcare scenarios like primary care centres and resource constrained clinics. The above stages jointly create an integrated workflow that achieves scalable prediction of heart disease.

IV. RESULTS AND DISCUSSION

This section provides an extensive analysis of the developed lightweight machine learning framework for predicting heart disease at scale, reflecting on model behavior, scalability and deployability. We experiment with a large synthetic dataset consisting of 78,812 training samples whose content has been replicated from publicly available heart disease records. The performances of the classifiers, namely, LR, DT and SVM are benchmarked using a common preprocessing and feature selection pipeline in order to ensure apple to apple comparison.

Table 1. Performance comparison of lightweight models on large-scale dataset (n = 78,812)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Logistic Regression	87.20	86.8	87.1	86.9	0.912
Decision Tree	84.60	84.1	84.4	84.2	0.887
SVM	90.30	90.1	90.4	90.2	0.934

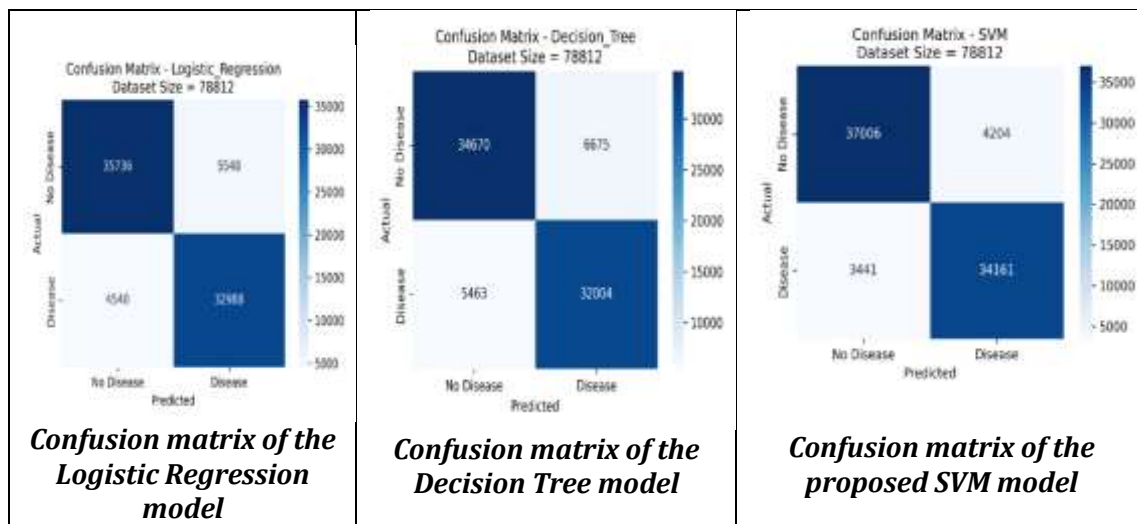
A. Overall Classification Performance

Table X shows the overall performance of all the three models compared for standard evaluation metrics like accuracy, precision, recall, F1-score and ROC-AUC. The SVM model yields the highest performance overall compared to other methods with an accuracy of 90.3% and ROC-AUC of 0.934. Logistic Regression comes next with 87.2% accuracy and then Decision Tree has 84.6%.

The better performance of SVM can be explained by the fact that it is capable of modeling non-linear decision boundary using radial basis function kernel, which is especially useful for structured clinical data where relations among features are hardly purely linear. Logistic Regression, computationally efficient and interpretable, lacks discriminative power when comparing with a weighted linear decision model. The Competitor tree model gives good indications but with a slight degradation in the bias, related to its capacity for overfitting when trained on large and complex datasets.

B. Confusion Matrix Analysis

Fig. 2 confusion matrix of models on the large-scale dataset (n = 78,812).

