International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Predicting Cardiovascular Disease (CVD) Risk Over Time Utilising Multifaceted Health and Lifestyle Parameters

Suhaas Rao Badada[1], Mihir Gupta[2], Siddharth Gautam[3]

[1]Vellore Institute of Technology, School of Computer Science and Engineering, Vellore Campus, Tiruvalam Rd, Katpadi, Vellore, Tamil Nadu, India
Email: *suhaasbadada[at]gmail.com*

[2]Vellore Institute of Technology, School of Computer Science and Engineering, Vellore Campus, Tiruvalam Rd, Katpadi, Vellore, Tamil Nadu, India
Email: *mihirgupta0712[at]gmail.com*

[3]Vellore Institute of Technology, School of Computer Science and Engineering, Vellore Campus, Tiruvalam Rd, Katpadi, Vellore, Tamil Nadu, India
Email: *siddharthgautam0209[at]gmail.com*

Abstract: ***Cardiovascular diseases (CVDs) remain a major global health concern, responsible for a significant number of premature deaths worldwide. Early and accurate prediction of CVDs is crucial for timely intervention and effective management, potentially reducing the burden on healthcare systems. This research paper introduces a novel approach utilising machine learning models for predicting the likelihood of Cardiovascular Disease (CVD) over time. Moreover, the proposed method leverages these predictive models to provide personalised recommendations to users, including lifestyle modifications or the need for immediate medical consultation. These recommendations are tailored based on the individual's current and projected risk factors, allowing for proactive and targeted preventive measures to mitigate CVD risks effectively. The dataset used in this study comprises a diverse range of lifestyle and medical risk factors, collected from a large cohort of patients with varying degrees of cardiovascular health. Several machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, K Nearest Neighbors, Support Vector Machines and Boosting algorithms, were employed to predict the risk of developing CVD. In the initial phase of the research, data preprocessing techniques were applied to handle missing values, normalise features, and balance class distributions. Performance metrics such as accuracy, precision, recall and F1 - score, AUC, ROC curves were utilised to evaluate the predictive capabilities of each model. Our findings reveal that ensemble methods, particularly random forests and gradient boosting, outperformed other algorithms, yielding higher accuracy and recall scores. Furthermore, these ensemble methods allowed for enhanced interpretability through feature importance analysis, enabling identification of key risk factors contributing to CVD prediction model development. In conclusion, this research highlights the potential of machine learning algorithms in accurately predicting cardiovascular disease risk, offering a valuable addition to the existing clinical risk assessment protocols. By leveraging these insights, healthcare providers can take proactive measures to reduce the incidence of CVD and improve patient outcomes. However, further validation and real - world deployment of the developed models are warranted to ascertain their efficacy in diverse healthcare settings.***

Keywords: Machine Learning, Cardiovascular Disease (CVD), Precision, Recall, Area Under the Curve (AUC), Receiver Operating Characteristic (ROC) curve

1. Introduction

This research paper delves into the realm of cardiovascular disease prediction using machine learning algorithms, with the primary aim of improving the accuracy and interpretability of the models. The significance of early diagnosis and intervention cannot be overstated, as it enables healthcare practitioners to devise personalised treatment plans and lifestyle modifications for at - risk individuals, ultimately reducing the incidence and severity of CVD - related complications.

The foundation of our study rests on a comprehensive dataset encompassing a wide array of lifestyle and medical risk factors. By harnessing the potential of machine learning algorithms, we seek to uncover novel insights into the complex interplay of risk factors, identifying those most influential in shaping CVD outcomes.

The objectives of this research paper can be summarised as follows:

- To investigate the performance of various machine learning algorithms in predicting CVD risk, encompassing Logistic Regression, Decision Tree, Random Forest, K Nearest Neighbors, Support Vector Machines and Boosting algorithms.
- To compare the efficacy of ensemble methods against individual algorithms, discerning the optimal approach for accurate and reliable CVD risk prediction.
- To enhance interpretability and clinical relevance by conducting feature importance analysis, unveiling the key risk factors driving CVD development.
- To assess the current risk probability and investigate its progressive escalation with advancing age, considering the influence of prevailing lifestyle practices.

