Explainable Machine Learning Algorithms to Predict Cardiovascular Strokes

Afia Fairooz Tasnim
Department of Public Health, California State University, Long Beach, USA
afia.tasnim009@gmail.com (corresponding author)

Rukshanda Rahman
Department of Computer Science, Westcliff University, Irvine, USA
r.rahman.562@westcliff.edu

Mani Prabha
Department of Business Administration, International American University, Los Angeles, USA
maniprabha612@gmail.com

Md. Azhad Hossain
Department of Business Administration, International American University, Los Angeles, USA
azhad17@gmail.com

Sadia Islam Nilima
Department of Business Administration, International American University, Los Angeles, USA
sadiaiislam94@gmail.com

Md. Abdullah Al Mahmud
Department of Business Administration, International American University, Los Angeles, USA
abdullahiau1@outlook.com

Timotei Istvan Erdei
Department of Vehicles Engineering, Faculty of Engineering, University of Debrecen, Hungary
timoteierdei@eng.unideb.hu

Received: 1 October 2024 | Revised: 30 October 2024 and 23 November 2024 | Accepted: 25 November 2024

 DOI: https://doi.org/10.48084/etasr.9152

ABSTRACT

Cardiovascular disease has been more common throughout the past several decades. Cardiovascular disease detection methods use machine learning algorithms to assess data and provide accurate cardiac diagnosis. An accurate and comprehensive assessment of cardiovascular risk is essential to improve cardiovascular protection and reduce the frequency and severity of heart attacks and strokes. This paper proposes a machine learning-based autonomous strategy for the diagnosis of cardiovascular disease. Some preprocessing methods were applied to improve the results and accuracy. Finally, lazy prediction was used to find the best model by applying a neural network and two ensemble models. The best accuracy of 99% was obtained with the HistGradientBoosting (ensemble) classifier, which obtained respectable results with a higher accuracy rate. This model can enhance the ability to predict cardiovascular disease with better accuracy.

Keywords-cardiovascular; healthcare system; machine learning; ensemble; social development

I. INTRODUCTION

Since the heart controls blood circulation, it is the most vital organ in the human body. A multitude of individuals suffer from various cardiac diseases. Cardiovascular Disease (CVD) is becoming a major risk factor for heart disease. CVD accounted for about half of all deaths in the United States and other countries [1]. Smoking, excessive alcohol, poor eating habits, and inactivity are the main behavioral risk factors for heart diseases and strokes [2, 3], while air pollution is one of the major environmental factors. People with behavioral risk factors may also be overweight or have high blood pressure, high blood sugar, or high cholesterol. In primary care settings, these intermediate risk factors can be assessed to determine an increased risk of heart attack, stroke, heart failure, and other consequences.

Identifying and treating CVD-vulnerable people can prevent premature deaths. All basic medical facilities must have basic medical technology and drugs to treat and guide patients [4]. If CVD is diagnosed early, people will be aware of this problem and motivated to treat it. Data analysis utilizing machine learning algorithms helps to diagnose CVD. Unfortunately, CVDs have increased worldwide in the past 40 years. More than four out of five CVD deaths are caused by stroke or coronary artery disease, and one-third of those under 70 years of age die too soon [4, 5]. In 2023, 51% of the respondents did not list heart disease as the leading cause of death [6]. Heart failure (9.1%), high cholesterol (13.4%), stroke (17.5%), numerous major cardiovascular causes (17.1%), chronic heart disease (40.3%), and arterial problems (2.6%) cause the majority of cardiovascular deaths [7, 8].

