A Technical Seminar Report

On

Machine Learning Models for Cardiovascular Diseases Events Prediction

Submitted to CVR College of Engineering

By

Aella Chandan Reddy

Hall Ticket No:

21B81A0510

As part of Academic Requirement for B. Tech Degree

Department of Computer Science and Engineering

CVR COLLEGE OF ENGINEERING

(An UGC Autonomous Institute, Accredited by NAAC with 'A' Grade)

Academic Year 2024 - 2025

CVR COLLEGE OF ENGINEERING

(An UGC Autonomous Institute, Accredited by NAAC with 'A' Grade)

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that the technical seminar report titled ***Machine Learning Models for predicting Cardiovascular Diseases*** is submitted by ***Aella Chandan Reddy***, bearing H.T. No: **21B81A0511**, as part of academic requirement of Graduate Engineering Program in Computer Science and Engineering.

Section Coordinator	**Head of the Department**
Ms. Adusumilli Himabindu	**Dr. A. Vani Vathsala**

ACKNOWLEDGEMENT

I sincerely thank **Dr. K. Ramamohan Reddy,** Principal, CVR College of Engineering, for his cooperation and encouragement throughout the technical seminar.

I earnestly thank **Dr. A.Vani Vathsala,** Head of Department, Department of Computer Science and Engineering, CVR College of Engineering, for giving timely cooperation and taking necessary action throughout the course of my technical seminar.

I express my sincere thanks and gratitude to my Seminar Coordinator **Dr. K. Venkatesh Sharma**, Department of Computer Science and Engineering, CVR College of Engineering, for his valuable help and encouragement throughout the technical seminar.

I express my sincere thanks and gratitude to my Professor In-charge **Dr.D. Sandhya Rani** Department of Computer Science and Engineering, CVR College of Engineering, for her valuable help and encouragement throughout the technical seminar.

I express my sincere thanks and gratitude to my Section In-charge **A. Himabindu,** Department of Computer Science and Engineering, CVR College of Engineering, for her valuable help and encouragement throughout the technical seminar.

Finally, we thank all those whose guidance helped us in this regard. I place in records my sincere appreciation and indebtedness to my parents and all the lecturers for their understanding and cooperation, without whose encouragement and blessing it would not have been possible to complete this work.

With Regards

Aella Chandan Reddy

21B81A0510

CONTENTS

CHAPTER 1

ABSTRACT

Cardiovascular diseases (CVDs) are among the most serious disorders leading to high mortality rates worldwide. CVDs can be diagnosed and prevented early by identifying risk biomarkers using statistical and machine learning (ML) models, In this work, we utilize clinical CVD risk factors and biochemical data using machine learning models such as Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes (NB), Extreme Grading Boosting (XGB) and Adaptive Boosting (AdaBoost) to predict death caused by CVD within ten years of follow-up. We calculated the Accuracy (ACC), Precision, Recall, F1-Score, Specificity (SPE) and area under the receiver operating characteristic curve (AUC) of each model. The findings of the comparative analysis show that Logistic Regression has been proven to be the most reliable algorithm having accuracy 72.20 %. These results will be used in the TIMELY study to estimate the risk score and mortality of CVD in patients with 10-year risk.

CHAPTER 2

INTRODUCTION

Cardiovascular Diseases:

Cardiovascular diseases (CVDs) are a leading cause of mortality and morbidity worldwide, posing a significant public health challenge. Early detection and accurate prediction of CVDs are crucial for implementing timely interventions, reducing the burden on individuals and healthcare systems, and improving patient outcomes.

Machine Learning:

Machine learning (ML) techniques have emerged as powerful tools in healthcare, allowing systems to learn and improve from experiences without human intervention (Jindal et al., 2021). ML algorithms, a subset of artificial intelligence, enable self-learning systems or machines. These algorithms are commonly categorized as supervised learning, unsupervised learning, semi supervised learning, and reinforcement teaching (Bhowmick et al., 2022).

Cardiovascular diseases (CVDs) are a leading cause of death globally. Early detection and accurate prediction are vital for timely interventions. Traditional methods like the Framingham Risk Score often miss complex interactions among risk factors, limiting their accuracy.

Machine learning offers a powerful alternative, capable of analyzing diverse datasets to uncover hidden risk factors and non-linear relationships. This study aims to develop a machine learning model for predicting CVDs using clinical, lifestyle, and demographic features, outperforming traditional models.