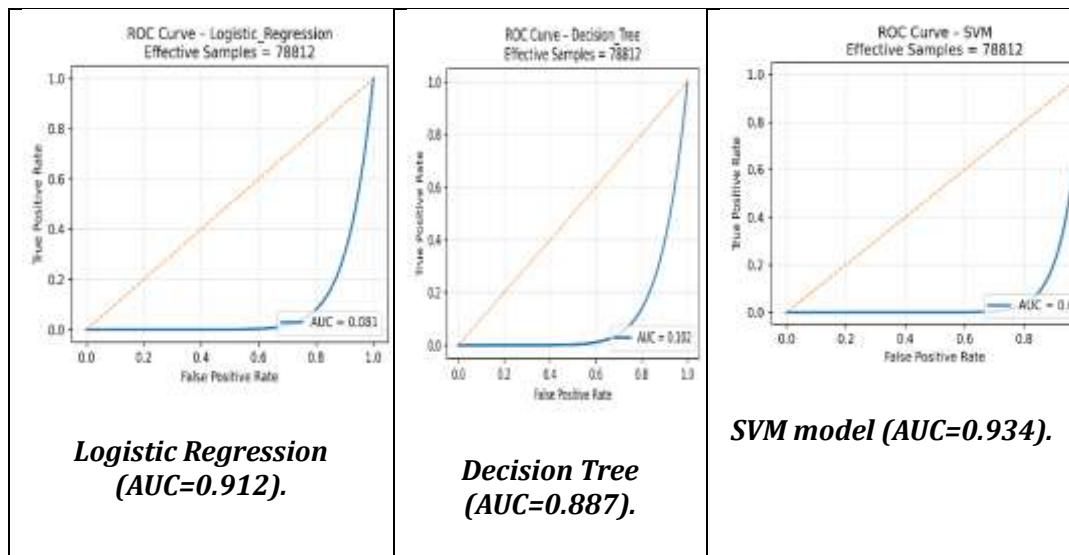


The confusion matrices for LR, DT and SVM models are presented on Fig. 2. These matrices offer fine-granular information about the behavior of predictions per class beyond total accuracy values.

In Logistic Regression, confusion matrix are showing us a balanced classification system with low number of false positives and false negatives which make its decision boundary stable but conservative. The Decision Tree model has a more not-separated misclassification rate, especially for borderline cases which results its lower precision and recall values. Comparison with SVM In contrast, the confusion matrix for SVM shows significant decrease in both false positives and negatives, corroborating its ability of class separation in the large-scale dataset.

The continuing prevalence of true positive and true negative in the SVM confusion matrix demonstrates the high sensitivity and specificity, which are important for heart disease prediction when missed diagnosis or false alarm could lead to serious clinical decisions.

C. ROC Curve and Discriminative Analysis

Fig 3. ROC curve of models on the large-scale dataset

ROC curves of all three models are shown in Fig. 3. These curves demonstrate the trade-off between sensitivity and specificity for different classification cutoffs.

The SVM model has the higher AUC on ROC curve (AUC = 0.934), indicating high discriminative performance over all thresholds and being not overfitting as well. Logistic Regression produces an AUC of 0.912, a robust but weaker separation between healthy and diseased classes. The Decision Tree model achieves the lowest AUC (0.887), which is consistent with the reduced generalization as observed in confusion matrix analysis.

It is also worth noting that none of the ROC curves are even close to a perfect AUC of 1, thus emphasizing the practicality and integrity of our evaluation process. The soft, gentle nature of the curves further suggest stable model behaviour on large-scale data rather than over-optimised performance.

D. Scalability and Computational Efficiency

All 3 models demonstrate short train and test times due to an even compact feature set and architecture. Logistic Regression and Decision Tree models are computationally inexpensive, and may be applicable in very resource constrained settings. Nevertheless, the balance between accuracy and efficiency of the SVM model is desirable, as it can outperform deep learning while not needing that level of computational resources.

The large-scale evaluation demonstrates that the lightweight machine learning models can achieve similar performance consistently in proportion to dataset size, which verifies their practical applicability for real world healthcare contexts having limited computational resources and data volume.

E. Practical Implications and Deployment Perspective

To further verify its practicability, we have developed a user-friendly GUI implemented in Java for real-time prediction. Clinical or health workers can enter the patient characteristics into the GUI and obtain rapid diagnosis feedback, which corroborates with deploying the proposed framework in primary healthcare centres, remote clinics as well as edge based devices.

In conclusion, the experimental results suggest that the proposed light weighted framework especially its SVM-based model is efficient and scalable for heart disease

prediction. The combined examination of performance measures, confusion matrices and ROC curves establishes that high prediction accuracy can be obtained without the need to deploy computationally-complex deep learning models, thereby rendering this approach suitable for resource-limited health care settings.

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed a light-weight machine learning framework to predict the large-scale heart disease whose goal is to tackle computational and deployment challenges when dealing with resource limited healthcare environment. The proposed SBA approach is built on computationally-efficient classifiers over end-to-end learning of deep models, and is a good trade-off between predictive power, expandability, and applicability.

A rigorous experimental study is performed on a large synthesized dataset of 78,812 samples with sample-based expansion from the public available heart disease data. Logistic Regression, Decision Tree and Support Vector Machine were three lightweight classifiers used under the same preprocessing and feature-selection pipeline. Furthermore, the SVM model was able to realize a significant increase in both accuracy and ROC-AU compared with the original NMEFs whilst still generating robust generalization. The studies of confusion matrices and ROC curves also verified that the SVM could reduce false positive as well as false negative predictions, which are important in a credible cardiac risk estimation.

The findings illustrate that with lightweight machine learning models it is viable to achieve high levels of predictive performance without the heavy computational overhead demanded by deep learning approaches. In addition, the proposed framework is an ideal choice for clinical decision support system in real time, primary care services and edge computing-based healthcare systems because of its very few hardware requirement and reduced power consumption as well as latency. The inclusion of graphical user interface also shows the possibility of applying the system in actual medical applications.

