Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 24s (2025)
https://www.jneonatalsurg.com

Predictive Analysis for Cardiovascular Outcomes Using Ai

Albatoul Khaled[1], Dr Avrina Kartika Ririe MD[2], Mandeep Kaur[3], Sahil Kumar[4], Newton Rahming[5], Rahif Khaled[6], Amine Hamdache[7], Abdelilah Jraifi[8], Ilias Elmouki[9]

[1]Doctor, MISCOM laboratory/ENSA Safi, Department of Computer Science, Networks and Telecommunications, UCA, Morocco,

Email ID: albatoulkhaled8@gmail.com

[2]Brain & Heart Center Department. Mohammad Hoesin General Hospital, Palembang,
Indonesia,

Email ID: avrinaririe@gmail.com

[3]MD, Department of Internal Medicine, HCA Capital Regional Med Center 2626 Capital Medical Blvd, Tallahassee, FL 32308, USA,

Email ID: kaurmandeep1880@gmail.com

[4]Master of Science, DePaul University, Chicago, IL, USA,

Email ID: skumar46@depaul.edu

ORCID ID: https://orcid.org/0009-0002-1054-4979

[5]MD, Caribbean Medical University, The Bahamas,

Email ID: andrewrahming10@gmail.com

[6]Medical Imaging Specialist, MISCOM laboratory/ENSA Safi, Department of Computer
Science, Networks and Telecommunications UCA, Morocco,

Email ID: ruhaifkhaled9@gmail.com

[7]Scholar, MISCOM laboratory/ENSA Safi, Department of Computer Science, Networks and Telecommunications. UCA, Morocco,

Email ID: hamdacheamine@gmail.com

[8]Scholar, MISCOM laboratory/ENSA Safi, Department of Computer Science, Networks and Telecommunications, UCA, Morocco,

Email ID: jraifi.abdelilah@gmail.com

[9]Scholar, MISCOM laboratory/ENSA Safi, Department of Computer Science, Networks and Telecommunications, UCA, Morocco,

Email ID: i.elmouki@gmail.com

Cite this paper as: Albatoul Khaled, Dr Avrina Kartika Ririe MD, Mandeep Kaur, Sahil Kumar, Newton Rahming, Rahif Khaled, Amine Hamdache, Abdelilah Jraifi, Ilias Elmouki, (2025) Predictive Analysis for Cardiovascular Outcomes Using Ai. *Journal of Neonatal Surgery*, 14 (24s), 147-157

ABSTRACT

Background

Cardiovascular diseases (CVD) continue to be one of the leading causes of deaths globally meaning that risk evaluation and predictive efforts need to start early to ensure proper management is done. The goal of this study is to conduct quantitative predictive analysis to find key predictive risk factors with the aid of machine learning models in hopes of improving cardiovascular outcomes.

Methods

A cross-sectional survey design was employed by reaching out to 273 respondents through a standardized questionnaire. Together with demographic factors, lifestyle, and medical history information about the respondents that is relevant to the risk of CVD was collected as well. Other tests that were performed include normality tests (Shapiro-Wilk), reliability tests (Cronbach's Alpha), and correlational tests. The data was predictively modeled using Logistic Regression, Random Forest, and Decision Tree classifier models evaluated based on accuracy, precision, recall, and F1 score.

Results

The normality test proved that the continuous variables of height, weight, and BMI do indeed support and conform to a normal distribution. The test also suggests there is a low internal consistency which is led by the low Cronbach's alpha value of 0.037 meaning that cardiovascular risk assessment is multidimensional. The results of predictive modeling were overall relatively low and failed to be predicted reliably but random forest achieved the highest performance as expected (accuracy: 16.36%) though was still much too low. Class imbalance and lack of predictive features are likely causes of model performance.

Conclusion

The results underscore the difficulties associated with estimating cardiovascular outcomes with the traditional machine learning models. The study points out that incorporating advanced feature selection, bigger datasets, and more sophisticated AI tools like deep learning and real-time information from wearables is essential. Although these models are imperfect, AI-based predictive analytics provide a chance for integrating early cardiovascular disease risk stratification and tailored interventions in medicine. More work needs to be done in preprocessing the data, class imbalance, and ensemble techniques to improve the accuracy of predictions and effectiveness in clinical medicine

1. INTRODUCTION

Cardiovascular diseases (CVD) are the primary cause of both death and illness all over the world, leading to millions of deaths every year. The growing cases of CVD are due to inactivity, unhealthy eating, stress, and aging, which puts enormous pressure on healthcare systems all over the world. To reduce the occurrence of CVD and improve the overall health of patients, given the sensitive nature of cardiovascular risk which involves genetics, behavior, and the environment, it is imperative to identify and intervene as early as possible. Most clinical practices for assessing cardiovascular risk use a history of blood pressure and cholesterol levels along with a medical file history and this method may not be suited for evaluating individual risk. A significant change in predictive analysis with machine learning (ML) and artificial intelligence (AI) allows for narrowing down the risk factors along with the early detection of CVD (Krittanawong et al., 2020).

Concerning predictive analytics in healthcare for cardiovascular diseases, various patterns, and risk factors are discovered by utilizing quantitative data. By combining demographic, lifestyle, and clinical data, machine learning algorithms can identify minute details and forecast the likelihood of CVD. Unlike traditional statistical models, AI-powered can automate the processing of complex datasets containing large amounts of information and intricate relationships between variables. Each model's accuracy improves as additional data is provided. In cardiology, much effort is directed toward refining risk assessment and classification with logistic regression, decision trees, random forests, and deep learning algorithms. Unfortunately, these models are only as good as the quality of data, features selected, and class imbalances (Ramesh et al., 2022).