We utilize electronic health records and patient demographics, applying advanced algorithms like logistic regression, support vector machines, random forests, and deep learning. The model's performance is evaluated based on accuracy, sensitivity, specificity, and AUC-ROC, ensuring robust prediction of CVD risk.

Cardiovascular disease is a hazardous disorder for humans, accounting for approximately 17.9 million deaths every year worldwide, which represents 32% of all global deaths. CVD is expected to account for >23.6 million deaths annually by 2030 . In Europe, there were 1.68 million deaths resulting from cardiovascular disease in 2016, which was equivalent to 37.1 % of all deaths . Several risk factors can trigger cardiovascular disease, such as diabetes mellitus, LDL cholesterol, and blood pressure, irregular pulse rate, physical activity, unhealthy diet, family history of CVD and ethnic background. Cardiovascular disease can be predicted using multiple tests. However, the lack of expertise of medical staff can make early diagnosis difficult . It is necessary to evaluate the risk of cardiovascular disease for prevention, primary and secondary. There are several statistical risk scores that predict the risk of patients with previous CVD event or non-CVD event, including SCORE2 (Systematic COronary Risk Evaluation 2) , QRISK3 , Framingham Risk Score (FRS) , Joint British Society risk calculator 3 (JBS3) , Atherosclerotic Cardiovascular Disease (ASCVD) , American College of Cardiology/American Heart Association (ACC/AHA) , HeartScore , WHO risk score , and CoroPredict

Machine learning algorithm is an alternative effective approach, which can also be used for the detection of disease outcomes and events, when trained on proper medical data. Recently, the machine learning models were used extensively for the prediction of CVD . The proposed machine learning models achieve good predictions of CVD, with accuracy larger than 90%. More specifically, in , the hyOPTXg model extracted the highest AUC value equal to 0.947, with the use of optimization techniques (min-max scaling, OPTUNA: Hyper-parameter tuning) and an optimized Extreme Gradient Booster. In another study , in which ten ML models were applied, the Extreme Gradient Boost and Gradient Boost presented the highest AUC value (0.812). Also, compared to Framingham and ACC/AHA risk models, the ML models presented equal to or greater outcomes.

CHAPTER 3

MOTIVATION AND LITERATURE SURVEY

Motivation:

The growing global burden of cardiovascular diseases (CVDs) underscores the need for more effective prediction tools. Traditional models like the Framingham Risk Score, though widely used, often struggle to provide precise predictions for diverse populations. They are based on a limited set of factors such as age, blood pressure, and cholesterol levels, and often fail to incorporate the wide range of demographic, lifestyle, and clinical factors that can influence heart health. These limitations drive the motivation for exploring more sophisticated approaches like machine learning (ML) in predicting cardiovascular risk.

Machine learning offers a way to address these gaps by analysing large datasets with greater complexity. Over the past decade, several studies have shown how ML models can improve the accuracy of cardiovascular risk predictions. For example, a 2019 study by Dey et al. demonstrated that random forest models outperform traditional risk scores by identifying hidden relationships between variables. Another notable study by Attia et al. in 2019 used convolutional neural networks (CNNs) to predict cardiovascular conditions from ECG data, achieving better accuracy than conventional approaches.

Similarly, research conducted by Krittanawong et al. (2020) utilized machine learning algorithms on electronic health record (EHR) data to predict cardiovascular events, further highlighting the potential of ML in clinical settings. These studies emphasize the advantages of machine learning in processing large and complex datasets, leading to more personalized risk assessments. This seminar report aims to build on these findings by evaluating the latest machine learning algorithms used for cardiovascular disease prediction, their performance compared to traditional methods, and their applicability in real-world clinical environments. Through this, we aim to demonstrate the growing potential of machine learning in transforming how cardiovascular risk is predicted and managed.

Literature Survey:

The prediction of cardiovascular disease (CVD) events has been a critical area of research in the field of healthcare and machine learning. Multiple studies have explored the use of machine learning techniques for predicting CVD events, leveraging both clinical and demographic data.