Careful feature selection and extraction are needed to develop a challenging task-performing artificial neural network, reducing data noise, eliminating unneeded information, and improving model accuracy and efficiency [9-14]. Cleveland Clinic built machine learning models to identify cardiovascular disease using 70,000 patient records [15]. The Cleveland dataset has 12 rows and 70000 columns. Decision Tree (DT), LGBM, XGB, neural network, Logistic Regression (LR), K-nearest Neighbors (KNN), and Naive Bayes (NB) have been used to identify heart disease [16, 17]. The XGB model achieved the highest accuracy with 77.35% when excluding BMI, while including BMI improved the accuracy to 83.52%. BMI is crucial to identifying cardiovascular disease [18, 19]. In [18], a machine learning-based CVD classification method was presented, utilizing 1025 patient records from a Kaggle dataset with 14 attributes. Data are preprocessed before modeling [21, 22]. In [23], machine learning was used to detect CVD. This study employed the Cleveland Heart Disease database using 13 factors. The best accuracy was achieved with neural networks, NB, and DT. In [25], machine learning was used to classify CVD. Data mining challenges were solved using KEEL while missing values were handled with All Possible-MV. The DT classifier with hypertuning was 86.7% accurate. In [26], machine learning was used to detect cardiovascular disease. This trial used 303 cases from the Cleveland Heart Disease dataset and additional patients were manually surveyed using all 13 input attributes, and RF achieved the highest accuracy of 88.16%. In [28, 29], machine learning models were used to detect CVD.

Many studies have used machine learning to diagnose CVD but most of them used older machine learning methods and datasets, while some omitted data preprocessing. This study used cutting-edge machine learning models and inventive data pretreatment approaches [31, 32]. To avoid overfitting, data mining, feature selection, binary classification, data cleaning, and other methods were used during preprocessing. The dataset was also analyzed using DT, LR, RF, and GB. This work makes the following notable contributions:

- Uses preprocessing techniques to the collected data, which include 12 features and 1000 subjects.
- Lazy Predict was used to identify the best model based on accuracy.

This study used a Lazy Predict technique, a neural network, and two ensemble models to build an automated system that uses the Cardiovascular_Disease_Dataset to identify CVD. The Lazy Predict technique and the dataset make it unique compared to other approaches.

II. METHODOLOGY

A. Dataset

The Cardiovascular_Disease_Dataset [33] has 1000 subjects and 12 characteristics. The target field predicts CVD based on heart health, which is 1 for heart disease or 0 for no heart disease. Figure 1 shows some key dataset properties:

- Patient ID contains a patient's unique serial number.
- Age consists of numerical values in years.

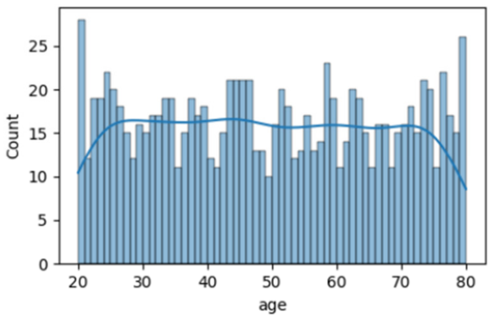

Fig. 1. Age columns visualization.

- Gender: There are 76.5% male and 23.5% female. Figure 2 represents the age feature, which is a categorical column.
- Chest pain type: There are four types of chest pain. 0 is for typical angina, 1 is for atypical angina, 2 is for non-anginal pain, and 3 is for asymptomatic pain. Figure 3 represents the chest pain feature.
- Fasting blood sugar contains two values: True indicates people who are fasting for blood sugar, and False indicates people who are not. Figure 4 represents the fasting blood sugar feature.

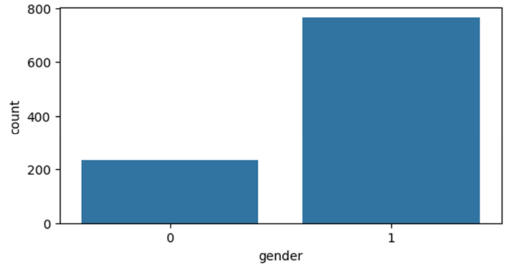

Fig. 2. Gender columns visualization.

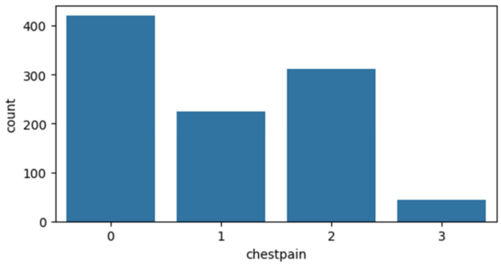

Fig. 3. Chest pain columns visualization.