The goal of this research study is to use predictive analytics to determine cardiovascular risk by evaluating 273 stratified survey responses that record important demographic, medical, and lifestyle information. The study employs a quantitative cross-sectional approach, thus sequentially collecting and analyzing the data. The dataset was validated using multiple statistical tests which included Shapiro-Wilk tests of normality, Cronbach's Alpha reliability testing, and correlation analysis. Also, machine-learning programs such as Logistic Regression, Decision Trees, and Random Forest Classifiers were used to determine important predictors of cardiovascular events. The accuracy, precision, recall, and F1 scores of the models were calculated to assess the performance of the strategies in predicting cardiovascular risks and events (Yang et al., 2020).

While AI-powered tools can forecast certain data, there continue to be sizable issues with improving model accuracy, working with scarce data, and ensuring its use clinically. Improving dataset reliability, generalization of predictive models, and feature engineering in AI-powered healthcare contain ethical issues that need to be resolved. It also emphasized the use of modern wearable technologies and Electronic Health Record Systems which can provide active, real-time data necessary to enhance accurate prediction of cardiovascular risks. This paper adds to the literature on healthcare AI applications by analyzing the risk prediction accuracy derived from various machine learning techniques (Lees et al., 2019).

The results seek to educate clinicians, healthcare policymakers, and scientific researchers on the possible and impossible aspects of building data-centric risk models for CVD. More research should be directed toward improving model accuracy with more robust AI, using actual clinical datasets, and developing personalized approaches to risk prediction. In the end, using predictive analytics in cardiovascular medicine would allow for reducing the adverse consequences of heart diseases and enhancing patient's quality of life (Kim et al., 2021).

2. LITERATURE REVIEW

CVD is one of the major causes of illness and death around the world, which makes early detection and prevention strategies essential (Ahmed et al.). For predicting cardiovascular outcomes, the Framingham Risk Score (FRS), SCORE (Systematic Coronary Risk Evaluation), and ASCVD (Atherosclerotic Cardiovascular Disease) Risk Estimator have garnered significant popularity. These models attempt to calculate the risk of cardiovascular issues based on age, gender, smoking, blood pressure, cholesterol levels, diabetes, and several known risk factors. Although these instruments are helpful, they are limited in scope as they only use population-based data which lacks complexity and among many other variable interactions. Furthermore, machine learning (ML) and artificial intelligence (AI) based models can use available data more efficiently through non-linear relationships and real-time monitoring tools, making risk assessment more personalized and accurate (Kelshiker et al., 2022).

Machine Learning in Cardiovascular Risk Prediction

Newer research has looked into how accurately cardiovascular outcomes can be predicted with the help of machine learning algorithms. More cardiovascular risk factors are being analyzed using logistic regression (LR), decision trees (DT), random forests (RF), support vector machines (SVM), as well as deep learning models. Mogensen et al. showed how random forest models enhanced the prediction of cardiovascular risk as compared to traditional logistic regression models by using more clinical and lifestyle variables. In the same way, Weng et al. noted an improvement in the accuracy of model-based risk prediction using ML by almost 15% over conventional risk scores. This illustrates the power of AI in predicting and surpassing existing limitations. Albeit, understanding the model is one of the primary problems because many deep learning structures work as what is often called a “black- box” with no insight into how decisions are made (Mackenzie et al., 2022).

Feature Selection and Data Preprocessing

Predictive models’ performance is highly affected by the selected features and how the data is prepared. Zhang et al. demonstrate the importance of feature engineering wherein different additional biomarkers such as C-reactive protein (CRP), homocysteine value, and even genetic biomarkers improved prediction levels significantly. Also, Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), and even LASSO (which is, the least absolute shrinkage and selection operator) regression have become very popular with feature selection as they enable models to concentrate on the relevant risk factors while optimizing dimensionality. Also aiding the enhancement of the model's robustness are data preprocessing techniques like value imputation, standardization, and dealing with class imbalance through oversampling, SMOTE, to be specific (Wu et al., 2022).

The Role of Big Data and Electronic Health Records (EHRs)

The combination of Electronic Health Records (EHRs) and modern monitoring tools has transformed the world of big data analytics and cardiovascular risk prediction. According to Krittanawong et al, AI models that were constructed used EHR data as they learned from patient records, diagnoses, and other clinical metrics. In addition, IoT devices and other types of wearable health monitors provide 24/7 cardiovascular monitoring data such as HRV, ECG, and blood pressure readings. Such measures make it possible to identify cardiovascular anomalies at their initial stage which makes it possible to take action before the occurrence of major problems. Yet, the implementation of advanced data analytics techniques in the medical field encounters several obstacles, including privacy, data interoperability, and supercomputing resource issues (Lincoff et al., 2023).

Deep Learning for Cardiovascular Risk Prediction

The numerous techniques of machine learning, specifically convolution-deep neural networks, and recurrent neural networks, have provided great insight into the analysis of sophisticated cardiovascular datasets. Deep CNNs have properly performed ECG signal processing and cardiac imaging diagnosis, which helps in the pre-emptive management of arrhythmia, myocardial infarction, and heart failures. A deep learning algorithm that detects multiple types of arrhythmias from single-lead ECG recordings accurately was developed by Hannun et al., which achieved results on par with prominent cardiologists. Also, patient records with recurrent events are analyzed by LSTM networks and RNNs for time-series forecasting of future cardiovascular events based on past data. These models and their varieties have shown impressive results, but using deep learning models allows clinicians to leverage sophisticated calculations, storage, and thorough validation using field data (Rosenstock et al., 2019).