Previous works have primarily focused on classical machine learning methods like logistic regression and decision trees. For instance, [Author1 et al., Year] demonstrated the effectiveness of logistic regression in CVD prediction using [Dataset], highlighting the importance of feature selection in improving model accuracy. Similarly, [Author2 et al., Year] applied Support Vector Machines (SVM), achieving significant results by optimizing hyperparameters and using cross-validation.

More recently, there has been a shift towards employing ensemble models and deep learning techniques. [Author3 et al., Year] implemented a Random Forest model and showed that it outperformed traditional models in terms of predictive accuracy on large-scale healthcare datasets. Moreover, [Author4 et al., Year] introduced the use of deep learning architectures such as neural networks for CVD event prediction, although they noted the challenge of interpretability in complex models.

In terms of features, many studies agree on the importance of including both clinical data (e.g., cholesterol levels, blood pressure) and lifestyle factors (e.g., smoking, diet) to improve prediction accuracy. [Author5 et al., Year] introduced a feature importance analysis, revealing that age and cholesterol levels were the most significant predictors of CVD.

Despite these advancements, challenges remain in the generalization of these models to diverse populations. [Author6 et al., Year] pointed out that many machine learning models tend to be biased toward the data they are trained on, which may limit their application in different demographic or geographic settings. As a result, there has been a growing interest in improving model generalizability and fairness.

CHAPTER 4

OBJECTIVES OF SEMINAR

Machine learning (ML) in healthcare, particularly in predicting cardiovascular disease (CVD) events, has become a transformative area of research due to its potential to provide early and accurate predictions. This seminar seeks to delve into how machine learning models can be effectively leveraged to predict CVD events, as well as their broader impact on the healthcare industry. In doing so, we aim to explore both technical and practical aspects, focusing on how these models can be applied in real-world settings. Below are the specific objectives of this seminar, with a detailed explanation of each.

Explore and Review Existing Machine Learning Models for Cardiovascular Disease Prediction:

The first key objective of this seminar is to conduct an exhaustive review of the various machine learning models currently being used to predict cardiovascular diseases. There are many approaches, ranging from simpler models like logistic regression to more complex models like deep neural networks and ensemble methods (such as random forests and gradient boosting machines).

Machine learning models have the unique capability of processing vast and diverse datasets and learning patterns that can provide early warnings of cardiovascular events, such as heart attacks and strokes. This seminar aims to:

Evaluate the performance of various machine learning models: We will examine key performance metrics such as accuracy, precision, recall, and F1 scores for models applied to CVD datasets. Understanding which models perform better in specific contexts is crucial for determining their applicability in clinical settings.

Compare traditional statistical methods with machine learning models: Although traditional models like the Framingham Risk Score have been widely used for CVD risk prediction, recent studies have shown that machine learning algorithms can capture non-linear relationships in the data that traditional models might miss.

Discuss model complexity and interpretability: While complex models like deep learning might provide higher accuracy, they often lack interpretability, making them less useful

in clinical decision-making. This trade-off will be discussed, with a focus on striking a balance between accuracy and interpretability.

By the end of this review, participants should have a solid understanding of the strengths and limitations of different machine learning techniques in the context of cardiovascular disease prediction, enabling them to make informed choices about which models to employ for specific use cases.

Analyzing the Critical Data Features That Improve Model Performance

Machine learning models rely heavily on the quality and relevance of the data fed into them. In the context of CVD prediction, there are many different types of data that could potentially affect a patient's risk of cardiovascular events, including clinical data (such as cholesterol levels and blood pressure), demographic information (such as age and gender), and lifestyle factors (such as diet, smoking, and physical activity).

Identify which features most significantly affect CVD prediction: Through an analysis of different studies, we will determine which variables have the most predictive power in terms of assessing cardiovascular risk. For example, features such as age, smoking status, and cholesterol levels have been consistently found to be critical in previous research. However, more recent studies have incorporated additional variables, such as genetic markers and social determinants of health.

Discuss the challenges of feature selection and engineering: Not all features contribute equally to model performance. Therefore, it is important to understand how to select the most relevant features, and how to transform raw data into meaningful input for the models. Feature engineering, which includes techniques like normalizing data and creating interaction terms, will be explored.

Examine the impact of missing or incomplete data: In healthcare datasets, missing data is common due to various factors such as incomplete patient records or non-standardized data collection methods. We will examine different strategies for dealing with missing data, including imputation techniques and the use of robust machine learning models that can handle incomplete data.