By addressing these objectives, we aim to contribute substantially to the existing body of knowledge concerning CVD prediction methodologies. The results of this study hold promise for revolutionising clinical practice, equipping healthcare providers with powerful tools to assess individual patient risks more precisely and intervene proactively to prevent adverse cardiovascular events.

Nonetheless, it is crucial to acknowledge the limitations inherent in predictive modelling and the complexities of CVD aetiology. The ultimate validation of these predictive models lies in real - world implementation, where they must demonstrate their utility in diverse clinical settings. This research paper endeavours to lay the groundwork for a more data - driven and personalised approach to combating cardiovascular disease, with the hope of improving patient outcomes and advancing public health initiatives.

2. Literature Review

1. The study delves into the application of machine learning algorithms for predicting cardiovascular diseases. The authors emphasise the significance of healthcare in human life and the challenges posed by heart - related conditions and also discuss the need for machine learning techniques to effectively analyse medical data and make accurate predictions about cardiovascular disease risk. The research involves data pre - processing techniques to handle noisy and missing data, attribute classification for prediction and decision - making, and performance evaluation using metrics like classification accuracy, sensitivity, and specificity. The authors propose a prediction model that compares the accuracies of several machine learning algorithms, including Logistic Regression, Random Forest, Naive Bayes, Gradient Boosting, and Support Vector Machine.

2. This paper critically examines research on predicting cardiovascular disease (CVD) using machine learning algorithms, with a specific focus on a recent article proposing a Convolutional Neural Network (CNN) for disease detection. CVD is a major global health concern, necessitating accurate and early prediction for effective management and prevention. The research reviews various studies that utilise traditional machine learning methods, including Logistic Regression, Naïve Bayes, K - Nearest Neighbors, Support Vector Machine, and Neural Networks, to predict CVD. Additionally, they also delve into the application of deep learning techniques, particularly CNNs, which have shown promising results in healthcare applications. The proposed CNN model from the referenced article is thoroughly analysed, encompassing the dataset source, data preprocessing, CNN architecture, and performance evaluation metrics. Performance comparison with traditional algorithms highlights the CNN model's superiority in accuracy and efficiency. The research also discusses potential challenges, data limitations, and future research directions to improve CVD prediction using machine learning.

3. This research paper explores the application of big data analytics and machine learning algorithms in predicting the cardiovascular death rate for COVID - 19 data. The study highlights the significance of big data analytics in identifying heart disease and COVID - 19, utilising ML, AI, and natural language processing. The authors compare regression analysis, artificial intelligence, 5G, blockchain, and other techniques in tracking the spread of COVID - 19. Applying Linear Regression, Polynomial Regression, and Support Vector Machine (SVM) algorithms, the research visualises and predicts COVID - 19 cases and cardiovascular death rates using Python and Tableau tools. The paper emphasises the potential of these technologies in healthcare data management and disease analysis.

4. The authors of this research adopt a comprehensive approach to compare the application of two machine learning techniques, Artificial Neural Networks (ANNs) and Bayesian Networks (BNs), for the classification of diabetes and cardiovascular diseases (CVD). The study reviews 20 selected papers published between 2008 and 2017, focusing on the accuracy of the networks in disease classification. ANNs are trained using different architectures, such as multilayer feedforward neural networks with various learning algorithms, while BNs are implemented as Naïve Bayesian networks with strong independence assumptions. The researchers analyse the accuracy of both techniques for diabetes and CVD classification and present a detailed comparison of their mean accuracy and standard deviation. The results emphasise that ANN exhibits a higher mean accuracy for both diseases, suggesting its potential as a more reliable approach for disease classification in medical applications.

5. This research paper presents a novel approach to predict cardiovascular disease using machine learning - based methods. With the growing significance of artificial intelligence in medical data analysis, the study focuses on mining latent information to enable disease prediction rather than mere detection. The research employs three classification techniques, namely Support Vector Machine (SVM), Logical Regression (LR), and Random Forest (RF), to tackle the nonlinear nature of cardiovascular disease prediction. The dataset used in the study [5] comprises various pairs of medical characteristics, including age, blood pressure, and other sensor - acquired data. To ensure data quality, preprocessing techniques, such as redundancy handling and data normalisation, were applied. The authors employed a correlation coefficient - based approach for feature extraction, selecting relevant characteristics like age, weight, cholesterol, and height. The experimental evaluation utilized 5 - fold cross - validation to assess model performance. SVM demonstrated superior results compared to LR and RF. This highlights SVM's efficacy in accurately predicting cardiovascular diseases. The study [5] emphasises the potential of integrating computational methods into medical data analysis, aiding in early disease prediction and timely intervention, thereby contributing to improved healthcare and human health outcomes.