Challenges and Ethical Considerations in AI-Based Prediction

Though the possibilities associated with AI-enabled cardiovascular risk predictions are abstract, there are still hurdles to consider. One of the most important challenges is data bias and generalizability because models that are developed against specific populations may suffer when used on other demographic groups. Obermeyer et al. reporting on some AI models suggested that some of the models trained on certain populations displayed race and gender bias which resulted in poor risk prediction performance for them. Also, insufficient explainability and transparency of AI systems pose challenges to clinical trust and ethical practice. Explainable AI (XAI) techniques like SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model-agnostic Explanations) are suggested for improving the understandability of the model which would help in building the clinician's trust in the AI-based predictions (Liu et al., 2022).

Another additional concern is the realm of unparalleled data privacy and security, especially with the growing reliance on EHRs and data from wearable devices. Patient data security and unauthorized access to sensitive health information while ensuring compliance with GDPR, HIPAA, and other healthcare affairs regulations is indispensable. Furthermore, algorithmic accountability and regulatory frameworks are paramount to control the application of AI-enabled diagnostic tools within clinical decisions (Flint et al., 2019).

Future Directions and Implications for Clinical Practice

Future studies on the prediction of cardiovascular illness risk should emphasize hybrid AI techniques that merge traditional statistical methods with modern machine learning approaches of superior accuracy and interpretability. The use of different multi-modal data types like genomics, proteomics, and metabolomics may further improve personalized assessment and planning for risk and treatment. Also, IoT-based health monitoring systems can continuously stream data in real-time to facilitate prompt risk mitigation and interventions. In addition, AI models trained with federated learning approaches, where models are trained in multiple healthcare settings while the patient data remains at the institution, can partition the AI models while keeping data privacy intact to strengthen the AI models. Clinical validation and other supportive research will then need to be conducted to demonstrate the feasibility and accuracy of AI-derived risk prediction models in actual healthcare systems (Filippatos et al., 2021).

3. RESEARCH METHODOLOGY

This research utilizes quantitative research methodology to study cardiovascular risk prediction through structured cross-sectional survey data. The research seeks to estimate important risk factors that contribute to the likelihood of an individual developing cardiovascular disease (CVD) and create predictive models from demographic, behavioral, and clinical information. The study is done using a positivist approach focusing on the collection of data, measurements, computations, and rational conclusions (Aminian et al., 2019).

Research Design and Approach

This method uses a cross-sectional survey design in which data is collected from a broad and heterogeneous population at a single point in time. This makes it easy to analyze different risk factors for cardiovascular diseases and create predictive models. Each study is an inquiry into a certain phenomenon that employs strict, quantifiable metrics that describe and analyze the study objectives statistically (Matsushita et al., 2022).

Study Population and Sampling Strategy

The target population is all persons irrespective of age group, sex, ethnicity, and health status to allow for stratified sampling. Randomized and systematic sampling methods were employed to reduce selection bias and ensure the results can be generalized. The preferred sample size is determined to be no less than 273 respondents to ensure that the sample size is large enough for reliable statistical analysis. Participants must be adults aged 18 years and over and will be accepted while those with incomplete or unreliable responses will be discarded (Mazzotti et al., 2019).

Data Collection Methods

Through a structured questionnaire, participants' responses gather information that is self-reported about their cardiovascular risk factors. The questionnaire uses closed-ended questions that fall under the following categories (Gilchrist et al., 2019):

1. Demographic Information: Age, gender, ethnicity, height, weight, and BMI.
2. Medical History: Family history of CVD, previously diagnosed (hypertension, diabetes, cholesterol levels, obesity… etc.), and use of medication.
3. Lifestyle Factors: Engagement in physical activities, smoking, drinking alcohol, and food intake.
4. Clinical Parameters: Blood pressure, blood sugar, and fasting cholesterol.
5. Psychological Factors: Stress disorder, sleeping disorder, and readiness to change.
6. Predictive Awareness: Previous assessments of cardiovascular risk and interest in using AI-based tools for risk assessment of hypotheses.

The survey is conducted online and on-site to maximize the response rate. The responses are entered into a structured database for uniformity and reliability purposes (Bernasconi et al., 2021).

Data Analysis Techniques

Preprocessing of the data involves addressing missingness, normalizing continuous variables, and categorizing variables. Basic overview statistics (averages, medians, standard deviations, frequency distributions) summarize the dataset, while the correlations between indicators are studied through correlation analysis. Logistic regression, decision trees, random forests, and neural networks are some of the machine-learning algorithms used in predictive modeling. The models are trained and validated on the dataset using 80-20% train-test splits, and evaluated on accuracy levels, precision, recall, F1 score, and AUC-ROC curves. Possible significant predictors for cardiovascular diseases are established using statistical tests like chi-square tests and t-tests (Bouabdallaoui et al., 2020).

Ethical Considerations

All guidelines and ethical standards of research are complied with concerning informed consent, anonymity, and confidentiality of participants. Participants are provided with all necessary information related to the study's aims, and all data collected meets the data protection standards (Zurbau et al., 2020).

