Original Article
Deep learning models for predicting heart disease risk using the UCI database: methods, performance, and clinical context

Reza Khademi[1*], Golnaz Yazdanpanah[2*], Ghazaleh Rouhparvarzamin[3*], Vida Hafezi[4], Shayesteh Haghighi[5], Parham Panahi[6], Aida Bakhshi[7,8], Bita Faridnia[9], Ramin Ahangar-Sirous[10], Yasaman Tavakoli[11], Mohammad Amin Karimi[12], Farzad Sheikhzadeh[13], Ata Akhtari Kohnehshahri[14], Amir Abdi[15], Shekoufeh Safarbeiranvand[16], Mahsa Asadi Anar[17], Parisa Alsadat Dadkhah[18]

[1]*Student Research Committee, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran; [2]School of Medicine, Shiraz University of Medical Sciences, Shiraz, Iran; [3]Student Research Committee, School of Nursing and Midwifery, Shahid Sadoughi University of Medical Sciences, Yazd, Iran; [4]School of Medicine, Alborz University of Medical Sciences, Karaj, Iran; [5]Department of Nursing, School of Nursing and Midwifery, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran; [6]Kazan Federal University, Kazan, Russian Federation; [7]Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran; [8]Clinical Research Development Unit, Ghaem Hospital, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran; [9]Department of Internal Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran; [10]Student Research Committee, Tabriz University of Medical Sciences, Tabriz, Iran; [11]Medical Student, Department of Medicine, Mazandaran University of Science, Sari, Iran; [12]School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran; [13]School of Medicine, Iran University of Medical Sciences, Tehran, Iran; [14]Student Research Committee, Faculty of Medicine, Tabriz Medical Sciences, Islamic Azad University, Tabriz, Iran; [15]Student Research Committee, School of Medicine, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran; [16]Student Research Committee, School of Medicine, Islamic Azad University, Kerman, Iran; [17]College of Medicine, University of Arizona, Tucson, AZ, USA; [18]Student Research Committee, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. *Equal contributors.

Abstract: Objectives: To develop and evaluate deep learning models for predicting heart disease using the University of California, Irvine (UCI) heart disease dataset, and to contextualize model performance against classical machine learning approaches. Method: Data were extracted from the University of California Irvine (UCI) heart disease dataset, including information from Cleveland, Hungary, Switzerland, and Long Beach V, collected in 1988. The dataset comprises 1,025 patients and 14 key attributes. Deep learning models were used to analyze the data and predict heart disease risk. Results: The deep learning models demonstrated high accuracy in predicting heart disease risk. The Random Forest model achieved an accuracy of 99%. Significant predictors included exercise-induced angina and downsloping ST segments. The data revealed that 72% of females and 42% of males experienced heart attacks. There was a 79% chance that atypical angina and a 77% chance that non-anginal pain would lead to a heart attack. Exercise-induced angina had a 67% chance of resulting in a heart attack, while downsloping of the peak exercise ST segment had a 72% chance. Additionally, a 71% chance was observed for heart attacks in patients with no major coronary artery blockage (ca=0), and a 75% chance for those with a potentially reversible thalassemia-related defect (thal=2). Age groups 40-44 and 50-54 had a 76% and 61% risk of heart attacks, respectively. Conclusion: Deep learning models can significantly enhance heart disease risk prediction, leading to improved treatment strategies. These findings can aid in early diagnosis and timely interventions, improving clinical outcomes for heart disease patients.

Keywords: Heart disease, deep learning, machine learning, risk prediction, cardiovascular risk factors, random forest

Introduction

Heart disease, one of the most common chronic diseases worldwide, is the leading cause of death across all genders and racial groups and affects millions of people in the United States alone, where approximately 695,000 people lost their lives to heart disease in 2021, according to the American Heart Association [1]. In 2020, an estimated 19.05 million people died from cardiovascular diseases worldwide, an increase of 18.71% since 2010, while in the United States the direct costs associated with these diseases reached $103.5 billion in 1996-1997 increased to $251.4 billion in 2018-2019 [1]. Heart diseases include a range of conditions such as coronary artery disease, congenital heart disease, cardiac arrhythmias, dilated cardiomyopathy, myocardial infarction, heart failure, hypertrophic cardiomyopathy, mitral valve regurgitation, mitral valve prolapse and aortic stenosis, the early detection and intervention of which are essential to treat and prevent the serious consequences of these diseases [2]. The availability of comprehensive open source platforms for accessing patient records has greatly expanded the potential for integrating advanced computing technologies into medical diagnostics, enabling more accurate disease detection and intervention strategies to prevent diseases from becoming life-threatening. In this context, the transformative role of machine learning (ML) and artificial intelligence (AI) in the healthcare industry has become increasingly evident, with these technologies facilitating the development of innovative models capable of diagnosing diseases, classifying medical conditions, and predicting outcomes with remarkable precision. By leveraging ML, comprehensive genomic data analysis can be performed seamlessly, providing previously unattainable insights and paving the way for more personalized and effective medical interventions. Additionally, the ability to train models for pandemic prediction and deep medical record analysis provides unprecedented opportunities to refine predictive capabilities, optimize resource allocation, and improve patient care through data-driven insights [3-5].