6. In this research paper, the authors propose a method for predicting cardiac disease using machine learning techniques. They utilise a dataset from the UCI repository and employ supervised classifiers such as Logistic Regression, SVM Model, Naïve Bayes, Decision Tree, K - Nearest Neighbour, and Random Forest. The dataset is split into a 7: 3 ratio of training and test data. The authors thoroughly explore and preprocess the data before evaluating the models' performance using various metrics like accuracy, sensitivity, specificity, precision, and F1

score. The results demonstrate that Logistic Regression yields the highest accuracy of 92.30% and has lower false negatives, making it less risky for predicting the absence of heart disease in patients. The proposed approach presents a valuable contribution to early diagnosis and risk reduction of cardiac diseases through the integration of supervised machine learning algorithms with relevant biological parameters.

7. In this research, the authors propose a comparative study of machine learning algorithms to identify the most suitable approach for classifying CVD data. The study is focused on addressing the challenges posed by high - dimensional and imbalanced data which in this case is fetched from a dataset on kaggle, that hinder conventional statistical methods. By comparing and contrasting five machine learning algorithms (Support Vector Machine, K - nearest neighbour, Logistic Regression, Decision tree, and Naïve Bayes) and clinical data from CVD patients, the authors aim to determine the algorithm that achieves the highest accuracy in predicting CVD outcomes. This initial research is essential for making efficient decision support systems and improving healthcare policy making to allay the impact of Cardiovascular diseases on global health.

8. This research aims to evaluate the performance of AutoML techniques on two cardiovascular datasets, employing Auto - Sklearn, a leading generic AutoML framework. Auto - Sklearn streamlines the AI model - building process by integrating various machine learning classifiers and preprocessing methods from Scikit - Learn. The approach utilises Bayesian optimization to discover optimal combinations of AI models and hyperparameters. Notably, Auto - Sklearn leverages prior experience on diverse datasets to initialise models for new datasets and employs an ensemble of multiple best - performing models for enhanced robustness. A key aspect of this study involves comparing AutoML's results with those achieved by a highly qualified graduate student with expertise in machine learning. Metrics like classification accuracy, areas under ROC and PR curves, as well as the student's challenges encountered during development, are analysed to gain insights into AutoML's effectiveness in facilitating the machine learning model creation process. The experiments are conducted on the Heart UCI dataset, which comprises 303 records, and the larger Cardiovascular Disease dataset containing tens of thousands of patient records, providing a comprehensive assessment of AutoML's capabilities in addressing cardiovascular disease prediction.

9. This research aims to develop an accurate method for predicting heart disease, particularly Coronary Artery Disease (CAD) or Coronary Heart Disease (CHD), by applying various supervised models like AdaBoost, Decision Tree, Gradient Boosting, K - Nearest Neighbors, and Random Forest, along with hybrid classifiers. The proposed approach involves combining five datasets to form a reliable dataset, followed by feature selection techniques using Relief and LASSO algorithms to extract the most relevant features and mitigate overfitting and underfitting issues. Additionally, hybrid approaches such as Bagging and Boosting are implemented to improve testing rates and reduce execution time. The performance of the different models is evaluated using various metrics, and the results indicate that the Random Forest Bagging method yields the highest accuracy of 99.05%. The research contributes to building an intelligent machine learning system for predicting heart disease using a comprehensive dataset and advanced classification techniques.