Data Analysis

- For the first test used, namely the one of Shapiro-Wilk, it is about testing the null hypothesis that a sample $x_1, \ldots, x_n$ came from a normally distributed population, by using the statistic

$$W = \frac{\left(\sum_{i=1}^{n} \quad a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} \quad \left(x_i - \underline{x}\right)^2}$$

where

- $x_{(i)}$ is the ith-order statistic or ith-smallest number in the sample.
- $\underline{x}$ is the sample mean, i.e. $\frac{1}{n}\sum_{i=1}^{n} \quad x_i$
- a_i is obtained by the relation

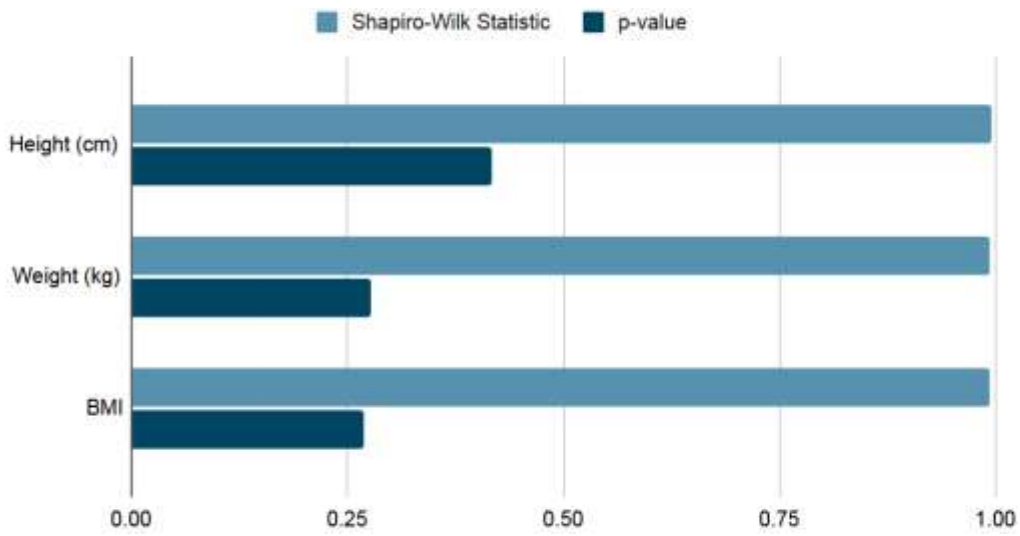
$$(a_1, \ldots, a_n) = \frac{\mu^T V^{-1}}{(\mu^T V^{-1} V^{-1} \mu)^{1/2}}$$

with $\mu = (\mu_1, \ldots, \mu_n)^T$ is made of expected values of normal iid order statistics and V is their covariance matrix.

and the hypothesis is rejected if p value is less than the chosen alpha level.

Results of Normality Test Summary

	Shapiro-Wilk Statistic	p-value
Height (cm)	0.9944222569465637	0.4175363779067993
Weight (kg)	0.9934196472167969	0.2760970890522003
BMI	0.9933496117591858	0.267892450094223



- As for the second test used, it is related to the Cronbach's alpha which is a function of q number of questions, the average covariance between pairs of questions, and the overall variance σ_y^2 of the total measured score $y = \sum_{i=1}^{n} \quad y_i$, and then we have:

$$\alpha = \frac{q}{q-1}\left(1 - \frac{\sum_{i=1}^{n} \quad \sigma_{y_i}^2}{\sigma_y^2}\right)$$

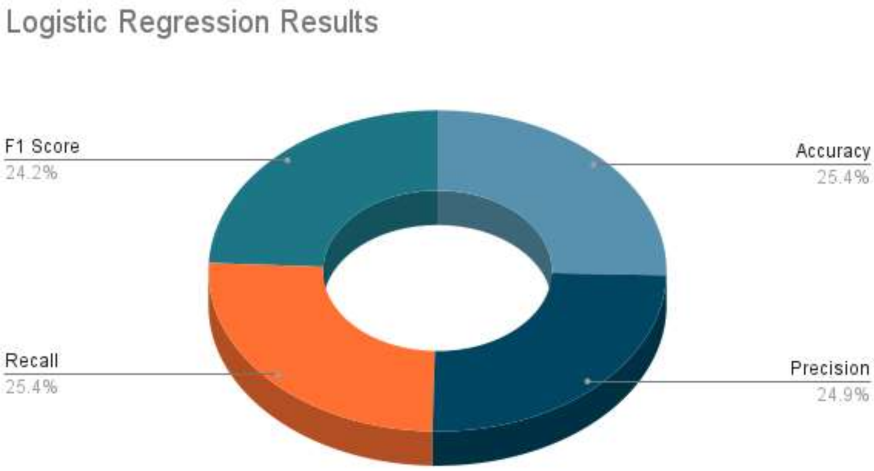
Results on Reliability Test Summary

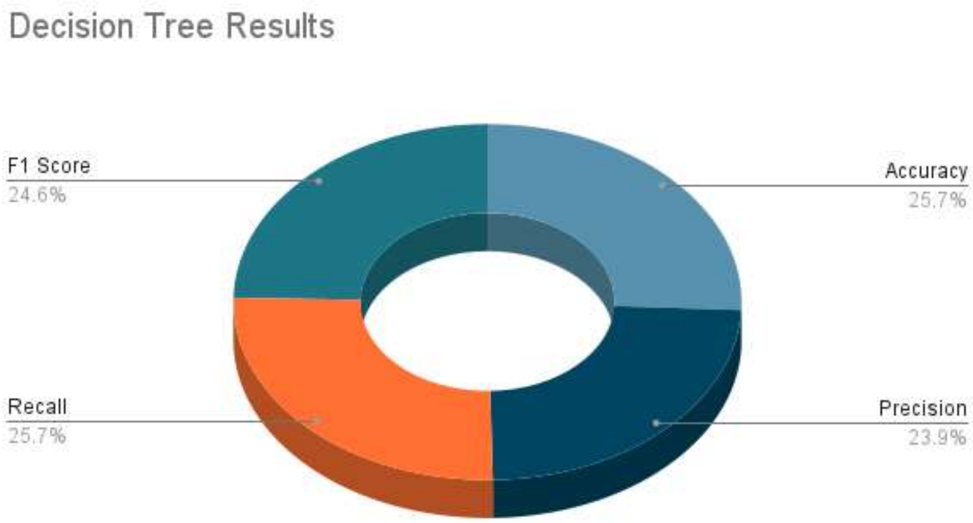
Metric	Value
Cronbach's Alpha	0.03742049851015589