Further, AI and machine learning (ML) have been increasingly used in clinical decision support in recent years. Applications for these range from automatic interpretation of medical imaging (echocardiography, CT, MRI) to ECG sig-nal analysis for arrhythmia detection, predictive analytics for clinical outcomes, and natural language processing (NLP) to extract information from the electronic health records. They enhance diagnostic accuracy, facilitate early detection, and inform clinical decisions through models of high-dimensional data that are not manageable using classical statistics. A number of researchers have investigated the application of machine learning techniques to the classification/prediction of heart disease. For instance, Melillo et al. [6] developed an automated classifier for detecting congestive heart failure, distinguishing high-risk from low-risk patients using the Classification and Regression Tree (CART) algorithm. This approach achieved a sensitivity of 93.3% and a specificity of 63.5%. To further enhance performance, Rahhal et al. [7] proposed an electrocardiogram (ECG)-based method that leverages deep neural networks to select and utilize optimal features for improved diagnostic accuracy. Similarly, Guidi et al. [8] introduced a clinical decision support system aimed at early detection of heart failure, comparing various machine learning and deep learning models, including support vector machines (SVM), random forests, and CART. Among these, random forests and CART achieved the highest classification accuracy of 87.6%. Zhang et al. [9] demonstrated the integration of natural language processing (NLP) with rule-based techniques to extract NYHA heart failure classes from unstructured clinical notes, achieving an accuracy of 93.37%. Furthermore, Parthiban and Srivatsa [10] utilized SVM techniques to predict heart disease in diabetic patients, achieving an impressive 94.60% accuracy by incorporating common clinical features such as blood sugar levels, patient age, and blood pressure data. These advancements highlight the growing potential of ML and deep learning models in transforming heart disease diagnosis and management. Heart failure has been the focus of more extensive research due to the complexity of its diagnostic process [11], and this has made Computer-Aided Decision Support Systems particularly valuable in this field, as demonstrated in a study by Krishnaiah et al., where data mining techniques were employed to significantly reduce the time required for accurate disease prediction [12].

Cardiology has a uniquely focused utility with DL. In contrast to classic ML models that rely

Table 1. Study variables

Variable	Type	Range/Description
Age	Numeric	Patient's age in years
Sex	Categorical	0: Female, 1: Male
Chest Pain Type	Categorical	0-3: Various pain classifications
Resting Blood Pressure	Numeric	Measured in mm Hg
Serum Cholesterol	Numeric	Triglyceride levels
Fasting Blood Sugar	Categorical	≤120 mg/dL or >120 mg/dL
Resting ECG Results	Categorical	Normal/Abnormal variations
Maximum Heart Rate	Numeric	Highest heart rate achieved
Exercise-Induced Angina	Categorical	Presence/Absence
ST Depression (Oldpeak)	Numeric	Exercise-induced
ST Segment Slope	Categorical	Upsloping/Flat/Downsloping
Major Vessel Count	Numeric	0-3 vessels
Thalassemia Status	Categorical	Defect variations

on manually engineered features, DL automatically learns intricate features from high-dimensional data, which can account for the non-linear association between CVD risk factors in many cases [13]. DL has demonstrated better results in tasks, such as automated image segmentation in echocardiography and cardiac MRI, arrhythmia detection by ECG signals, and predicting major adverse cardiac events [14]. Moreover, DL models can incorporate multi-modal data (clinical characteristics, images, and physiological signals), enabling a more comprehensive risk estimation and precision in cardiology [15]. For these reasons, numerous researchers have developed methods to assist in detecting heart disease and possible associated risk factors by considering various factors. Many of these approaches utilize ML techniques to overcome the limitations of traditional statistical analysis methods, which often struggle to capture prognostic information within large datasets that involve complex, multi-dimensional interactions [16-18]. In this study, we aim to use a deep learning model to accurately predict heart disease risk by identifying and analyzing the most critical cardiovascular risk factors using an open source dataset from the UCI Database.