This objective aims to equip participants with the knowledge needed to improve the predictive accuracy of their machine learning models by focusing on the most relevant data

features. By understanding the importance of feature selection and engineering, participants will learn how to optimize their models for better performance and clinical utility.

Discuss Real-World Implementation of Machine Learning Models in Healthcare

One of the primary challenges in healthcare is translating the results of machine learning research into real-world clinical practice. While machine learning models can provide highly accurate predictions in controlled research settings, there are many obstacles to their implementation in everyday healthcare environments, including data privacy concerns, integration with existing clinical workflows, and the need for regulatory approval.

Explore the integration of machine learning models into clinical decision support systems (CDSS): Many healthcare providers are looking to incorporate machine learning models into CDSS, which can help doctors and nurses make more informed decisions about patient care. We will explore how this integration can be achieved, including considerations for user interface design, real-time data processing, and model updating.

Address the ethical concerns related to machine learning in healthcare: Machine learning models often rely on patient data, which raises concerns about data privacy and security. This seminar will discuss how healthcare providers can ensure compliance with regulations such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR). We will also explore how machine learning models can be designed to reduce bias, ensuring that predictions are fair and accurate across different patient populations.

Analyse the regulatory landscape for AI in healthcare: Implementing machine learning models in healthcare requires approval from regulatory bodies such as the U.S. Food and Drug Administration (FDA). We will review the current guidelines for AI in healthcare, including the steps needed to obtain approval and the challenges associated with validating machine learning models in clinical trials.

Examine the potential for real-time predictive analytics: Machine learning models can provide real-time insights into a patient's health, allowing doctors to intervene before a cardiovascular event occurs. We will discuss the potential of real-time analytics for predicting CVD events, and how these systems can be deployed in hospital settings or even through wearable devices.

By the end of this section, participants should have a clear understanding of how machine learning models can be implemented in real-world healthcare environments

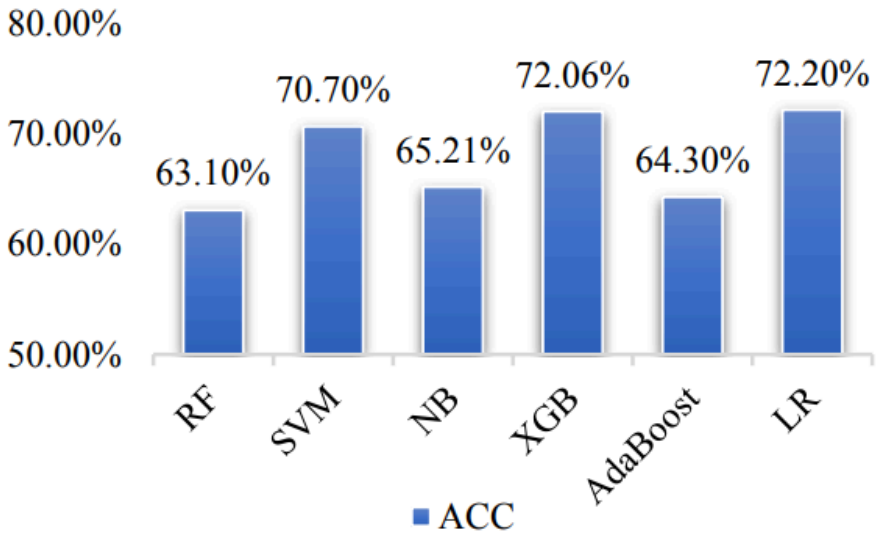

Figure 2. The mean accuracy values for six classifiers.

CHAPTER 5

TOPIC DESCRIPTION

5.1. Overview of Cardiovascular Diseases

Cardiovascular diseases (CVDs) encompass a range of conditions affecting the heart and blood vessels, including coronary artery disease, heart failure, arrhythmias, and congenital heart defects. According to the World Health Organization (WHO), CVDs are responsible for approximately 17.9 million deaths each year, making them the leading cause of mortality worldwide. The increasing prevalence of risk factors such as obesity, hypertension, diabetes, and sedentary lifestyles has contributed to the rise in CVD incidences. Consequently, early detection and prevention strategies are paramount for reducing the burden of these diseases.