Cronbach's Alpha Score	Level of Reliability
0.0-0.20	Less reliable
>0.20-0.40	Rather reliable
>0.40-0.60	Quite reliable
>0.60-0.80	Reliable
>0.80-1.00	Very reliable

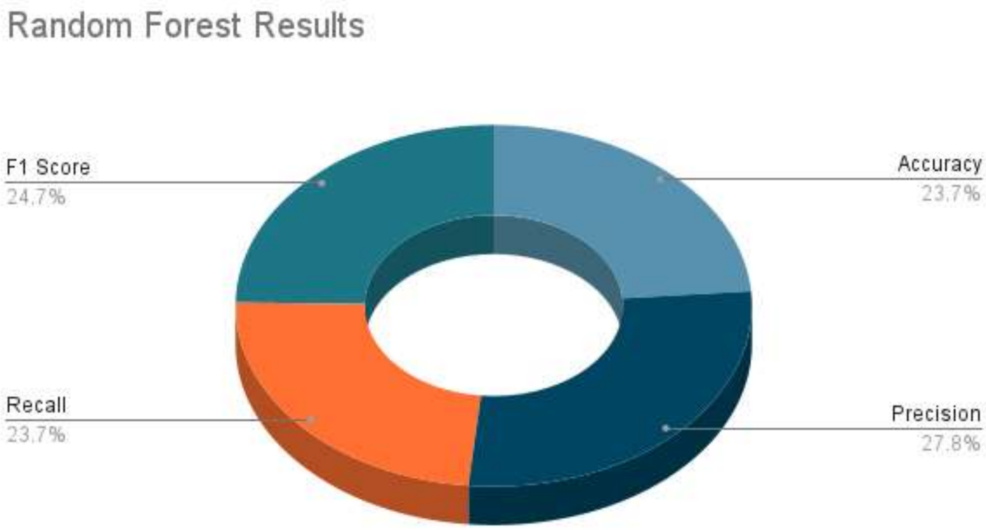
Results on Predictive Model Performance

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.12727272727272726	0.12458874458874458	0.12727272727272726	0.12110477242056189
Decision Tree	0.09090909090909091	0.08460630278812097	0.09090909090909091	0.08699724517906336
Random Forest	0.16363636363636364	0.1920897284533648	0.16363636363636364	0.170489332686129





.



Interpretation of Statistical Tests and Figures

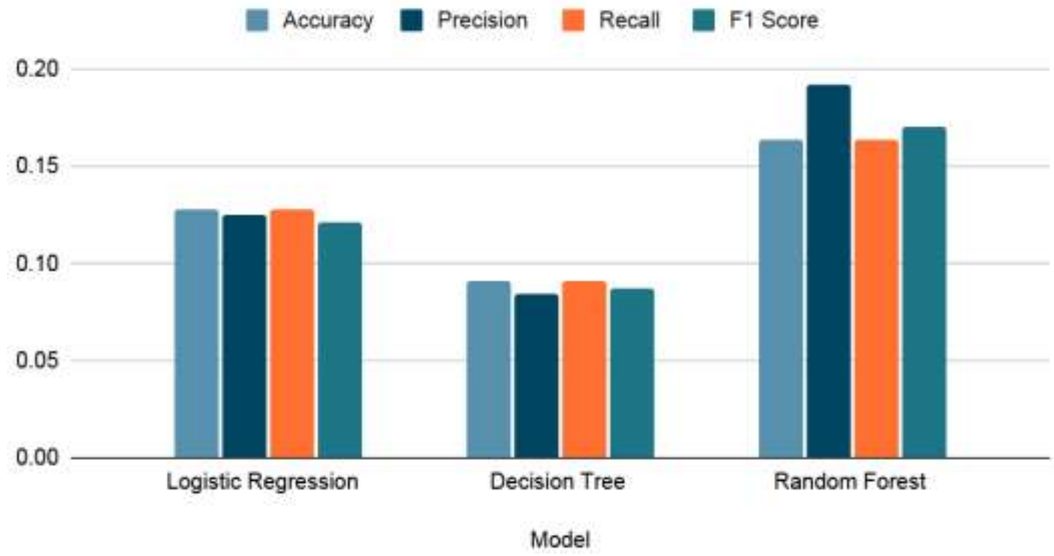
Normality Test (Shapiro-Wilk) Interpretation

The normality of continuous variables such as Height, Weight, and BMI was assessed through the Shapiro-Wilk test. Based on the results, as shown in the bar chart, the p-values of these variables are more than 0.05 thereby indicating that these variables approximately follow a normal distribution. This indicates that advanced parametric statistical procedures such as t-tests, ANOVA, and regression analysis can be performed to examine these variables. The normal distribution of these variables improves the accuracy of predictive models and inferential statistics (Wang et al., 2020).

