Feature Selection and Optimization for Cardiovascular Disease Prediction

Dr. Poorna Priya Putcha[1], Dr. P. Pavitra[2], Lakshmi Amulya Chellapilla[3], Vasantha Dhuli[4], Uma Malleswari Boddeti[5], Sandhya Atmakuri [6]

Dadi institute of Engineering& Technology (A) ,Anakapalle,India

putchapoornapriya@gmail.com[1],pavitray015@gmail.com[2], chlamulya08@gmail.com[3],
dhuli.vasantha@gmail.com[4],umaboddeti1234@gmail.com[5],sandhyaatmakuri4@gmail.com[6]

Abstract.

This Emergent technology has eased the burden on hospital resources and has allowed for efficient monitoring and management of patients. Cardiovascular diseases, which account for the highest cause of death, by the statistics of the World Health Organization, contribute to almost 20,000 deaths in this nation every year. This leads in a high death rate over the world, which helps to boost the current healthcare infrastructure. The primary aim of this research is to incorporate advanced Machine Learning (ML) classification algorithms to identify abnormal human vital signs, like Electrocardiogram (ECG), Heart Rate (HR), and Blood Pressure (BP), in near real-time. Given that network latency is crucial for this application, enhancing the overall Quality of Service is essential.

Keywords. Heart and blood vessel problems, Blood pressure, Pulse rate, electrical activity recording and others.

1. INTRODUCTION

The world's deadliest diseases are cardiovascular diseases. They claim over 17.9 million lives every year. It traditionally relies on traditional risk factors, but classically cannot predict cardiac events. Cardiac risk assessment can be enhanced using machine learning algorithms. It analyses large data and extracts subtle patterns. With such algorithms, individual risk scores are generated.

These include supervised, unsupervised, and deep learning methods, which improve the accuracy of prediction and allow for earlier intervention to prevent disease. Improved patient outcomes, resource utilization, and identification of people with higher risk profiles result from this. Personalized medicine is very crucial since it provides tailored treatment plans according to these risk profiles, especially through supervised learning data.

Challenges still remain, with data quality and availability issues, interpretability of the models, and how to integrate them into clinical workflows Thus, the equal use of machine learning-based cardiac risk assessments will call for addressing healthcare disparities in their utilizations. The integration thus requires collaboration from several actors, including healthcare professionals, data scientists, and policymakers, among others.

Machine learning (ML) offers an opportunity to analyse large data sets and reveal hidden patterns in patient information. This model can predict the risk of heart disease more precisely because it considers age, cholesterol levels, blood pressure, family history, and lifestyle choices. ML can also assess real-time data from wearable devices and diagnostic tools like

ECG, which are challenging for clinicians to interpret in detail. The integration of ML in heart risk assessment is a significant advancement.

Dataset 1: Heart Attack Risk

It gives information about all patients who have ever had an ML and includes demographics, clinical history, and characteristics of lifestyle of the patients. All data sets include variables like age, sex, blood pressure, levels of cholesterol, and smoking status. This data set provides comprehensive information on a number of risk factors associated with having an ML.

Dataset 2: COVID-19 Vaccination

The second dataset gives information regarding those who have been vaccinated against COVID-19. It includes demographic characteristics, vaccination details, and health outcomes. The data contains variables such as age, sex, type of vaccination, and adverse reactions.

2. RELATED WORK

Machine learning is revolutionizing how we assess cardiac risk. It enhances prediction accuracy, improves diagnoses, and allows for a more personalized approach. Unlike traditional methods that rely on fixed formulas like the Framingham Risk Score, machine learning relies on large datasets and complex algorithms to uncover underlying patterns. It uses supervised learning algorithms. Indeed, with current advances in research, it is shown that machine learning enhances cardiac risk assessment through diverse data analysis algorithms.

2.1 Supervised Learning:

Supervised Learning is a key area of ML whereby the algorithm learns from labelled examples for the prediction of results. In essence, data applied supervised learning should comprise input features such as images, texts, or numbers. The data also includes correct answers, known as labels or target values. The complete idea is that the algorithm learns connections between inputs and outcomes, which make predictions on the outcomes for new, unseen data. The process trains the model by feeding a dataset, it adjusts itself to commit fewer errors in its predictions. The model is evaluated separately on another set of data and is said to perform like this. Supervised learning is extensively applied to classification tasks, determining whether emails are spam, and regression problems, such as predicting the house price as a function of size and location features.

2.2 Logistic regression:

This method models how a binary outcome relates to one or more predictors. The outcome variable can take two values 0 and 1.Machine Learning can predict the responses of patients to therapies, propagating personalized therapy through optimized medication plans.

ML further innovates cardiac care: faster and more accurate personalized risk assessments likely to prevent multiple deaths.

$$P(y=1|X) = 1 / (1 + e^{-(b0 + b1x1 + b2x2 + \cdots + bnxn)}) \tag{2.2.1}$$

Where: • P(y=1|X) is the conditional probability of the dependent variable being 1, given the input features X.