- Maximum heart rate achieved: The range of values is 71-202. Figure 5 represents the max heart rate feature.
- Target: 0 indicates the absence of heart disease, and 1 denotes its presence. Figure 6 represents the target feature.

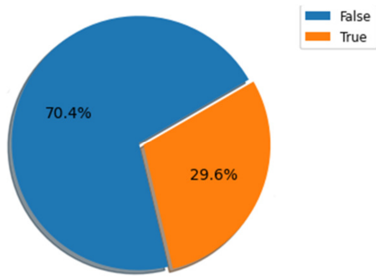

Fig. 4. Fasting blood sugar columns visualization.

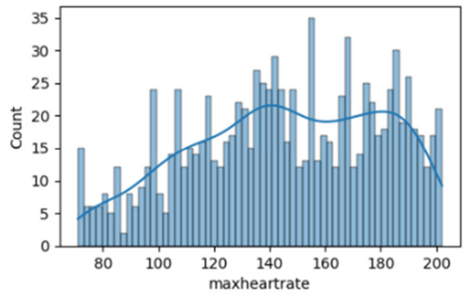

Fig. 5. Max heart rate columns visualization.

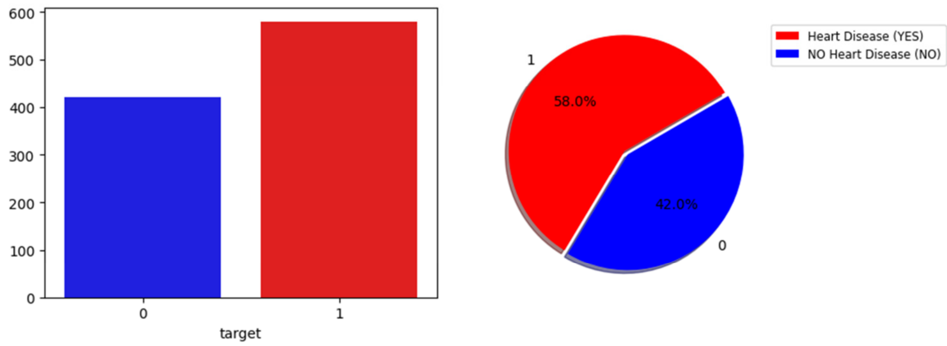

Fig. 6. Target columns visualization.

B. Preprocessing

The process of preprocessing raw data for machine learning models is also known as data preparation. Unaltered data collected from multiple sources is not suitable for analysis. Data preparation is the most difficult and time-consuming part of data science. The main purposes of data preparation are to eliminate data problems, such as missing numbers, enhance their quality, and prepare them for machine learning. This study employed a range of preprocessing techniques.

- Enhancing data quality: In machine learning, the preparation of data is crucial to improve data quality and lay the groundwork for reliable conclusions. EDA is most important, helping to create a good picture of the dataset.
- Data manipulation: Data manipulation was performed via pandas.get_dummies(). This function transforms indicator or dummy variables from categorical data. This technique was applied to all categorical columns of the dataset.

As the dataset did not contain missing or duplicate values, no such technique was used.

C. Feature Selection

Understanding feature selection is essential to build an automated learning model. Including every parameter in a dataset is rarely beneficial when developing a model for real-world data science applications. Repetitive factors weaken a classifier's capacity to generalize in addition to lowering its overall accuracy. Furthermore, the total complexity of a model increases with the addition of new variables. Feature selection is the process of selecting a specific set of qualities from the initial set of features to reduce the feature space as much as possible while still meeting predetermined criteria. A correlation matrix was created using the dataset, which helped to understand the relation. The correlations between all features and the target were analyzed. Chest pain, slope, no major vessels, and resting BP have a high correlation with the target variable. Figures 7 and 8 represent the correlation matrix between features and the correlations between the features and the target variable, respectively.