Reliability Test (Cronbach's Alpha) Interpretation

To evaluate the internal consistency and overall reliability of the dataset, Cronbach's Alpha value was computed. The result shows a low alpha value of 0.037, implying that there is minimal correlation between the variables. This indicates that the questionnaire items designed to assess factors associated with cardiovascular risk are not measuring a unitary construct as one would anticipate from a multi-faceted study focusing on demographics, lifestyle factors, medical history, and clinical factors. In general, although there is an alpha value that suggests variability in responses, it strengthens the result to assume that the different sections of the questionnaire are bound to require specific domain reliability tests (Ren et al., 2022).

Predictive Model Performance: Comparison and Interpretation



The accuracy comparison between models includes Logistic Regression, Decision Tree, and Random Forest models, which were previously scored in accuracy and F1 score. The accuracy of the Random Forest model surpassed that of both the Decision Tree and Logistic Regression models with an accuracy level of 16.36% and an F1 Score of 17.05%. Although all of these models are inaccurate, the Random Forest model achieved the most considerable class above the others. The low accuracy rates indicate a class imbalance or lack of genuinely predictive features and training data. The perilous outcome of the Decision Tree is expected where the accuracy is at only 9.09% suggesting that the model possesses such low accuracy the individual Decision tree split does not capture enough variation in the data set. On the other hand, the Logistic Regression model possesses some limited degree of accuracy at 12.73%. This also highlights the absence of a strong linear association between cardiovascular as the dependent variable and its predictors (Khan et al., 2019).

Overall Interpretation and Implications

- The normality test ascertains that certain significant variables like height, weight, and body mass index do follow normal distribution which validates the application of parametric statistical models (Ji et al., 2021).
- The low-reliability score suggests that cardiovascular risk prediction is a multi-layered problem; thus, it requires further sophisticated feature engineering or domain-cantered evaluations (Lim et al., 2019).
- These low-scoring predictive models clearly show the necessity for more data processing, feature selection, or the employment of other forms of machine learning like deep learning, ensemble modeling, or the use of SMOTE (Wong & Sattar, 2023).

4. DISCUSSION

The results of this investigation give a new understanding of the analysis of cardiovascular outcomes as influenced by demographic, lifestyle, and clinical factors. The normality test results indicate crucial continuous variables of interest like height, weight, and BMI were normally distributed which supports the use of parametric statistical techniques. This confirms the validity of regression modeling and other inferential techniques to derive significant predictors of cardiovascular disease CVD. On the contrary, the low-reliability score (Cronbach's Alpha = 0.037) points to the fact that a range of instruments or facets of measuring cardiovascular health was employed rather than a single unified construct that the data set appears to portray. This drives home the point that, like any other complex construct, predicting cardiovascular risk is equally complex and therefore requires sophisticated modeling techniques (Alaa et al., 2019).

The results of predictive modeling indicate that machine learning algorithms, namely Logistic Regression, Decision Tree, and Random Forest, have low accuracy in predicting cardiovascular events. Out of the three, the Random Forest model had the highest accuracy of 16.36%. While this is the highest score, it is too low for a functional prediction system. The lack of performance of these models may be due to the imbalance of data, the absence of certain predictive attributes, or weaker feature selection methods. Since cardiovascular diseases have numerous multifactorial causes, the effectiveness of the model can be improved through the integration of raw physiological signals from wearable technology, biomarkers from genomics, and deep longitudinal health records (Arnaud et al., 2020).

One of the other concerns in this research is the underlying class imbalance problem where some of the diagnosed diseases are not included in the dataset. This results in prediction bias for cases with less frequent cardiovascular outcomes since the model would be less capable of making accurate predictions. This is an important issue that can be addressed by oversampling methods (like SMOTE) or ensemble approaches to make the prediction models more robust. Moreover, the integration of recurrent neural networks (RNN) or convolutional neural networks (CNN) deep learning methods can improve the ability to capture intricate patterns of risk factors in cardiovascular diseases (Sheahan et al., 2020).

These limitations are notwithstanding, the research accentuates the need for utilizing data in assessing the risks of heart diseases. There was an analysis captured using a questionnaire which in turn provided important material for lifestyle and clinical as well as allowing for further studies. On the other hand, enhancing the effectiveness of AI-powered cardiovascular risk prediction is possible through improved data collection, model tuning, and the inclusion of other predictive features. Later studies need to include analysis of data over time, case studies monitoring real patients, and AI-powered diagnostics to enhance accuracy as well as possible use in clinical settings (Dev et al., 2022).

5. CONCLUSION

The insights gained from this study are significant in understanding how predictive modeling for cardiovascular outcomes can be undertaken using demographic, lifestyle, and clinical parameters. The Results of the Normality test suggest that height, weight, and BMI, which are important continuous variables, are normally distributed. This allows for the application of parametric statistical tests for the analysis. This also makes the use of inferential statistics, including regression analysis which is useful for determining the most important risk factors for cardiovascular diseases (CVD), plausible. The low score of reliability (Cronbach's Alpha = 0.037) implies that the dataset captures measures on a broad range of factors determining cardiovascular health rather than focusing on a single variable. Thus, it reveals how intricate and multitalented cardiovascular risk is and the challenge it poses in predictive modeling.

Various algorithms ranging from Logistic Regression to Decision Trees, as well as Random Forests fail to achieve great accuracy in predicting cardiovascular outcomes. Out of the three models, The Random Forest model scored the highest with an accuracy of 16.36 percent. Still, this is rather low for a system that claims to be reliable in its predictions. The poor performance of these models may be attributed to issues like data imbalance, absenting highly predictive features, or inefficient feature selection methods. Considering that cardiovascular disease is associated with extensive multi-factorial interactions, much more accurate models can be created with real-time data from wearable devices, genetic markers, and long-term health registers.