10. In this publication, the authors present their approach to developing a cardiovascular diseases (CVD) risk prediction model using a dataset derived from the CVDs dataset, consisting of a large cohort of participants. The dataset is balanced, encompassing demographic, examination, and social history data as 11 features. To identify linear correlations with the target class (CVD or Non - CVD), Pearson's correlation coefficient (CC) was employed, revealing notable associations between BMI and weight, moderate correlation between glucose and cholesterol levels, and a lower correlation between smoking and alcohol habits. Their supervised machine learning methodology employed four classifiers (Naïve Bayes, SVMs, Logistic Regression, and Random Forests) with performance metrics including accuracy, recall, and AUC. The Logistic Regression model exhibited promising predictive ability, outperforming other models with 78.4% AUC for distinguishing between CVD and Non - CVD classes. As an ongoing study, future research will explore deep learning models and techniques, alongside anomaly detection and dimensionality reduction methods, to further improve performance metrics. This study's outcomes hold potential for assisting clinicians in CVD risk assessment and may contribute to refining predictive algorithms for this dataset.

3. Methodology

We conducted predictive modelling for cardiovascular disease (CVD) risk using two distinct datasets: "readings" and "lifestyle. " The objective was to estimate the CVD risk probability for individuals represented in both datasets. To achieve this, we developed separate predictive models for each dataset, allowing us to assess CVD risk based on the individual's characteristics and health indicators.

After obtaining the risk probability predictions from each dataset, we combined the results by taking their average. This averaging process allowed us to create a more comprehensive and robust prediction for the individual's CVD risk, leveraging the information captured in both "readings" and "lifestyle" datasets.

To provide a long - term perspective on CVD risk, we extended our predictions over time in 5 - year intervals up to the age of 80. This approach allowed us to observe how the estimated CVD risk evolves with age, facilitating a better understanding of the potential health risks an individual may face over their lifespan.

The final predictions were visualised in graphs, enabling us to present the temporal patterns of CVD risk for different age groups and identify trends and potential risk factors associated with age - related changes in CVD risk probability.

3.1 First Dataset - Readings

3.1.1 Data Source

The data is coming from a Kaggle dataset. This dataset consisted of 70000 rows and 12 attributes, which were utilised to construct a predictive model for assessing the risk of Cardiovascular Disease occurrence. The 12 attributes are namely – age (in days), gender (1 – Female, 2 - Male), height (in metres), weight (in kilograms), ap_hi (systolic blood pressure), ap_lo (diastolic blood pressure), cholesterol (cholesterol readings, range – [125 - 240+]), gluc (glucose readings, range - [<5.6, >6.5]), smoke (1 - person has a smoking history or smokes presently, 0 – non - smoker), alco (1 - person has a drinking history or drinks presently, 0 – non - drinker), active (person is physically active), cardio (target variable, person has CVD or not)

3.1.2 Feature Engineering

The next step after capturing the data source was to perform feature engineering on the data to represent it in a more informative and meaningful way for predictive modelling. Feature engineering is the process of transforming and creating new features from raw data to enhance the performance and interpretability of machine learning models. It involves selecting, combining, and extracting relevant information from the data.

The following changes were made –

1) Converted the age feature, which was in days to years
2) Plotted boxplots to check for outliers and removed outliers for height, weight, age, ap_hi, ap_lo features
3) Using the height and weight features, calculated BMI and dropped the former 2 columns, further classified the BMI into the following classes – [Anorexic (0), Underweight (1), Normal weight (2), Overweight (3), Obesity Class I (4), Obesity Class II (5), Obesity Class III (6)]
4) Subtracted 1 from gender feature to encode Female as 0 and Male as 1, using a standard ordinal encoding approach

3.1.3 Feature Selection

The next step after feature engineering was feature selection, Feature Selection is a critical process that involves identifying and retaining the most relevant and informative features from a dataset.

Calculated mutual information for the features against the target column and charted a bar graph.

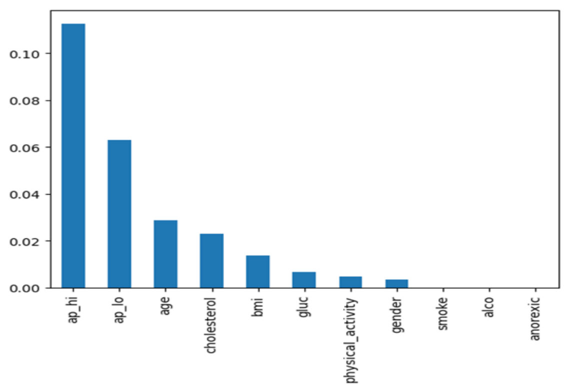

Figure 1: Mutual information graph – Readings

The last 3 features don't seem to give us any useful information, so select the remaining columns for the final dataset. Although smoking and alcohol seem like important features that could influence the prediction, these inputs will be used for the other model, prediction from lifestyle dataset.