Method

Data source and dataset characteristics

The data are based on the broad heart disease dataset from the University of California Irvine (UCI) Machine Learning Repository [19]. The medical records selected from four different databases include Cleveland, Hungary, Switzerland, and Long Beach V; this is a compilation based on data collected in 1988. In this refined dataset, there are 14 key features among 76 attributes of the original dataset that have been proved necessary for the prediction of heart diseases.

Inclusion and exclusion criteria

Inclusion criteria: All adult patient records with the 14 most used clinical attributes of the UCI heart disease dataset (age, sex, chest pain type, resting blood pressure, serum cholesterol, fasting blood sugar, resting ECG, maximum heart rate, exercise induced angina, ST depression, ST slope, number of major vessels, thalassemia, and target output).

Exclusion criteria: Records with any of these variables missing or with incomplete values, Duplicate entries; Physiologically impossible/implausible data values (e.g., negative, out-of-range clinical measurements) and Cases detected as extreme outliers during preprocessing (detected using the Isolation Forest detection).

Variables and risk factors

Thirteen relevant features-both categorical and numerical-presented the wide physiological and medical signs that might be related to the risk of heart disease. Each of these variables was a demographic, clinical, or physiological measurement that gave a different view of cardiovascular health (detailed characteristics are available in **Table 1**).

Proposed deep learning model

There are two main ways to implement deep learning models: with the Sequential API or the Functional API. Sequential is for representing simple architectures, where the layers are stacked in a linear manner, and the functional API is for dynamic architectures, that is, architectures, not necessarily so straightforward

(for example, architectures containing more than one input and/or output, skip connections, shared layers). In this study, we used a Keras Sequential model due to its lightweight and relevancy for clinical tabular data analysis. The design comprised three fully connected dense layers with ReLU activation, dropout layers to reduce overfitting and a flatten layer. We trained the model on binary cross-entropy loss with an Adam optimizer, using the mini-batch learning and early stopping based on a validation loss. The results of this deep learning model were compared to legacy machine learning models (e.g. Random Forest) and reported in the results section.

Evaluation metrics employed

Some of the different metrics of model evaluation include the confusion matrix, accuracy score, precision, recall, sensitivity, and F1 score. A confusion matrix is a kind of tabular structure that categorizes true and predicted values into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

1. True Positive (TP): Instances correctly identified as positive.

2. False Positive (FP): Instances incorrectly identified as positive.

3. False Negative (FN): Instances incorrectly identified as negative.

4. True Negative (TN): Instances correctly identified as negative.

To evaluate the model performance, the accuracy score is determined by the formula:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Specificity is the measure of true negative cases correctly identified, or the true negative rate, and is determined by the formula:

$$specificity = \frac{TN}{TN + FP}$$

The percentage of real positive cases that are accurately predicted to be positive is measured by sensitivity, also referred to as recall:

$$sensitivity = \frac{TP}{TP + FN}$$

By comparing the models' efficacy in different situations, these metrics offer a thorough assessment of the models' performance.

Computational methodology

The data were processed using a strong computational environment of Python 3 and SPSS Package 27. For the analyzed data to be presented, they have to show strict standards when compared using less than 0.05 *p*-value thresholds of significance. On machine learning algorithms, these increase accuracy for the predictions.

Ethical considerations

This is ensured through a dataset that guarantees total anonymity of patients, with all personally identifiable information removed in compliance with stringent medical data privacy regulations. This corresponds to strict ethical guidelines in medical research.

Results
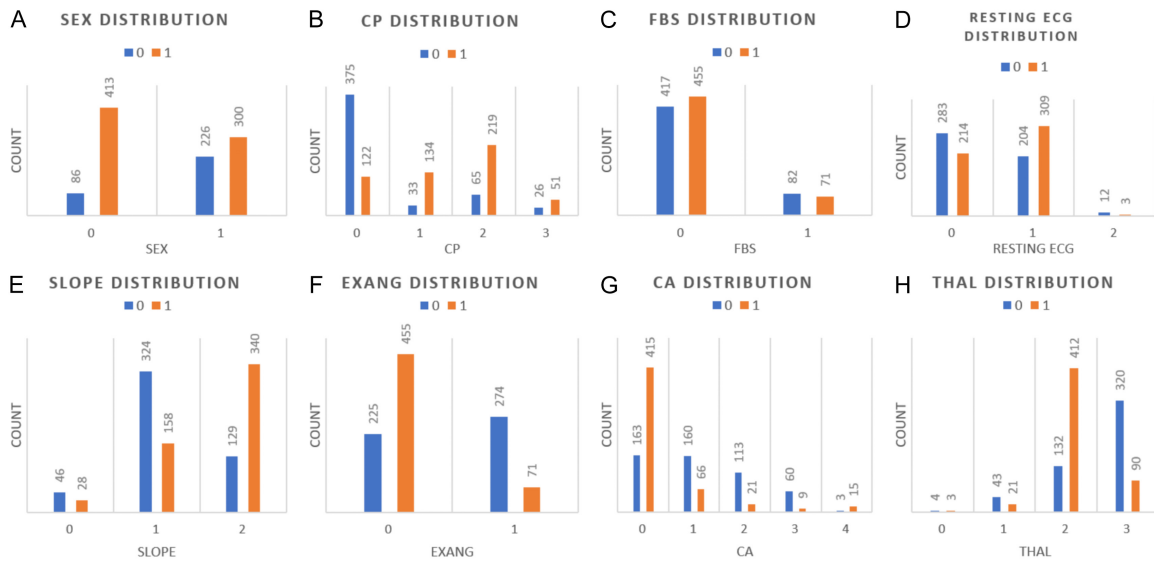
Checking the distribution of the data