•b0b_0b0 is the bias term (intercept).

• b1, b2,…, bn-1, bn are the weights linked to the features x1, x2,…, xn.

• x1, x2,…, xn are feature predictors or the independent var.

• “e” is the foundation of the natural log, approx. 2.71828.

3. PROPOSED MODEL

The Cardiac Risk Assessment System proposed helps doctors predict the risk of heart disease and help manage this risk during heart attack diagnosis by integrating patient data that encompasses traditional medical history, test results, and even information brought on by wearable devices. The system harnesses machine learning and data analysis techniques to derive trends that lead to the proposal of an escalated risk towards heart disease. With this groundwork then, the system aims at giving the prediction of heart disease risk, an individualized treatment plan, as well as lifestyle recommendations. Besides, the system is integrated with electronic health records for continuous patient information updates. The interface is user-friendly to doctors, making it intuitive for use.

4. RESULTS AND DISCUSSION

This area of research has displayed good potential in improving the cardiac risk assessment. In this study, it explains how machine learning algorithms may effectively predict cardiovascular disease with clinical and demographic factors. Logistic regression reached an accuracy of 65% on the training set and 60% on the testing set.

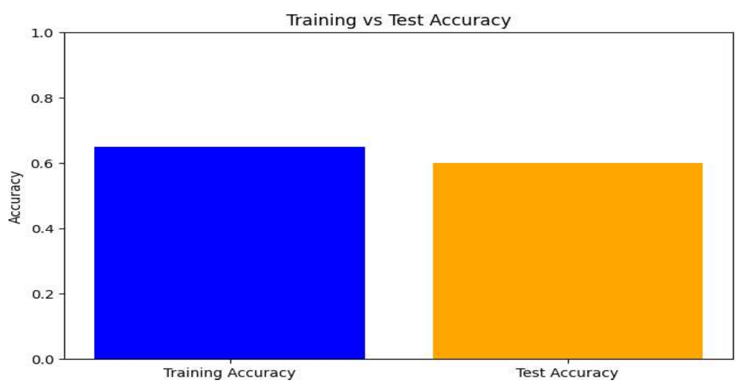

Figure 4.1: Training and Testing accuracy

The graph illustrates the training and test accuracy of a machine learning model. The blue bar indicates the training accuracy, which is just over 60%, showing how well the model performs on the training dataset. In contrast, the orange bar shows the test accuracy, slightly lower at about 55%, which indicates the model's capability to generalizes to new, untested data. The negligible difference between the training and test accuracy implies that there is minimal overfitting, although the absolute accuracy values are still low.

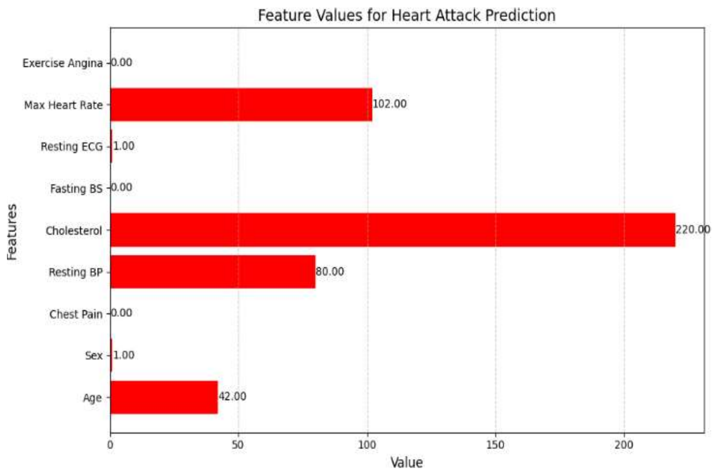

Figure 4.2: Feature Values of Heart Attack Prediction

Graph representation Feature values of heart attack prediction Some key health indicator graphs are presented here, too. Cholesterol level of a subject is recorded to be a high level of 220 mg/dL in border line high risk categories of heart disease. Record value of the heart is given by 102 bpm of bpm, which is an extreme low record value suggesting demising cardiac functions or cardio-fitness. Resting blood pressure is normal at 80 mmHg. Fasting blood sugar levels are also below the critical threshold.