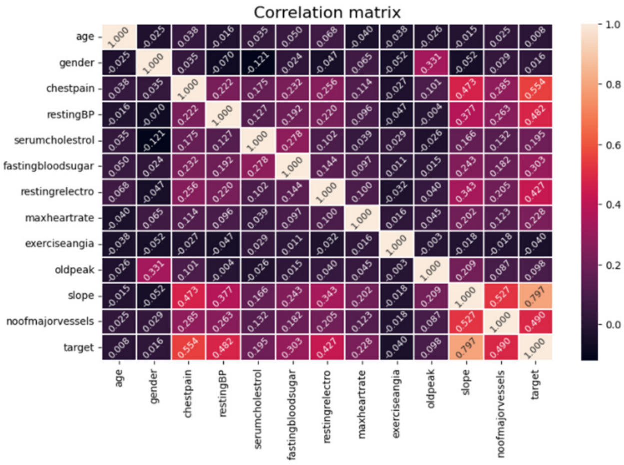

Fig. 7. Correlation matrix.

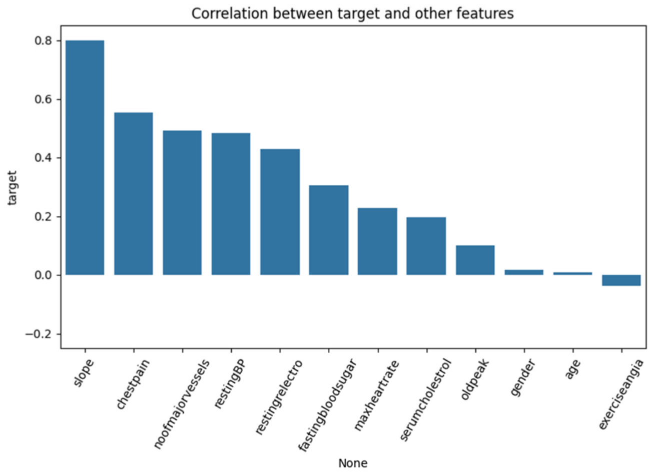

Fig. 8. Correlation between target and other features.

D. Data Splitting

The dataset was divided into a 70:30 ratio for training and testing. The training set was used to build the model and the testing set was used to assess how well the predictive model performs.

E. Machine Learning Algorithms

1) Lazy Predict

The dataset was processed using Lazy Predict, which is a Python library that simplifies machine learning by providing a single interface for several algorithms. With just a few lines of code, users may quickly construct and evaluate many models, allowing them to identify the most promising solutions for a given problem without wasting time on systems that are unlikely to succeed. The dataset was divided using train_test_split(). Then LazyClassifier() was utilized to fit the models using a random state of 42.

2) Gradient Boosting (GB)

GB is a robust boosting strategy that combines several weak learners into powerful ones by training each new model to reduce loss functions such as the mean square error or cross-entropy using gradient descent. This technique calculates the variance function slope compared to the ensemble's expectations in each iteration and trains a new model to reduce this gradient [34]. The new model predictions are added to the groups, and the cycle continues until an interruption threshold is met. This process of adding models continues until the model can reliably predict a complete training dataset.

3) Multi-Layer Perceptron (MLP) Classifier

MLPs are multilayered neural networks with a feedforward function. The MLP classifier offers options for network construction, model learning, and effectiveness assessment. Text, image, and time series classification can be performed using it.

4) Histogram-Based Gradient Boosting

The GB ensemble machine learning technique builds models one after another to remedy the flaws of the previous model. Because of its high prediction accuracy, it is often used for regression and classification problems. Histogram-based GB (HistGB) improves algorithm efficiency by discretizing continuous input data in bins (histograms). This strategy dramatically reduces memory and computational complexity. Histograms help the HistGBClassifier handle large datasets better than other GB methods. This is excellent for handling hundreds or tens of thousands of samples. As this classifier supports missing values, it can handle partial datasets without imputation. This technique scales well with samples and features, making it suitable for high-dimensional data. GB works well in reality, but simulation training is arduous. Trees must be built and added sequentially, unlike RF, which can be trained simultaneously using multiple CPU cores. Numerous strategies have been developed to improve the efficiency of GB training.