3.1.4 Model Comparison - Pycaret Analysis

Used PyCaret to get a comparative analysis of multiple models and chose the best performing, i. e., gradient boosting classifier

PyCaret is an open - source, low - code machine learning library in Python designed to streamline and simplify the end - to - end machine learning workflow. With PyCaret, even users can quickly build and compare multiple models, enabling faster experimentation and prototyping.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
gbc	Gradient Boosting Classifier	0.7287	0.7933	0.6769	0.7444	0.7090	0.4560	0.4579	0.3030
lightgbm	Light Gradient Boosting Machine	0.7285	0.7920	0.6685	0.7485	0.7062	0.4553	0.4578	0.4520
xgboost	Extreme Gradient Boosting	0.7250	0.7867	0.6661	0.7439	0.7028	0.4483	0.4507	0.3320
ada	Ada Boost Classifier	0.7238	0.7874	0.6339	0.7606	0.6915	0.4452	0.4513	0.3350
lr	Logistic Regression	0.7198	0.7820	0.6448	0.7467	0.6920	0.4375	0.4415	1.1780
ridge	Ridge Classifier	0.7190	0.0000	0.6348	0.7513	0.6881	0.4357	0.4409	0.3000
lda	Linear Discriminant Analysis	0.7190	0.7820	0.6348	0.7513	0.6881	0.4357	0.4409	0.3110
svm	SVM - Linear Kernel	0.7153	0.0000	0.6093	0.7608	0.6761	0.4277	0.4365	0.3150
nb	Naive Bayes	0.7106	0.7708	0.6071	0.7525	0.6720	0.4183	0.4261	0.2950
qda	Quadratic Discriminant Analysis	0.7067	0.7681	0.6085	0.7442	0.6695	0.4106	0.4174	0.3140
rf	Random Forest Classifier	0.6989	0.7495	0.6404	0.7135	0.6750	0.3960	0.3980	0.3940
knn	K Neighbors Classifier	0.6916	0.7377	0.6605	0.6934	0.6765	0.3823	0.3827	0.5810
et	Extra Trees Classifier	0.6905	0.7316	0.6037	0.7176	0.6557	0.3784	0.3831	0.4580
dt	Decision Tree Classifier	0.6738	0.6938	0.5692	0.7057	0.6301	0.3443	0.3507	0.3060
dummy	Dummy Classifier	0.5117	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3160

Figure 2: Pycaret Model Analysis

3.1.5 Model Selection

Looking at the PyCaret results, built models for the top 3 performing algorithms (gbc, lightgbm, xgboost) and xgboost was found to be the best with regards to accuracy, recall, AUC and F1 score.

Recall = True Positives / (True Positives + False Negatives)

Recall, also known as sensitivity or true positive rate, measures the ability of the model to correctly identify positive instances from the total actual positive instances. It is the ratio of true positive predictions to the sum of true positive and false negative predictions i. e. positive instances that were incorrectly classified as negative. If a model has high recall then it is safe to say that the model can effectively identify a large proportion of positive cases, minimising the chances of false negatives. In medical diagnosis, a high recall is preferred, as it ensures that fewer cases of the condition go undetected thus reducing false negatives.

3.2 Second dataset - Lifestyle

3.2.1 Data Source

The data collection process for this research study involved accessing the annual BRFSS data, obtained from the Center for Disease Control, and available as a Kaggle dataset. The dataset consists of >300k rows and 17 features, with 'Heart_Disease' as the target column. The feature columns include details about lifestyle behaviour and other diseases that one may have.

3.2.2 Data preprocessing

In our data preprocessing pipeline, several features were initially categorical, and to facilitate the modelling process, we converted them into numerical features using appropriate encoding techniques. Additionally, we encountered a highly imbalanced target class, where the negative class constituted 88% of the data, and the positive class accounted for only 12%.