Data distribution is important in any problem for prediction or classification. There is a slight imbalance, which needs to be resolved so as not to result in overfitting of the model. Balancing the dataset will enable the model to determine the pattern leading to heart disease and those that are not.

Dataset preprocessing

The dataset did not contain any null values, but it did need some corrections in the distribution of data and handling a lot of outliers. In the first attempt, the desired result was not achieved since it lacked feature selection and handling of outliers. Later, promising results were obtained using normalization to handle overfitting of the dataset and applying an Isolation Forest method for outlier detection. Plotting graphs was done by applying various kinds of plotting for studying the distribution of data, identification of outliers, and computing data skewness. The above preprocessing is necessary in terms of classification and prediction analysis.

Investigating the data's skewness

Plots of the attribute values and the skewness (asymmetry) of the data distribution were

Figure 1. Distribution of key health metrics by target outcome. Eight subplots (A to H) showing the distribution of various health metrics among individuals with and without heart disease: (A) Gender Distribution (Sex), (B) Chest Pain Type Distribution (cp), (C) Fasting Blood Sugar Distribution (fbs), (D) Resting ECG Results Distribution (restecg), (E) Slope Distribution (slope), (F) Exercise-Induced Angina Distribution (exang), (G) Calcium Distribution (ca), (H) Thalassemia Distribution (thal).

established. These plots offer insights into the distribution of multiple variables, including sex, chest pain (cp), fasting blood sugar (fbs), resting electrocardiogram (restecg), exercise-induced angina (exang), slope of the peak exercise ST segment (slope) and number of major vessels (ca), and thalassemia-related defect (thal) and target. Examining these distributions, as illustrated in **Figure 1A-H**, yields a thorough summary of the information.

Description of the dataset

This dataset includes information from 1,025 patients. This variable is a target variable with 0 indicating no heart disease and 1 indicating a case of heart disease. **Table 1** provides a detailed view of the study variables: age, sex, type of chest pain, resting blood pressure, serum cholesterol level, fasting blood sugar level, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise relative to rest, slope of the peak exercise ST segment, number of major vessels colored by fluoroscopy, thalassemia-related defect, and target.

Visualization of attribute distributions

Figure 2 presents the distribution of various attributes, which are age, sex, chest pain, rest-

ing blood pressure, cholesterol, fasting blood sugar, resting electrocardiogram results, maximum heart rate, exercise-induced angina.