5) Decision Tree (DT)

DTs are useful in statistics, data mining, and machine learning, as they can demonstrate the connection between parts, supporting data-driven decision-making. A flowchart-like framework predicts or makes decisions. The nodes represent attribute judgments or tests, the branches show their outcomes, and the leaf nodes reflect the determinations or forecasts. Each leaf node represents a category or continuous value, each internal node an attribute test, and each branch the test's result. DTs start at the root node to predict dataset categories. This method branches to the next node by comparing the initial property values with the data (actual dataset) attribute values. Again, comparing attribute values with previous subnodes leads to the next node. It does this until it reaches the tree's leaf node. DT learning divides and conquers by greedily searching for the optimal tree split points. After classifying all or most records, this division is recursive and top-down. The DT complexity is crucial when classifying data points as homogeneous. Smaller trees can fit more information in one class, or pure leaf nodes.

III. RESULTS AND DISCUSSION

This study applied machine learning techniques to diagnose CVD. The Cardiovascular_Disease_Dataset consists of 1000 subjects. The key features are obtained through feature selection and those that were not needed were removed. Then Lazy Predict was used to examine the models that worked better with the dataset, as shown in Table I. Some models achieved 97% accuracy, which is satisfactory. A neural network and two ensemble models were applied to enhance accuracy. As shown in Table II, the HistGB classifier achieved the highest F1-score of 0.987 and was finally chosen.

TABLE I. LAZY CLASSIFIER RESULTS

Model	Accuracy	Balanced accuracy	ROC AUC	F1 score	Time taken
LR	0.94	0.94	0.94	0.94	0.27
LinearSVC	0.94	0.94	0.94	0.94	0.16
SVC	0.96	0.96	0.96	0.96	0.07
SGD	0.94	0.94	0.94	0.94	0.60
Calibrated CV	0.94	0.94	0.94	0.94	0.37
Passive-Aggressive	0.92	0.93	0.93	0.92	0.05
ExtraTrees	0.93	0.93	0.93	0.93	0.45
KNN	0.95	0.95	0.95	0.95	0.05
RF	0.97	0.97	0.97	0.97	0.82
XGB	0.97	0.97	0.97	0.97	0.32
AdaBoost	0.96	0.96	0.96	0.96	0.99
LGBM	0.97	0.97	0.97	0.97	0.79
GaussianNB	0.95	0.95	0.95	0.95	0.20
QDA	0.97	0.97	0.97	0.97	0.13
Bagging	0.95	0.95	0.95	0.95	0.46
BernoulliNB	0.93	0.93	0.93	0.93	0.42
Label Spreading	0.95	0.95	0.95	0.95	2.91
Label Propagation	0.95	0.95	0.95	0.95	2.58
ExtraTree	0.91	0.91	0.91	0.91	0.21
NuSVC	0.95	0.95	0.95	0.95	1.69
DT	0.96	0.96	0.96	0.96	0.45
RidgeCV	0.95	0.95	0.95	0.95	0.10
Ridge	0.99	0.99	0.99	0.99	0.10
LDA	0.94	0.94	0.94	0.94	0.14
Nearest Centroid	0.92	0.92	0.92	0.92	0.16
Perceptron	0.93	0.94	0.94	0.93	0.11
Dummy	0.54	0.50	0.50	0.38	0.34

TABLE II. THREE MODELS WITH F1 SCORES

Models	F1 score
MLPClassifier	0.973
GradientBoostingClassifier	0.983
HistGradientBoostingClassifier	0.987

A. Confusion Matrix

A confusion matrix can be used to accurately assess an AI classification model. Totals include False Positives (FP), True Positives (TP), False Negatives (FN), and True Positives (TP). This provides a comprehensive overview of the various error types and performance metrics associated with the classification model. Figure 9 represents the confusion matrix using the HistGB classifier.