NAME	Age	Gen	Chol	RESTBP	Covid	Exang	ECG	CP	FBS	MAX HR	Output
Sample 1	21	0	160	80	1	0	0	0	0	70	0
Sample 2	20	0	150	80	1	0	0	0	0	72	0
Sample 3	20	0	170	80	1	0	0	0	0	73	0
Sample 4	22	0	180	80	1	0	0	0	0	71	0
Sample 5	39	0	0	80	1	0	1	0	0	75	0
Sample 6	43	1	270	80	1	0	1	0	1	177	1
Sample 7	32	0	0	0	1	0	0	0	1	130	0
Sample 8	28	0	180	80	1	0	0	0	1	90	0
Sample 9	55	1	190	130	1	0	0	0	0	94	1
Sample 10	27	0	200	80	1	0	1	0	0	100	0
Sample 11	41	1	200	80	1	0	0	0	1	120	0
Sample 12	37	1	200	80	1	0	0	0	0	104	1
Sample 13	28	0	200	80	1	0	0	0	1	135	0
Sample 14	44	1	220	80	1	0	1	0	0	98	1
Sample 15	70	1	210	80	1	1	1	1	0	84	1
Sample 16	27	0	210	80	1	1	1	0	1	89	1
Sample 17	38	0	0	90	1	0	1	0	0	77	1
Sample 18	53	1	180	98	1	0	0	0	1	81	1
Sample 19	38	1	220	100	1	0	1	0	0	97	1
Sample 20	72	1	0	110	1	0	0	0	1	101	0
Sample 21	42	1	220	80	1	0	1	0	0	102	0

Table 4.1: Real time data

This is a data set of medical records of various individuals. Every row in the table represents a sample, and that sample is captured through the column labelled as NAME, describing the characteristic of the individual. These characteristics include Age, Gen: Gender, Chol: Cholesterol, RESTBP: Resting Blood Pressure, and whether Covid was experienced or not by the person. Other columns are exercise-induced angina, electrocardiographic results, type of chest pain experienced, fasting blood sugar level recorded, and maximum heart rate reached.

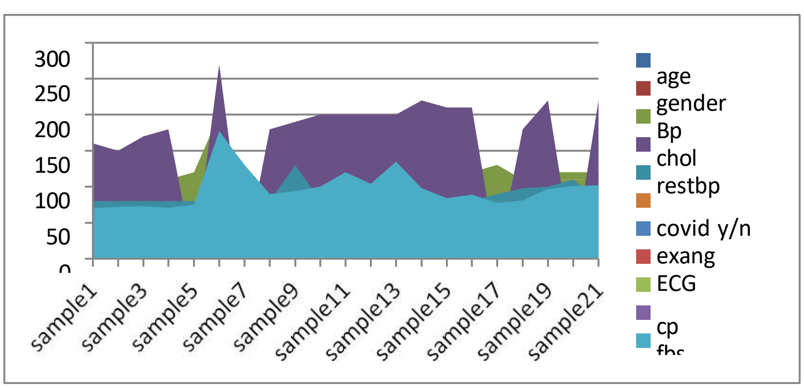

Figure 4.3: Real Time Graph

The chart shows the contribution of the key health indicators such as cholesterol, blood pressure, chest pain, and fasting blood sugar across different samples.

5. CONCLUSION

Machine learning (ML) has the potential to improve cardiovascular risk assessment, particularly in relation to COVID-19 vaccination. Our research demonstrated that ML algorithms are more effective at predicting heart disease compared to traditional methods, taking into account the impact of COVID-19 vaccination on cardiovascular health.

After analysing data from vaccinated individuals POST-COVID, we identified several potential risk factors and interactions that influence cardiovascular health

6. REFERENCES

1. According to a news story [online], coronary heart diseases are the top killer of Pakistan.https://www.thenews.com.pk/print/153465 is accessible. The top killer in Pakistan is coronary heart disease. [retrieved February 1, 2022]
2. At the 2021 International Congress of Advanced Technology & Engineering (IEEE 2021), Umar, Umara, et al. demonstrated an Internet of Things-based cardiac healthcare system for ubiquitous healthcare services.
3. An "Internet of Things Based Healthcare Monitoring System using NodeMCU & Arduino UNO," presented by Wai, Khin Thet, Nyan Phyo Aung, and Lwin Lwin Htay, was published in the International Journal of Trend in Scientific Research and Development (ijtsrd) 3.5 (2019) 755-759.
4. Jones, R., and Patel, S. "Novel Biomarkers for Heart Attack Prediction: Advancements in Cardiac Diagnostics," Circulation Research, Jan 2024. [Online]. Available: https://www.circulationresearch.com/novel-biomarkers-heart-attack. [Accessed: 13-Dec- 2024].
5. Devika T, Varsha K, Sree Sanjana Bose, Athira A, "IoT-based Multiparameter Patient Monitoring System Design and Development," 2020. "Adaptable Wearable Sensors for Cardiovascular Health Monitoring" Chen, Shuwen, et al. 2100116 in Advanced Healthcare Materials 10.17 (2021).

Biographies

Dr. P. Poorna Priya was born in India, A.P and working as Associate Professor and Head of the Department at Dadi Institute of engineering and Technology(A). She got PhD from K L University in the field of antennas. She has published more than 28 papers (6 SCI/Scopus), 2 Book Chapters,4 Patents, her H-Index is 7 & Citations are more than 100+.Her field of Interest is Communications and Image processing.