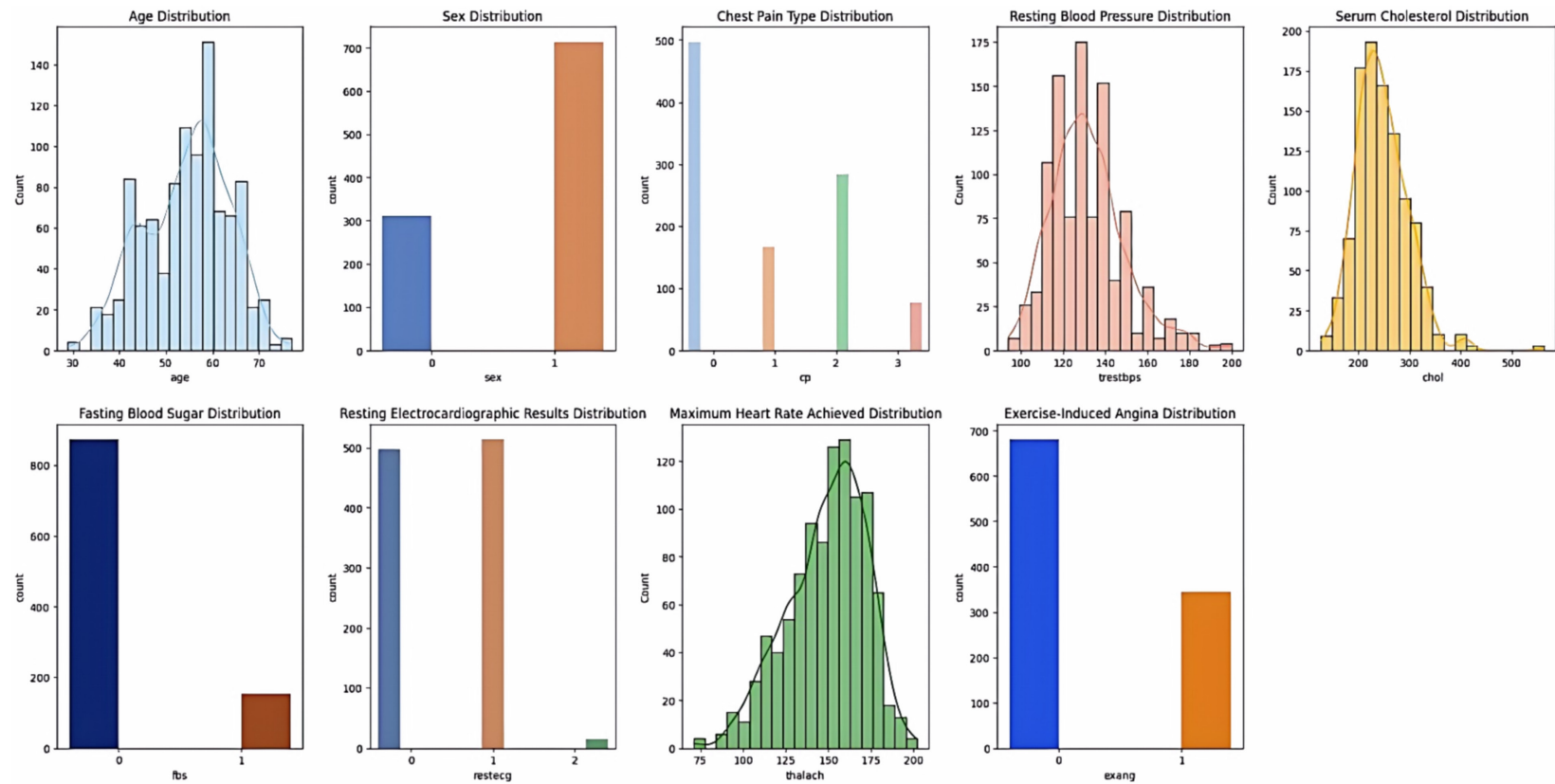*Attribute distributions and heart attack statistics*

In this dataset, 72% of females and 42% of males experienced heart attacks. The data indicates a 79% chance that atypical angina and a 77% chance that non-anginal pain will lead to a heart attack. There is a 67% chance that exercise-induced angina can result in a heart attack. Additionally, there is a 72% chance that downsloping of the peak exercise ST segment can lead to a heart attack. Moreover, there is a 71% chance that having no major coronary artery blockage (ca=0) may lead to a heart attack. Furthermore, there is a 75% chance that a potentially reversible thalassemia-related defect (thal=2) may lead to a heart attack. The age groups 40-44 and 50-54 are 76% and 61% prone to heart attacks, respectively.

Random forest classification report

Table 2 summarizes the Random Forest classification report, giving an overview of the model's performance metrics. In general, the classifier has an accuracy of 99%. The macro average

Figure 2. Comprehensive attribute distributions. Visualizations of various health metrics among individuals with and without heart disease: Age Distribution (age), Gender Distribution (sex), Chest Pain Type Distribution (cp), Resting Blood Pressure Distribution (trestbps), Serum Cholesterol Distribution (chol), Fasting Blood Sugar Distribution (fbs), Resting ECG Results Distribution (restecg), Maximum Heart Rate Distribution (thalach), Exercise-Induced Angina Distribution (exang), ST Depression Distribution (oldpeak), Slope of Peak Exercise ST Segment Distribution (slope), Number of Major Vessels Distribution (ca), Thalassemia Distribution (thal).

Table 2. Random forest classification report

Class	Precision	Recall	F1-score	Support
0	0.97	1.00	0.99	102
1	1.00	0.97	0.99	103
Accuracy			0.99	205
Macro avg	0.99	0.99	0.99	205
Weighted avg	0.99	0.99	0.99	205

Performance metrics of the Random Forest classifier, including precision, recall, F1-score, and support for each class (0 = no heart disease, 1 = heart disease), demonstrating nearly perfect classification accuracy and robustness in predicting heart disease.

and weighted average of precision, recall, and F1-score are also 0.99, reflecting balanced and robust model performance. Indeed, the Random Forest Classifier performs very well, with almost perfect precisions, recalls, and F1-scores for both classes. The model is quite effective at an accuracy of 99% while performing the given classification task. There is a slight difference in recall for class 1, but it does not make much difference in the results. These results emphasize two important aspects: comprehensive pre-processing and strong modeling are keys to high classification accuracy in medical datasets.