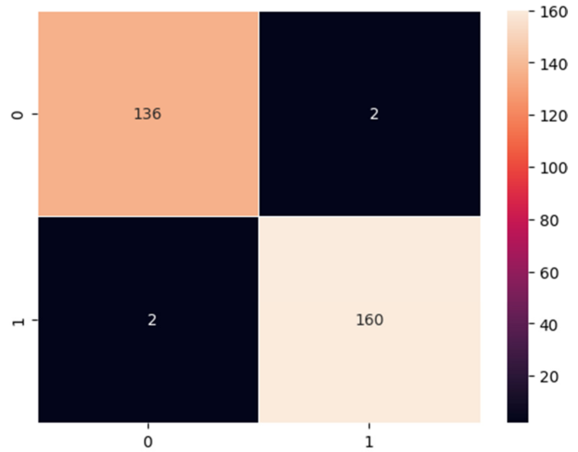

Fig. 9. Confusion matrix of the HistGB classifier.

B. Accuracy

Accuracy measures how often a machine learning model predicts the outcome correctly. The computation involves dividing the total number of guesses by the number of accurate forecasts. The accuracy of the model is a statistic that may be used to describe its performance in each class when all classes are of equal importance. The HistGB classifier achieved 0.99 accuracy with 0.99 precision, 0.99 recall, and 0.99 F1 score. Figure 10 shows the classification report for the HistGB classifier model.

```
              precision    recall  f1-score   support

           0       0.99      0.99      0.99       138
           1       0.99      0.99      0.99       162

    accuracy                           0.99       300
   macro avg       0.99      0.99      0.99       300
weighted avg       0.99      0.99      0.99       300
```

Fig. 10. Classification report of the HistGB classifier.

IV. CONCLUSIONS

This study used a large dataset and built three machine learning-based models to increase accuracy in predicting CVD. A variety of preprocessing techniques were used to prevent overfitting. The Lazy classifier was used to train the dataset, which made it easier to obtain high-quality models. Using a neural network and two ensemble models, an automated system that recognizes CVD using the Cardiovascular Disease Dataset was developed. Respectable results were obtained with the HistGB classifier. Multiple preparation techniques were used to ensure that the dataset was noise-free. The HistGB classifier achieved the highest accuracy of 0.99. This model has the potential to revolutionize medical research by rapidly and affordably diagnosing CVD for the sustainable development of the social healthcare system. Future studies will attempt to strengthen the accuracy and create software using those models along with advanced hyperparameter tuning.

ACKNOWLEDGEMENT

This research was supported by the University of Debrecen Program for Scientific Publication.

REFERENCES

[1] "Cardiovascular diseases," *World Health Organization*. https://www.who.int/health-topics/cardiovascular-diseases.

[2] A. Khosla, Y. Cao, C. C. Y. Lin, H. K. Chiu, J. Hu, and H. Lee, "An integrated machine learning approach to stroke prediction," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, Washington, DC, USA, Jul. 2010, pp. 183–192, https://doi.org/10.1145/1835804.1835830.

[3] M. W. Ullah, R. Rahman, S. I. Nilima, A. F. Tasnim, and M. B. Aziz, "Health Behaviors and Outcomes of Mobile Health Apps and Patient Engagement in the USA," *Journal of Computer and Communications*, vol. 12, no. 10, pp. 78–93, 2024, https://doi.org/10.4236/jcc.2024.1210007.

[4] N. G. B. Amma, "Cardiovascular disease prediction system using genetic algorithm and neural network," in *2012 International Conference on Computing, Communication and Applications*, Dindigul, Tamilnadu, India, Feb. 2012, pp. 1–5, https://doi.org/10.1109/ICCCA.2012.6179185.

[5] L. T. Braun *et al.*, "Palliative Care and Cardiovascular Disease and Stroke: A Policy Statement From the American Heart Association/American Stroke Association," *Circulation*, vol. 134, no. 11, Sep. 2016, https://doi.org/10.1161/CIR.0000000000000438.

[6] L. Soares-Miranda, D. S. Siscovick, B. M. Psaty, W. T. Longstreth, and D. Mozaffarian, "Physical Activity and Risk of Coronary Heart Disease and Stroke in Older Adults: The Cardiovascular Health Study," *Circulation*, vol. 133, no. 2, pp. 147–155, Jan. 2016, https://doi.org/10.1161/CIRCULATIONAHA.115.018323.