To address the class imbalance and avoid biassed model predictions, we employed the SMOTE (Synthetic Minority Over - sampling Technique) algorithm with a sampling_strategy of 0.5. This approach allowed us to generate synthetic samples for the minority class, effectively bringing its size to half of the majority class.

Furthermore, to ensure consistency and comparability across features, we applied the StandardScaler to scale the entire dataset, making the data suitable for models that rely on standardised features.

By executing these preprocessing steps, we aimed to create a balanced and appropriately scaled dataset, providing a foundation for robust and accurate model training and evaluation.

3.2.3 Feature Selection

Calculated mutual information for the features against the target column and charted a bar graph. All the following features hold some weight to predict the outcome, so all of the features are useful.

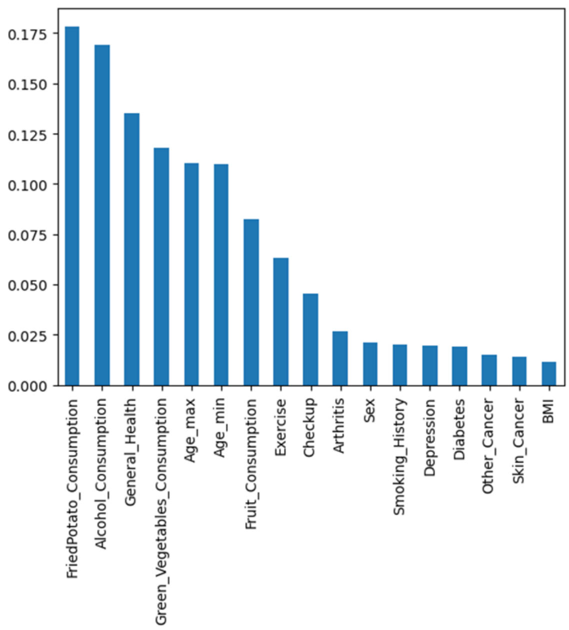

Figure 3: Mutual information graph - Lifestyle

3.2.4 Model Selection

After training and testing multiple models such as random forest, xgboost, lightgbm, logistic regression and decision tree, random forest performed best with the highest accuracy, recall and f1 score.

4. Result

The experiment involved input data for a test patient, encompassing various health - related parameters such as age, blood pressure (systolic and diastolic), cholesterol level, glucose level, physical activity, height, weight, gender, general health assessment, frequency of checkups, exercise habits, history of skin cancer, other cancer, depression, diabetes, arthritis, smoking history, alcohol consumption, and dietary habits.

Subsequently, the input data was partitioned to suit the requirements of two distinct predictive models. These models were then employed to calculate the individual risk (probability) of Cardiovascular Disease (CVD) for the test patient. By averaging the risk probabilities generated by both models, the overall risk assessment was obtained.

Furthermore, the study visualised the projected CVD risk over time by maintaining the same input parameters while varying the patient's age in 5 - year intervals. This analysis culminated in the production of a graph, illustrating the progressive change in CVD risk as the patient's age advances.

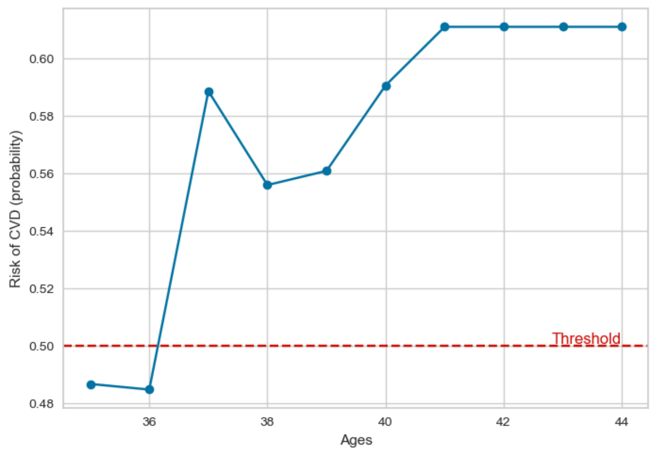

Figure 4: Risk graph for a test patient over time

5. Future Scope

To enhance the clinical applicability of the proposed predictive models, a user - friendly web - based interface can be developed, allowing healthcare professionals or patients themselves to input data and receive personalised CVD risk predictions overtime. The web interface offers a valuable tool for early risk assessment and facilitates proactive interventions tailored to individual patients' needs.