Discussion

Summary of findings

This comprehensive study applies advanced ML techniques to discover new patterns and risk factors of heart disease. The major findings point to the future role of ML in bringing a paradigm shift in the early detection and individualized treatment strategies to improve clinical outcomes of patients with heart disease.

The research was primarily conducted to analyze data from a really heterogeneous dataset, containing clinical reports from different parts of the world to find out the most significant predictors of heart attack. We used algorithms such as Random Forest, which performed much better compared to other classical methods in high performance for risk prediction.

Higher prevalence of heart attacks in females compared to males

The results of the study show that there is a significant difference in the frequency of heart attacks between women and men. From this point of view, the data analyzed showed that 72% of women suffered a heart attack, in contrast to 42% of men. This is consistent with recent statistics from the study by Rodgers et al. (2019), which indicate that more women are suffering from cardiovascular disease, especially at older ages [20].

This observation has undoubtedly challenged the traditional view that heart disease is a man's disease and called for a more gender-specific approach to the concept of cardiovascular health. The basis for such gender differences in heart attack risk is multifactorial, linking biological, hormonal and socioeconomic factors [21, 22]. Further research in this direction will help determine the exact mechanism for such a difference and could lead to more personalized prevention strategies and treatment protocols in female patients.

Although the findings are consistent with some studies [20, 23], the results of this study are completely at odds with some previous studies that have traditionally highlighted the higher prevalence in men [24]. This may actually be due to differences in the population, diagnostic criteria, and risk factor disease assessment methods used in previous studies. Heart disease presents differently in women and also has different and unique risk profiles that may not be given due importance or may even be underreported [22].

Therefore, the increased risk for heart attack in women also requires awareness and vigilance on the part of the health professional to ensure that cardiovascular screening, risk assessment, and management strategies are appropriately directed toward the specific needs and vulnerabilities of women [25]. These may offer opportunities for early detection, timely interventions, and better outcomes for those patients at risk from cardiovascular complications.

High risk associated with atypical and non-anginal chest pain

Unexpectedly, very high heart attack risks were found to be 79 percent associated with atypical

angina and 77 percent associated with non-anginal chest pain, which defied the established convention that typical angina is the most significant risk factor for coronary artery disease. This finding underscored the significance of a comprehensive assessment of chest pain and risk stratification.

Additional research should be done because this is a significant and high risk factor for atypical and non-anginal chest pain. In contrast to this study, the majority of previous research has linked these forms of chest pain to a low prevalence of coronary artery disease [26]. The differences could have been caused by many factors, including populations, diagnostic criteria, or specific clinical characteristics of the patients studied [27].

Further exploration of the potential mechanisms that explain the increased risk of heart attack associated with atypical and non-anginal chest pain is needed. The pathophysiology of the conditions, patient demographics, comorbidities, and other confounding factors that could influence these observed relationships should all be looked into [28]. Clarifying this interaction will enable the doctor to better identify patients who present with non-anginal or atypical chest pain and to provide them with appropriate, potentially earlier intervention [29].

The findings underscore the need for a more comprehensive and advanced method of cardiovascular risk assessment and the limitations of relying solely on traditional classifications of chest pain. A comprehensive assessment of patient-reported symptoms combined with the use of cutting-edge diagnostic techniques like functional imaging or biomarkers may improve risk stratification and direct the development of effective treatment plans for patients with a range of chest pain presentations [30, 31].

Significant predictors: exercise-induced angina and downsloping ST segments

In addition, the investigation found that exercise-induced angina and descending ST segments are strong predictors of myocardial infarction. All of these findings highlight the importance of considering clinical markers outside the traditional spectrum when assessing and treating cardiovascular disease [32].

This makes it reasonable when weighed against the well-established findings that exercise frequently reveals coronary artery disease and causes ischemic episodes [33]. However, a significant correlation found in this study would make it more important for medical professionals to stress testing and symptom assessment as a crucial, standard component of cardiovascular screening and stratification for the right interventions [34].

In the same way, downsloping ST segments associate with the risk of a heart attack in concert with the clinical relevance of this electrocardiographic finding [35]. Long regarded as a predictor of cardiovascular events and a sign of coronary artery disease, downsloping ST segments are suggestive of myocardial ischemia [36]. This study confirms the diagnostic importance of this parameter and points to the need for a comprehensive evaluation of electrocardiographic abnormalities within the general cardiovascular risk context.