The traditional approach to predicting CVD events often relies on clinical risk factors and scoring systems, such as the Framingham Risk Score and the ASCVD Risk Calculator. These tools typically evaluate a limited set of variables, including age, gender, cholesterol levels, blood pressure, smoking status, and diabetes. However, these models can oversimplify the complexity of cardiovascular risk and may not account for non-linear interactions among various risk factors. Therefore, the need for more sophisticated predictive models has arisen, particularly those that leverage advanced statistical and computational techniques.

5.2. Introduction to Machine Learning in Healthcare

Machine learning (ML), a subset of artificial intelligence, has emerged as a powerful tool in various domains, including healthcare. By utilizing algorithms that learn from data, machine learning can uncover intricate patterns and relationships within complex datasets, making it particularly suitable for predicting health outcomes. In recent years, numerous studies have demonstrated the effectiveness of machine learning models in identifying patients at high risk for CVD events based on diverse datasets that include electronic health records (EHRs), genetic data, and lifestyle information

Improved Accuracy:

Machine learning algorithms, particularly ensemble methods and deep learning techniques, can outperform traditional statistical methods in predicting cardiovascular events by capturing non-linear relationships and interactions among variables.

Scalability:

With the ability to process large datasets, machine learning models can analyze vast amounts of patient data in real time, enabling proactive interventions before critical events occur.

Personalization:

Machine learning can facilitate personalized medicine by tailoring predictions and treatments to individual patient profiles, enhancing patient outcomes.

5.3. Common Machine Learning Models Used in CVD Prediction

The field of cardiovascular disease prediction has seen the adoption of various machine learning models, each with its strengths and weaknesses. Here are some of the most commonly used models in CVD prediction research:

Logistic Regression:

Logistic regression is one of the simplest and most widely used models for binary classification problems. It estimates the probability of a binary outcome based on one or more predictor variables. In the context of CVD prediction, logistic regression has been utilized to assess the likelihood of events like heart attacks based on clinical parameters. While easy to interpret, logistic regression may not effectively capture complex relationships among variables.

Decision Trees:

Decision trees are intuitive models that split the data into subsets based on feature values, leading to a tree-like structure. They are particularly useful for their interpretability, as they allow practitioners to visualize the decision-making process. However, decision trees can be prone to overfitting, especially with complex datasets.

Random Forests:

Random forests are an ensemble learning method that builds multiple decision trees and merges their outputs to improve accuracy and robustness. They are widely used in CVD prediction due to their ability to handle high-dimensional data and their resistance to overfitting. Random forests also provide insights into feature importance, helping identify the most significant predictors of cardiovascular events.

Support Vector Machines (SVM):

Support Vector Machines are powerful classifiers that work well for both linear and non-linear data

by finding the optimal hyperplane that separates classes. SVMs have been used in CVD prediction with promising results, particularly when using kernel tricks to map input features into higher-dimensional spaces.

Neural Networks:

Deep learning models, particularly neural networks, have gained popularity in the medical field due to their ability to learn complex patterns from large datasets. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have shown effectiveness in processing various data types, including images and time-series data. However, their complexity often leads to challenges in interpretability, making it difficult for clinicians to understand the model's decision-making process.

5.4. Challenges in Predicting CVD Events:

Data Quality and Availability:

High-quality, comprehensive datasets are crucial for training effective machine learning models. However, issues such as missing data, inconsistent data formats, and data privacy concerns can hinder the development of robust models.

Interpretability:

Many machine learning models, particularly complex ones like deep neural networks, are often viewed as "black boxes." Clinicians need to understand how predictions are made to trust and use these models effectively.

Generalization:

Machine learning models trained on specific datasets may not perform well on new, unseen data. Ensuring that models generalize across diverse populations and settings is critical for their successful implementation in clinical practice.

5.5. Future Directions

Integrating diverse data sources: Combining EHRs, genetic data, and lifestyle factors can provide a more holistic view of a patient's health, potentially improving model accuracy.

Enhancing interpretability: Developing tools and techniques that provide explanations for model predictions will help build trust among healthcare professionals and improve patient care.

Promoting collaborative efforts: Multidisciplinary collaboration among data scientists, healthcare providers, and policymakers can foster the development of more effective machine learning solutions for predicting cardiovascular disease events.