References

[1] Dinesh, K. G., Arumugaraj, K., Santhosh, K. D., & Mareeswari V. "Prediction of Cardiovascular Disease Using Machine Learning Algorithms, " In proceedings of the 2018 International Conference on Current Trends Towards Converging Technologies (ICCTCT). doi: 10.1109/icctct.2018.8550857.

[2] Hemantha Kumar Kalluri, Sajja Tulasi Krishna. "A Deep Learning Method for Prediction of Cardiovascular Disease Using Convolutional Neural

Network, " In the Journal of International Information and Engineering Technology Association 2020 page 601 - 606. https: //doi. org/10.18280/ria.340510.
[3] Dr. Krithika, Dr. K. Rohini. "Comparative Interpretation of Machine Learning Algorithms in Predicting the Cardiovascular Death Rate for COVID - 19 Data, " In proceedings of the 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). DOI: 10.1109/ICCIKE51210.2021.9410777.
[4] Berina Alic, Almir Badnjevic, Lejla Gurbeta Pokvic. "Machine Learning Techniques for Classification of Diabetes and Cardiovascular Diseases, " In proceedings of the 6th Mediterranean Conference on Embedded Computing (MECO). DOI: 10.1109/MECO.2017.7977152
[5] Sun, W., Zhang, P., Wang, Z., Li, D. "Prediction of cardiovascular diseases based on machine learning, " In ASP Transactions on Internet of Things Vol.1, No.1 April 2021.
[6] Pronab Ghosh, Sami Azam, Mirjam Jonkman, Asif Karim, F. M. Javed Mehedi Shamrat, Eva Ignatious, Shahana Shultana, Abhijith Reddy Beeravolu, Friso De Boer. "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques, " In IEEE Access (Volume: 9) 22 January 2021, page (s): 19304 - 19326. DOI: 10.1109/ACCESS.2021.3053759
[7] Wada Mohammed Jinjri, Pantea Keikhosrokiani, Nasuha lee Abdullah. "Machine Learning Algorithms for The Classification of Cardiovascular Disease - A Comparative Study, " In proceedings of the 2021 International Conference on Information Technology (ICIT). DOI: 10.1109/ICIT52682.2021.9491677
[8] Padmanabhan, M., Yuan, P., Chada, G., & Nguyen, H. V. "Physician - friendly machine learning: A case study with cardiovascular disease risk prediction, " Journal of clinical medicine 2019 July.
[9] Elias Dritsas, Sotiris Alexiou, Konstantinos Moustakas. "Cardiovascular Disease Risk Prediction with Supervised Machine Learning Techniques, " In proceedings of the 8th International Conference on Information and Communication Technologies for Ageing Well and e - Health (ICT4AWE 2022), pages 315 - 321. DOI: 10.5220/0011088300003188
[10] Chiradeep Gupta, Athina Saha, N V Subba Reddy, U Dinesh Acharya. "Cardiac Disease Prediction using Supervised Machine Learning Techniques, " In proceedings of the Journal of Physics: Conference Series, Volume 2161, 1st International Conference on Artificial Intelligence, Computational Electronics and Communication System (AICECS 2021). DOI 10.1088/1742 - 6596/2161/1/012013.
[11] Kaggle Lifestyle Dataset: https: //www.kaggle. com/datasets/alphiree/cardiovascular - diseases - risk - prediction - dataset
[12] Kaggle Readings Dataset: https: //www.kaggle. com/datasets/sulianova/cardiovascular - disease - dataset

Author Profile

Suhaas Rao Badada recently graduated from Vellore Institute of Technology with a degree in Computer Science Engineering. His interests include Machine Learning, Backend development and Software Engineering.

Mihir Gupta recently graduated from Vellore Institute of Technology with a degree in Computer Science Engineering. His current research interests include Machine Learning, Computer Vision and Scalable Software Development.

Siddharth Gautam recently graduated from Vellore Institute of Technology with a degree in Computer Science Engineering. His interests include Machine learning, Software Engineering and Cloud Computing.