These findings illustrate the need to look beyond traditional risk factors toward a comprehensive approach in assessing cardiovascular disease. A diagnostic modality, such as stress testing and electrocardiographic analysis, performs a much more precise diagnosis and thus completes the status of cardiovascular health so that the potential attack can be identified in order to carry out appropriate interventions in a timely manner [37].

Unexpected high risk in younger age groups

According to these findings, comparatively younger age groups have a surprisingly high risk of heart attacks. This deviates from the study by Díez-Villanueva et al. (2022), which indicate that cardiovascular diseases are on the rise among the elderly [38].

The disproportionately high risk of heart attack observed in younger adults is noteworthy. Disparities in study populations, risk factor profiles, or the unique clinical features of the younger patients included in the analysis are just a few of the possible causes of this [39].

Such an unexpected pattern emphasises how useful a data-driven ML approach is for reveal-

ing nuanced risk profiles that conventional epidemiological research might have missed [8, 40]. Indeed, such findings suggest that health care providers should remain extremely vigilant and target screening and prevention strategies toward younger individuals who may not fit the classic cardiovascular disease risk profile [39].

Indeed, only if the root causes are more clearly explained and the observation is validated in different population segments can the scientific community develop a better understanding of why the risk of heart attack continues to decline across age groups [39]. In fact, it would provide one with the knowledge to develop effective, tailored prevention and intervention programs that help improve cardiovascular health in this vulnerable population segment.

Association between reversible thalassemia-related defects and heart attack risk

These results also indicated that thalassemia-related defects of reversible nature are significantly associated with an increased risk of heart attack. This observation underlines the fact that some cardiac abnormalities, other than the traditional risk factors, may be highly relevant in the development of cardiovascular complications [41].

Thalassemia is a group of inherited blood disorders characterized by the production of abnormal hemoglobin [42] and usually presents with various clinical symptoms, including cardiac complications [43]. The fact that the study found a high risk of heart attack in people with reversible thalassemia-related defects [44] suggests that these hematological and cardiovascular factors may be closely linked [42].

While the mechanisms linking the presence of thalassemia-related defects to a heart attack are still unexplained, this observation indicates a need for a probable contribution of genetic and hematologic factors in the overall evaluation of cardiovascular health [45]. Health professionals need to be vigilant in screening and follow-ups of these unusual risk factors, especially in populations of higher prevalence of thalassemia or other associated hemoglobin disorders.

Compared to earlier discussions that have more traditionally focused on traditional car-

diovascular risk factors, the relationship between heart attack risk and reversible thalassemia-related defects may be less obvious in this study. This discrepancy underlines the values of ML-driven analyses in uncovering unexpected new patterns within complex medical data sets that inform a more comprehensive and personalized approach to cardiovascular risk assessment and management [46, 47].

These insights, when translated into clinical practice, will help in the early detection and timely intervention to improve outcomes in patients with underlying hematological conditions predisposing them to a high risk of cardiovascular complications [47]. Further studies are needed to confirm these findings and explain the exact mechanisms by which thalassemia-related defects may lead to the development of heart attacks.

Lack of visible coronary artery blockage as a predictor of heart attack

According to the study's findings, a surprising correlation of no obvious blockage of the main coronary arteries exists with an elevated risk of heart attack. The above observation brings under question the current belief of the existence and severity of coronary artery disease, assessed by angiographic examination, being major causes of cardiovascular events [48].

This is an unexpected and counterintuitive finding that even without any apparent narrowing or blockage of the main coronary arteries, a significant risk of heart attack may be present. The limitations of angiographic techniques in identifying functional or subclinical abnormalities [49], the part played by microvascular dysfunction [50], and the impact of additional pathophysiological mechanisms not captured by conventional imaging techniques are some of the factors that may be responsible for this disparity [51].

The reasons for this apparently unexpected association need to be pursued. This partly may relate to the presence of non-obstructive coronary artery disease leading to myocardial ischemia with increased cardiovascular risk in the absence of significant stenosis [52]. The study population could also have included individuals who had atypical presentations of coro-

nary artery disease, such as coronary artery spasm or microvascular dysfunction, both of which are not readily evident on angiographic examination [53, 54].

The discrepancy between these study results and conventional wisdom emphasizes a limitation in the assessment of cardiovascular risk based solely on angiographic evaluation. It reflects that detailed definition of the underlying pathophysiology and guiding appropriate management will be required with a more comprehensive, multi-modality approach to cardiovascular imaging and functional assessment using tests such as stress testing, intravascular imaging, or sophisticated computational analysis [55].

These findings clearly need replication, possible mechanisms underlying the association between no visible blockage of the coronary arteries and the risk of heart attack explored, and clinical implications for risk stratification and treatment decision-making in patients with atypical or nonobstructive coronary artery disease presentations investigated.