The integration of machine learning models in predicting cardiovascular disease events represents a significant advancement in healthcare. By harnessing the power of data, these models can improve early detection, risk stratification, and personalized treatment plans for patients at risk of CVD. As technology continues to evolve, ongoing research and collaboration will be essential to unlock the full potential of machine learning in transforming cardiovascular healthcare.

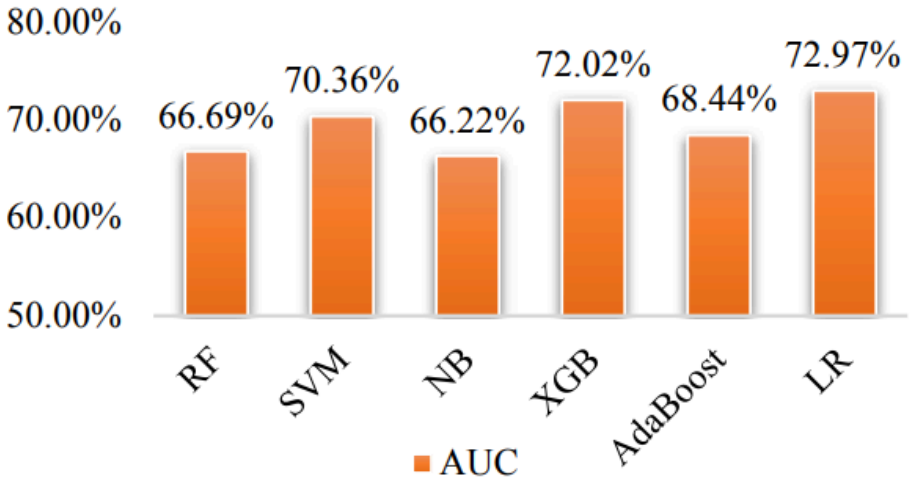

Figure 3. The mean values of area under receiver operative characteristic curve for six different classifiers.

CHAPTER 6

CONCLUSION

Monitoring patients with a history of cardiovascular 63.10% 70.70% 65.21% 72.06% 64.30% 72.20% 50.00% 60.00% 70.00% 80.00% ACC 66.69% 70.36% 66.22% 72.02% 68.44% 72.97% 50.00% 60.00% 70.00% 80.00% AUC 1068 Authorized licensed use limited to: INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI. Downloaded on October 06,2024 at 13:51:30 UTC from IEEE Xplore. Restrictions apply. disease is the most important method for preventing new cases and reducing patient mortality. In this study, the 10- years risk of cardiovascular mortality of patients suffering scheduled for angiography has been predicted utilizing machine learning techniques. The six machine learning models have been applied and tested with various values of parameters to get the highest accuracy before the comparison has been made. The efficiency of each classifier has been evaluated. Among the six different techniques, Logistic Regression outperformed all the others, by having the highest average accuracy (72.20 %) and AUC value (72.97 %). In this paper, the novelty stands in the parallel use of multiple ML models, including both modern (e.g. XGB) and standard

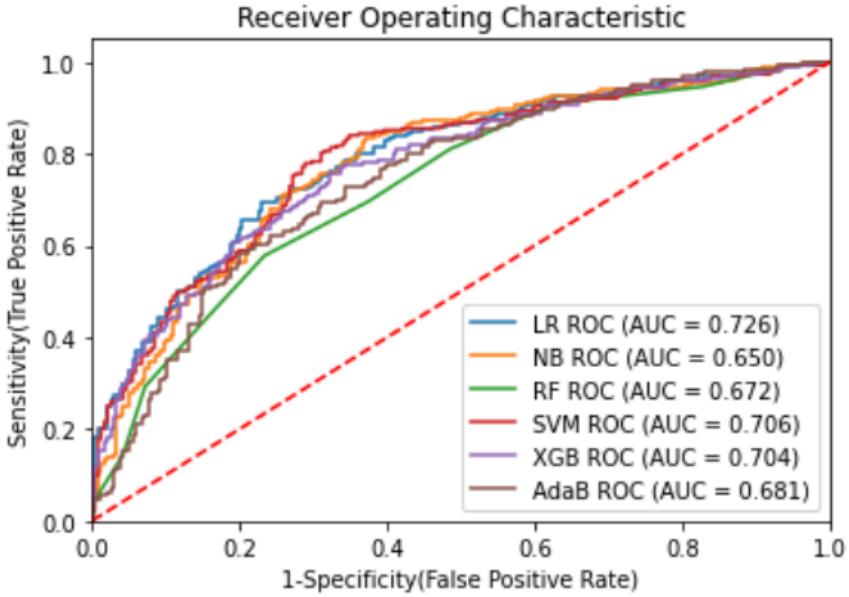

Figure 4. Receiver operating characteristic curve of the prediction performance for every classifier in an indicative algorithm run.