[7] R. L. Sacco *et al.*, "The Heart of 25 by 25: Achieving the Goal of Reducing Global and Regional Premature Deaths From Cardiovascular Diseases and Stroke: A Modeling Study From the American Heart Association and World Heart Federation," *Circulation*, vol. 133, no. 23, Jun. 2016, https://doi.org/10.1161/CIR.0000000000000395.

[8] A. Morgentaler, M. M. Miner, M. Caliber, A. T. Guay, M. Khera, and A. M. Traish, "Testosterone Therapy and Cardiovascular Risk: Advances and Controversies," *Mayo Clinic Proceedings*, vol. 90, no. 2, pp. 224–251, Feb. 2015, https://doi.org/10.1016/j.mayocp.2014.10.011.

[9] M. Rahman, Y. Wen, H. Xu, T. L. Tseng, and S. Akundi, "Data mining in telemedicine," *Advances in Telemedicine for Health Monitoring: Technologies, Design and Applications*, Jan. 2020, https://doi.org/10.1049/PBHE023E_ch6.

[10] A. Nikam, S. Bhandari, A. Mhaske, and S. Mantri, "Cardiovascular Disease Prediction Using Machine Learning Models," in *2020 IEEE Pune Section International Conference (PuneCon)*, Pune, India, Dec. 2020, pp. 22–27, https://doi.org/10.1109/PuneCon50868.2020.9362367.

[11] H. Kamel and J. S. Healey, "Cardioembolic Stroke," *Circulation Research*, vol. 120, no. 3, pp. 514–526, Feb. 2017, https://doi.org/10.1161/CIRCRESAHA.116.308407.

[12] N. N. Islam Prova, "Healthcare Fraud Detection Using Machine Learning," in *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, Coimbatore, India, Aug. 2024, pp. 1119–1123, https://doi.org/10.1109/ICoICI62503.2024.10696476.

[13] K. G. Dinesh, K. Arumugaraj, K. D. Santhosh, and V. Mareeswari, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms," in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, Coimbatore, Mar. 2018, pp. 1–7, https://doi.org/10.1109/ICCTCT.2018.8550857.

[14] K. A. Hicks *et al.*, "2017 Cardiovascular and Stroke Endpoint Definitions for Clinical Trials," *Circulation*, vol. 137, no. 9, pp. 961–972, Feb. 2018, https://doi.org/10.1161/CIRCULATIONAHA.117.033502.

[15] M. Juchartz, "Plan comparison of Hybrid Dynamic Conformal Arc Therapy (HDCAT) and 3-Dimensional Conformal Radiation Therapy (3DCRT) in palliative treatment of thoracic spine tumors," *Culminating Experience Projects*, Aug. 2024, Art. no. 453.

[16] B. H. Huang, M. J. Duncan, P. A. Cistulli, N. Nassar, M. Hamer, and E. Stamatakis, "Sleep and physical activity in relation to all-cause, cardiovascular disease and cancer mortality risk," *British Journal of Sports Medicine*, vol. 56, no. 13, pp. 718–724, Jul. 2022, https://doi.org/10.1136/bjsports-2021-104046.

[17] K. Adelborg *et al.*, "Migraine and risk of cardiovascular diseases: Danish population based matched cohort study," *BMJ*, Jan. 2018, Art. no. k96, https://doi.org/10.1136/bmj.k96.

[18] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Computers in Biology and Medicine*, vol. 136, Sep. 2021, Art. no. 104672, https://doi.org/10.1016/j.compbiomed.2021.104672.

[19] R. A. Bernstein *et al.*, "Effect of Long-term Continuous Cardiac Monitoring vs Usual Care on Detection of Atrial Fibrillation in Patients With Stroke Attributed to Large- or Small-Vessel Disease: The STROKE-AF Randomized Clinical Trial," *JAMA*, vol. 325, no. 21, Jun. 2021, Art. no. 2169, https://doi.org/10.1001/jama.2021.6470.

[20] J. M. Castellano *et al.*, "Polypill Strategy in Secondary Cardiovascular Prevention," *New England Journal of Medicine*, vol. 387, no. 11, pp. 967–977, Sep. 2022, https://doi.org/10.1056/NEJMoa2208275.