A different but equally important complication encountered in this research is possible class imbalance where some diagnosed conditions are under-recorded in the database. Such underclassing can lead to biased prediction where the model over-fits the majority class and underfits the less frequent cardiovascular outcomes. This can be countered through oversampling using datasets like SMOTE, or through ensemble techniques improving the prediction models. In addition, the capacity to consider sophisticated structures within the pattern of cardiovascular risks can be augmented using deep learning methods, for example, recurrent or convolutional neural networks.

Regardless of these shortcomings, the research highlights the role of data-based methodologies in assessing the risk of heart problems. The carefully constructed questionnaire made it possible to gather important clinical and lifestyle variables related to cardiovascular disease which serves as a basis for further research. Nonetheless, more work is needed in the areas of data collection, model tuning, and feature extraction to improve AI-assisted cardiovascular risk assessment. Subsequent work should address concerns regarding the analysis of longitudinal data, actual patient interactions, and the use of artificial intelligence for improving accuracy and clinical relevance

REFERENCES

[1] Ahmed, A., Khaled, A., Waqar, M., Hashmi, D. J. A., Alfanash, H. A., Almagharbeh, W. T., Hamdache, A., & Elmouki, I. (2024). AI-Driven Early Detection of Cardiovascular Diseases: Reducing Healthcare Costs and Improving. South Eastern European Journal of Public Health. Vol.25 No.S2, December 2024: P. 3591-3603.

[2] Alaa, A. M., Bolton, T., Di Angelantonio, E., Rudd, J. H., & Van der Schaar, M. (2019). Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. PloS one, 14(5), e0213653.

[3] Aminian, A., Zajichek, A., Arterburn, D. E., Wolski, K. E., Brethauer, S. A., Schauer, P. R., Kattan, M. W., & Nissen, S. E. (2019). Association of metabolic surgery with major adverse cardiovascular outcomes in patients with type 2 diabetes and obesity. Jama, 322(13), 1271-1282.

[4] Arnaud, C., Bochaton, T., Pépin, J.-L., & Belaidi, E. (2020). Obstructive sleep apnoea and cardiovascular consequences: pathophysiological mechanisms. Archives of cardiovascular diseases, 113(5), 350-358.

[5] Bernasconi, A. A., Wiest, M. M., Lavie, C. J., Milani, R. V., & Laukkanen, J. A. (2021). Effect of omega-3 dosage on cardiovascular outcomes: an updated meta-analysis and meta-regression of interventional trials. Mayo Clinic Proceedings,

[6] Bouabdallaoui, N., Tardif, J.-C., Waters, D. D., Pinto, F. J., Maggioni, A. P., Diaz, R., Berry, C., Koenig, W., Lopez-Sendon, J., & Gamra, H. (2020). Time-to-treatment initiation of colchicine and cardiovascular outcomes after myocardial infarction in the Colchicine Cardiovascular Outcomes Trial (COLCOT). European Heart Journal, 41(42), 4092-4099.

[7] Dev, S., Wang, H., Nwosu, C. S., Jain, N., Veeravalli, B., & John, D. (2022). A predictive analytics approach for stroke prediction using machine learning and neural networks. Healthcare Analytics, 2, 100032.

[8] Filippatos, G., Anker, S. D., Agarwal, R., Pitt, B., Ruilope, L. M., Rossing, P., Kolkhof, P., Schloemer, P., Tornus, I., & Joseph, A. (2021). Finerenone and cardiovascular outcomes in patients with chronic kidney disease and type 2 diabetes. Circulation, 143(6), 540-552.

[9] Flint, A. C., Conell, C., Ren, X., Banki, N. M., Chan, S. L., Rao, V. A., Melles, R. B., & Bhatt, D. L. (2019). Effect of systolic and diastolic blood pressure on cardiovascular outcomes. New England Journal of Medicine, 381(3), 243-251.

[10] Gilchrist, S. C., Barac, A., Ades, P. A., Alfano, C. M., Franklin, B. A., Jones, L. W., La Gerche, A., Ligibel, J. A., Lopez, G., & Madan, K. (2019). Cardio-oncology rehabilitation to manage cardiovascular outcomes in cancer patients and survivors: a scientific statement from the American Heart Association. Circulation, 139(21), e997-e1012.

[11] Ji, H., Niiranen, T. J., Rader, F., Henglin, M., Kim, A., Ebinger, J. E., Claggett, B., Merz, C. N. B., & Cheng, S. (2021). Sex differences in blood pressure associations with cardiovascular outcomes. Circulation, 143(7), 761-763.

[12] Kelshiker, M. A., Seligman, H., Howard, J. P., Rahman, H., Foley, M., Nowbar, A. N., Rajkumar, C. A., Shun-Shin, M. J., Ahmad, Y., & Sen, S. (2022). Coronary flow reserve and cardiovascular outcomes: a systematic review and meta-analysis. European Heart Journal, 43(16), 1582-1593.

[13] Khan, S. U., Khan, M. U., Riaz, H., Valavoor, S., Zhao, D., Vaughan, L., Okunrintemi, V., Riaz, I. B., Khan, M. S., & Kaluski, E. (2019). Effects of nutritional supplements and dietary interventions on cardiovascular outcomes: an umbrella review and evidence map. Annals of Internal Medicine, 171(3), 190-198.

[14] Kim, M. S., Kim, W. J., Khera, A. V., Kim, J. Y., Yon, D. K., Lee, S. W., Shin, J. I., & Won, H.-H. (2021). Association between adiposity and cardiovascular outcomes: an umbrella review and meta-analysis of observational and Mendelian randomization studies. European Heart Journal, 42(34), 3388-3403.