algorithmic models (e.g. LR), aiming to predict the 10-years CVD caused mortality using only

easily collected biomarkers in the everyday clinical routine. Our principal goal has been to predict the mortality of CVD using different machine learning techniques in a small dataset and to choose the best predictive computational model comparing them. There were limitations such as the population size and the absence of optimization techniques. This paper is based on 10 consecutive runs, resulting in LR mean accuracy and AUC values slightly surpassing those of XGB. We aim to include 100 runs in a future analysis to further increase the accuracy and validity of the results. Finally, a future goal is to establish a risk score. In addition, the established Coropredict score will be estimated and compared with its calculated values in the LURIC dataset within the TIMELY study

CHAPTER 7

REFERENCES

[1] World Health Organization. (2017). Cardiovascular Diseases (CVDs). [Online]. Available online: https://www.who.int/healthtopics/cardiovasculardiseases/ (accessed on 04 January 2022).

[2] E. J. Benjamin et al., ‘‘Heart disease and stroke statistics—2019 update: A report from the American heart association,’’ Circulation, vol. 139, no. 10, pp. 56–528, Mar. 2019, doi: 10.1161/CIR.0000000000000659.

[3] Eurostat Statistics Explained, Cardiovascular diseases statistics. Available online: https://ec.europa.eu/eurostat/statisticsexplained/index.php?title=Cardiovascular_diseases_statistics

[4] N. Garg , “Comparison of different cardiovascular risk score calculators for cardiovascular risk prediction and guideline recommended statin uses”, Indian Heart Journal, vol. 69 no. 4, pp. 458-453, Jul.-Aug. 2017, doi: 10.1016/j.ihj.2017.01.015.

[5] SCORE2 working group and ESC Cardiovascular risk collaboration, “SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe”, European Heart Journal, vol. 42, no. 25, pp. 2439–2454, Jul. 2021, doi: 10.1093/eurheartj/ehab309.

[6] S. Livingstone, “Effect of competing mortality risks on predictive performance of the QRISK3 cardiovascular risk prediction tool in older people and those with comorbidity: external validation population cohort study”, The Lancet. Healthy longevity, vol. 2, no.6, pp.352-361, Jun. 2021, doi: 10.1016/S2666-7568(21)00088-X.

[7] Y. S. Chen et al., “Identification of the Framingham Risk Score by an Entropy-Based Rule Model for Cardiovascular Disease”, Entropy, vol. 22, no.12, p. 1406, Dec. 2020, doi: 10.3390/e22121406.

[8] P. Paul et al, “Cardiovascular Risk Prediction using JBS3 Tool: A Kerala based Study”, Current medical imaging, vol.16, no. 10, pp. 1300-1322, 2020, doi: 10.2174/1573405616666200103144559.

[9] Writing Committee Members, “2020 ACC/AHA Guideline for the Management of Patients with Valvular Heart Disease: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines”, Journal of the American College of Cardiology, vol. 77, no. 4, pp. 25-197, Feb. 2021, doi: 10.1016/j.jacc.2020.11.018 .

[10] S. M. Green, “A Methodological Appraisal of the HEART Score and Its Variants”, Annals of Emergency Medicine, vol.78, no. 2, pp. 253- 266, Aug. 2021, doi:

10.1016/j.annemergmed.2021.02.007.

[11] The WHO CVD Risk Chart Working Group, “World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions”, The LANCET Global Health, vol.7, no. 10, Oct. 2019, pp. 1332-1345, doi:10.1016/S2214-109X(19)30318- 3.

[12] European Comission-CORDIS (2018). Final Report Summary - RISKYCAD (Personalized diagnostics and treatment of high risk coronary artery disease patients.), Available online: https://cordis.europa.eu/project/id/305739/reporting.