[21] I. Hussain and S. J. Park, "Big-ECG: Cardiographic Predictive Cyber-Physical System for Stroke Management," *IEEE Access*, vol. 9, pp. 123146–123164, 2021, https://doi.org/10.1109/ACCESS.2021.3109806.

[22] A. Khosla, Y. Cao, C. C. Y. Lin, H. K. Chiu, J. Hu, and H. Lee, "An integrated machine learning approach to stroke prediction," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, Washington DC USA, Jul. 2010, pp. 183–192, https://doi.org/10.1145/1835804.1835830.

[23] C. S. Dangare and S. S. Apte, "Improved study of heart disease prediction system using data mining classification techniques," *International Journal of Computer Applications*, vol. 47, no. 10, pp. 44–48.

[24] A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, Dec. 2019, Art. no. 211, https://doi.org/10.1186/s12911-019-0918-5.

[25] Purushottam, K. Saxena, and R. Sharma, "Efficient heart disease prediction system using decision tree," in *International Conference on Computing, Communication & Automation*, Greater Noida, India, May 2015, pp. 72–77, https://doi.org/10.1109/CCAA.2015.7148346.

[26] K. Battula, R. Durgadinesh, K. Suryapratap, and G. Vinaykumar, "Use of Machine Learning Techniques in the Prediction of Heart Disease," in *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, Mauritius, Mauritius, Oct. 2021, pp. 1–5, https://doi.org/10.1109/ICECCME52200.2021.9591026.

[27] D. Zhao, J. Liu, M. Wang, X. Zhang, and M. Zhou, "Epidemiology of cardiovascular disease in China: current features and implications," *Nature Reviews Cardiology*, vol. 16, no. 4, pp. 203–212, Apr. 2019, https://doi.org/10.1038/s41569-018-0119-4.

[28] Md. R. Ahmed, S. M. Hasan Mahmud, M. A. Hossin, H. Jahan, and S. R. Haider Noori, "A Cloud Based Four-Tier Architecture for Early Detection of Heart Disease with Machine Learning Algorithms," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, Chengdu, China, Dec. 2018, pp. 1951–1955, https://doi.org/10.1109/CompComm.2018.8781022.

[29] M. Y. Henein, S. Vancheri, G. Longo, and F. Vancheri, "The Role of Inflammation in Cardiovascular Disease," *International Journal of Molecular Sciences*, vol. 23, no. 21, Oct. 2022, Art. no. 12906, https://doi.org/10.3390/ijms232112906.

[30] E. L. Harshfield *et al.*, "Association Between Depressive Symptoms and Incident Cardiovascular Diseases," *JAMA*, vol. 324, no. 23, Dec. 2020, Art. no. 2396, https://doi.org/10.1001/jama.2020.23068.

[31] R. Eikelboom, R. Sanjanwala, M.-L. Le, M. H. Yamashita, and R. C. Arora, "Postoperative Atrial Fibrillation After Cardiac Surgery: A Systematic Review and Meta-Analysis," *The Annals of Thoracic Surgery*, vol. 111, no. 2, pp. 544–554, Feb. 2021, https://doi.org/10.1016/j.athoracsur.2020.05.104.

[32] D. Smajlovic, "Strokes in young adults: epidemiology and prevention," *Vascular Health and Risk Management*, Feb. 2015, Art. no. 157, https://doi.org/10.2147/VHRM.S53203.

[33] B. P. Doppala, D. Bhattacharyya, M. Janarthanan, and N. Baik, "A Reliable Machine Intelligence Model for Accurate Identification of Cardiovascular Diseases Using Ensemble Techniques," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–13, Mar. 2022, https://doi.org/10.1155/2022/2585235.

[34] Md. H. R. Sobuz *et al.*, "Optimization of recycled rubber self-compacting concrete: Experimental findings and machine learning-based evaluation," *Heliyon*, vol. 10, no. 6, Mar. 2024, Art. no. e27793, https://doi.org/10.1016/j.heliyon.2024.e27793.