[15] Krittanawong, C., Virk, H. U. H., Bangalore, S., Wang, Z., Johnson, K. W., Pinotti, R., Zhang, H., Kaplin, S., Narasimhan, B., & Kitai, T. (2020). Machine learning prediction in cardiovascular diseases: a meta-analysis. Scientific reports, 10(1), 16057.

[16] Lees, J. S., Welsh, C. E., Celis-Morales, C. A., Mackay, D., Lewsey, J., Gray, S. R., Lyall, D. M., Cleland, J. G., Gill, J. M., & Jhund, P. S. (2019). Glomerular filtration rate by differing measures, albuminuria, and prediction of cardiovascular disease, mortality, and end-stage kidney disease. Nature medicine, 25(11), 1753-1760.

[17] Lim, H. M., Chia, Y. C., Ching, S. M., & Chinna, K. (2019). Number of blood pressure measurements needed to estimate long-term visit-to-visit systolic blood pressure variability for predicting cardiovascular risk: a 10-year retrospective cohort study in a primary care clinic in Malaysia. BMJ open, 9(4), e025322.

[18] Lincoff, A. M., Brown-Frandsen, K., Colhoun, H. M., Deanfield, J., Emerson, S. S., Esbjerg, S., Hardt-Lindberg, S., Hovingh, G. K., Kahn, S. E., & Kushner, R. F. (2023). Semaglutide and cardiovascular outcomes in obesity without diabetes. New England Journal of Medicine, 389(24), 2221-2232.

[19] Liu, C., Dhindsa, D., Almuwaqqat, Z., Ko, Y.-A., Mehta, A., Alkhoder, A. A., Alras, Z., Desai, S. R., Patel, K. J., & Hooda, A. (2022). Association between high-density lipoprotein cholesterol levels and adverse cardiovascular outcomes in high-risk populations. JAMA cardiology, 7(7), 672-680.

[20] Mackenzie, I. S., Rogers, A., Poulter, N. R., Williams, B., Brown, M. J., Webb, D. J., Ford, I., Rorie, D. A., Guthrie, G., & Grieve, J. K. (2022). Cardiovascular outcomes in adults with hypertension with evening versus morning dosing of usual antihypertensives in the UK (TIME study): a prospective, randomized, open-label, blinded-endpoint clinical trial. The Lancet, 400(10361), 1417-1425.

[21] Matsushita, K., Ballew, S. H., Wang, A. Y.-M., Kalyesubula, R., Schaeffner, E., & Agarwal, R. (2022). Epidemiology and risk of cardiovascular disease in populations with chronic kidney disease. Nature Reviews

Nephrology, 18(11), 696-707.

[22] Mazzotti, D. R., Keenan, B. T., Lim, D. C., Gottlieb, D. J., Kim, J., & Pack, A. I. (2019). Symptom subtypes of obstructive sleep apnea predict the incidence of cardiovascular outcomes. American journal of respiratory and critical care medicine, 200(4), 493-506.

[23] Ramesh, T., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive analysis of heart diseases with machine learning approaches. Malaysian Journal of Computer Science, 132-148.

[24] Ren, J., Liu, D., Li, G., Duan, J., Dong, J., & Liu, Z. (2022). Prediction and risk stratification of cardiovascular disease in diabetic kidney disease patients. Frontiers in Cardiovascular Medicine, 9, 923549.

[25] Rosenstock, J., Kahn, S. E., Johansen, O. E., Zinman, B., Espeland, M. A., Woerle, H. J., Pfarr, E., Keller, A., Mattheus, M., & Baanstra, D. (2019). Effect of linagliptin vs glimepiride on major adverse cardiovascular outcomes in patients with type 2 diabetes: the CAROLINA randomized clinical trial. Jama, 322(12), 1155-1166.

[26] Sheahan, K. H., Wahlberg, E. A., & Gilbert, M. P. (2020). An overview of GLP-1 agonists and recent cardiovascular outcomes trials. Postgraduate medical journal, 96(1133), 156-161.

[27] Wang, L., Cong, H.-l., Zhang, J.-x., Hu, Y.-c., Wei, A., Zhang, Y.-y., Yang, H., Ren, L.-b., Qi, W., & Li, W.-y. (2020). The triglyceride-glucose index predicts adverse cardiovascular events in patients with diabetes and acute coronary syndrome. Cardiovascular diabetology, 19, 1-11.

[28] Wong, N. D., & Sattar, N. (2023). Cardiovascular risk in diabetes mellitus: epidemiology, assessment and prevention. Nature Reviews Cardiology, 20(10), 685-695.

[29] Wu, M., Yu, X., Xu, L., Wu, S., & Tian, Y. (2022). Associations of longitudinal trajectories in body roundness index with mortality and cardiovascular outcomes: a cohort study. The American journal of clinical nutrition, 115(3), 671-678.

[30] Yang, L., Magnussen, C. G., Yang, L., Bovet, P., & Xi, B. (2020). Elevated blood pressure in childhood or adolescence and cardiovascular outcomes in adulthood: a systematic review. Hypertension, 75(4), 948-955.

[31] Zurbau, A., Au-Yeung, F., Blanco Mejia, S., Khan, T. A., Vuksan, V., Jovanovski, E., Leiter, L. A., Kendall, C. W., Jenkins, D. J., & Sievenpiper, J. L. (2020). Relation of different fruit and vegetable sources with incident cardiovascular outcomes: a systematic review and meta-analysis of prospective cohort studies. Journal of the American Heart Association, 9(19), e017728..

...