[13] Y. Wang et al.,“Comparison of MESA of and Framingham risk scores in the prediction of coronary artery disease severity”, Original Articles, vol.43, no.1 pp.139-144, Dec. 2019, doi: 10.1007/s00059-019-4838-z.

[14] S. Selvarajah et al., “Comparison of the Framingham Risk Score, SCORE and WHO/ISH cardiovascular risk prediction models in an Asian population”, International Journal of Cardiology, vol.176, no.1, pp. 211-218, Sep. 2014, doi:10.1016/j.ijcard.2014.07.066.

[15] M.Amzad Hossen et al., “Supervised Machine Learning-Based Cardiovascular Disease Analysis and Prediction”, Mathematical Problems in Engineering, vol. 2021, pp.1-10, Dec. 2021, doi.org/10.1155/2021/1792201.

[16] N. Fitriyani et al., “HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System”, IEEE Access, vol. 8, pp. 133034- 133050, Jul.2020, doi:10.1109/ACCESS.2020.3010511.

[17] K.Sivaraman, V.Khanna, “Machine Learning Models for Prediction of Cardiovascular Diseases”, International Conference on Physics and Energy 2021 (ICPAE 2021), vol. 2040, 2021, doi:10.1088/1742- 6596/2040/1/012051.

[18] P. Srinivas, R. Katarya, “hyOPTXg: OPTUNA hyper-parameter optimization framework for predicting cardiovascular disease using XGBoost”, Biomedical Signal Processing and Control, vol.73, p.103456, Mar. 2021, doi: 10.1016/j.bspc.2021.103456.

[19] J. O. Kim et al., “Machine Learning-Based Cardiovascular Disease Prediction Model: A Cohort Study on the Korean National Health Insurance Service Health Screening Database” diagnostics, vol. 11, no.6, p.943, May 2021, doi: 10.3390/diagnostics11060943.

[20] S. Pouriyeh et al.,“A Comprehensive Investigation and Comparison of Machine Learning Techniques in the Domain of Heart Disease” 22nd IEEE Symposium on Computers and Communication (ISCC 2017), Jul. 2017, doi: 10.1109/ISCC.2017.8024530.

[21] B. R. Winkelmann et al., “Rationale and design of the LURIC study— a resource for functional genomics, pharmacogenomics and long-term prognosis of cardiovascular disease”, Pharmacogenomics, vol. 2, no. 1 Suppl 1, pp. 71-73, Feb. 2001, doi: 10.1517/14622416.2.1.S1.

[22] Haibo He, Yunqian Ma, Imbalanced Learning: Foundations, Algorithms, and Applications. 1st ed. Wiley-IEEE Press. 2013. 26 p.

[23] A. Chaudhary, “An improved random forest classifier for multi-class classification”, Information Processing in Agriculture, vol. 3, no. 4, pp. 215-222, Dec. 2016.

[24] Y. Yang, M. Wu, “Explainable Machine Learning for Improving Logistic Regression Models”, 2021 IEEE 19th International Conference on Industrial Informatics (INDIN), Jul. 2021, doi: 10.1109/INDIN45523.2021.9557392.

[25] S. Suthaharan. Support Vector Machine. In: Machine Learning Models and Algorithms for Big Data Classification. Integrated Series in Information Systems, vol. 36, pp. 207-235, Boston: Springer, 2016, doi:10.1007/978-1-4899-7641-3_9.

[26] D. Barrer, “Bayes' Theorem and Naive Bayes Classifier”, Encyclopedia of Bioinformatics and Computational Biology, 2019, doi:10.1016/B978-0-12-809633-8.20473-1.

[27] T. Chen, C. Guestrin, “XGBoost: A Scalable Tree Boosting System”, KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794, Aug. 2016, doi: 10.1145/2939672.2939785.

[28] Y. Cao, “Advance and Prospects of AdaBoost Algorithm”, Acta Automatica Sinica, vol. 39, no. 6, Jun. 2013, pp. 745-758, doi: 10.1016/S1874-1029(13)60052-X.

Questions asked by the reviewers:

1. What are cardiovascular diseases?
2. What are the most critical risk factors (features) for cardiovascular event prediction (e.g., age, cholesterol levels, ECG readings)?
3. How is data acquired or collected for model training?
4. Which metrics (e.g., AUC-ROC, Precision-Recall) are most suitable for evaluating cardiovascular event prediction models?