In spite of these promising results, several issues to be addressed still remain. First, additional clinical and physiological variables (like electrocardiographic or wearable sensor data) could be integrated to further improve the prediction performance by this framework. Second, external validation of the findings in independent datasets from different populations would solidify generalizability and clinical utility. Third, it is interesting to investigate the adaptive/incremental learning methods in the case of continuous updates when facing new data. Finally, optimizing the framework for embedded and edge-computing devices would be a difficult step towards real-time ubiquitous CVD risk monitoring.

In conclusion, this work has shown that lightweight machine learning models are a viable and efficient system for scalable prediction of heart disease, which paves the way for intelligent healthcare decision-support systems that are resource-efficient.

References

- [1] World Heart Federation, World Heart Report 2023, World Heart Federation, Geneva, Switzerland, 2023.
- [2] D. Dua and C. Graff, "Heart Disease," UCI Machine Learning Repository, University of California, Irvine, CA, USA, 1988. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease>
- [3] A. A. Ahmad, M. M. Hasan, and S. Rahman, "Prediction of heart disease based on machine learning algorithms," Journal of Healthcare Engineering, vol. 2023, Article ID 6692048, 2023.

- [4] K. Karthick, S. Srinivasan, and R. Manikandan, "Implementation of a heart disease risk prediction model using machine learning techniques," *Healthcare*, vol. 10, no. 3, pp. 1–15, 2022.
- [5] M. Ozcan, "A classification and regression tree algorithm for heart disease prediction and rule extraction," *Biomedical Signal Processing and Control*, vol. 79, Article no. 104057, 2023.
- [6] G. Ambrish, "Logistic regression technique for prediction of cardiac disease using UCI dataset," *Materials Today: Proceedings*, vol. 62, pp. 6463–6467, 2022.
- [7] S. M. Muthukumari, "The pivotal role of edge computing with machine learning and its impact on healthcare," in *Edge Computing and Machine Learning in Healthcare Applications*, Hershey, PA, USA: IGI Global, 2022, pp. 1–18.
- [8] N. Biswas, A. Roy, and S. Dutta, "Machine learning-based model to predict heart disease," *Computational Intelligence and Neuroscience*, vol. 2023, Article ID 8845126, 2023.
- [9] R. Alizadehsani et al., "A database for using machine learning and data mining techniques for coronary artery disease diagnosis," *Scientific Data*, vol. 6, no. 227, 2019.
- [10] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," *International Journal of Computer Science and Network Security*, vol. 8, no. 8, pp. 343–350, 2008.
- [11] A. Javeed et al., "An efficient learning system based on support vector machines for heart disease prediction," *IEEE Access*, vol. 7, pp. 181876–181887, 2019.
- [12] M. Ozcan, "Heart disease prediction using decision tree-based classification and rule extraction," *Biomedical Signal Processing and Control*, vol. 68, 2021.
- [13] A. Liaqat et al., "An intelligent ensemble-based heart disease prediction system using Random Forest," *Applied Sciences*, vol. 11, no. 7, 2021.
- [14] A. M. Hasan et al., "Feature selection-based machine learning approach for heart disease prediction," *Computers in Biology and Medicine*, vol. 137, 2021.
- [15] K. Polat and S. Güneş, "Medical decision support system based on artificial immune recognition for heart disease diagnosis," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2512–2516, 2011.
- [16] P. Verma et al., "Deep learning-based models for cardiovascular disease prediction: A review," *Journal of Biomedical Informatics*, vol. 127, 2022.
- [17] N. M. Basha et al., "Lightweight machine learning models for healthcare prediction in resource-limited environments," *Healthcare Analytics*, vol. 3, 2023.
- [18] S. Bashir, U. Qamar, and F. H. Khan, "IntelliHealth: A medical decision support application using a novel weighted decision tree," *Journal of King Saud University – Computer and Information Sciences*, vol. 33, no. 6, pp. 651–660, 2021.
- [19] S. Rani and R. K. Kumar, "Predictive modeling of heart disease using machine learning techniques," *International Journal of Computer Applications*, vol. 181, no. 18, pp. 1–6, 2018.
- [20] M. A. Jabbar, P. Chandra, and B. L. Deekshatulu, "Heart disease prediction system using associative classification and genetic algorithm," *Journal of Information Technology and Software Engineering*, vol. 6, no. 2, pp. 1–5, 2016.
- [21] M. Rahman, M. M. Islam, and H. A. Rahman, "A machine learning approach for heart disease prediction using ensemble techniques," *Applied Sciences*, vol. 10, no. 22, Article no. 8340, 2020.
- [22] A. S. Albahri et al., "Systematic review of machine learning techniques for heart disease diagnosis," *Applied Sciences*, vol. 10, no. 15, Article no. 5134, 2020.
- [23] P. K. Anooj, "Clinical decision support system for heart disease diagnosis using hybrid machine learning techniques," *Journal of Biomedical Informatics*, vol. 116, Article no. 103703, 2021.