Strengths and limitations

This study's key advantages are that it uses sophisticated machine learning algorithms, which have outperformed conventional statistical techniques in predicting the risk of heart disease, and it is enhanced by a varied dataset of clinical reports from various geographical areas.

This study is based on a specific population, which cannot fully represent the entire population. In addition, possible biases within the data set and the fact that the data are retrospective in nature can affect model performance and introduce confounding factors into the model. While the Random Forest and Decision Tree algorithms showed promise in this particular study, a broader comparison of the model with other high-performance ML techniques has not been conducted and external validation is required for its clinical implications.

Clinical implications

These results have important clinical significance as they highlight the need for individualized and thorough assessment and treatment of cardiovascular risk. Such identified risk factors, including atypical angina, exercise-induced angina, descending ST segments, reversible thalassemia-related defects and the absence of visible blockage in the coronary arteries, must be taken into account by the treating physician in assessment and follow-up, particularly in women and the relative younger population. Integrating ML-based risk prediction models into routine clinical care may dramatically improve early detection and focused interventions, improving outcomes while decreasing the cost of healthcare by virtue of better and more appropriate treatments. By adding this advanced analytics component to routine cardiovascular care, health systems will come closer to a more personalized and proactive approach to heart disease prevention and management.

For the authors, these results not only demonstrate the technical possibility of applying deep learning to routine clinical data but also underscore the importance of careful interpretation. Anymore, the apparent performance we would not generally generalize to be high due to internal validation and the source of the data, such as UCI, etc. However, we still think that deep learning can aid traditional ML methods by revealing intricate, non-linear relationships between cardiovascular risk profiles. Critically, deployment of this model into clinical workflows must be informed by concerns surrounding transparency, demographic fairness and practicality in the real world.

Future research

Future studies will be important, then, for the validation of these findings across populations, and in the investigation of mechanisms underlying the risk patterns observed - including higher rates of heart attacks among females and younger age groups, unexpected associations with atypical chest pain presentations, and the absence of visible blockage of the coronary arteries.

Secondly, more comparative studies can be performed with a variety of advanced ML algorithms. Additionally, emerging data sources could provide new insights into understanding better the risk of heart diseases, such as genomic and environmental data, including lifestyle habits. The interaction of hematologi-

cal factors, for instance, thalassemia-related defects with the cardio outcome, may further indicate complex contributions to heart diseases in general.

Conclusion

This groundbreaking study has shown how much Machine Learning stands to revolutionize the way heart diseases are studied and treated. The results point toward new risk factors and patterns that have challenged conventional wisdom and emphasized an individualized and holistic approach toward cardiovascular health.

It will certainly redefine heart disease prevention and treatment when all these ML-driven insights are translated into clinical practice, leading to early detection and intervention for better patient outcomes. As the scientific community moves forward to explore the frontiers of this rapidly evolving field, the results of this study point to new directions toward a future in which heart disease can be better predicted, treated and ultimately prevented.

Acknowledgements

The authors would like to thank the researchers whose work was included in this study.

Disclosure of conflict of interest

None.

Abbreviations

AI, Artificial Intelligence; AHA, American Heart Association; AMI, Acute Myocardial Infarction; CART, Classification and Regression Tree; ECG, Electrocardiogram; ML, Machine Learning; NLP, Natural Language Processing; NYHA, New York Heart Association; SVM, Support Vector Machines; UCI, University of California Irvine.

Address correspondence to: Parisa Alsadat Dadkhah, Student Research Committee, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran. Tel: +98-2122437293; E-mail: pardad7697@gmail.com; Mahsa Asadi Anar, College of Medicine, University of Arizona, Tucson, AZ, USA. Tel: +98-2122437293; E-mail: Mahsa.boz@gmail.com; Asadianar@arizona.edu

References

[1] Tsao CW, Aday AW, Almarzooq ZI, Anderson CAM, Arora P, Avery CL, Baker-Smith CM, Beaton AZ, Boehme AK, Buxton AE, Commodore-Mensah Y, Elkind MSV, Evenson KR, Eze-Nliam C, Fugar S, Generoso G, Heard DG, Hiremath S, Ho JE, Kalani R, Kazi DS, Ko D, Levine DA, Liu J, Ma J, Magnani JW, Michos ED, Mussolino ME, Navaneethan SD, Parikh NI, Poudel R, Rezk-Hanna M, Roth GA, Shah NS, St-Onge MP, Thacker EL, Virani SS, Voeks JH, Wang NY, Wong ND, Wong SS, Yaffe K and Martin SS; American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics-2023 update: a report from the American Heart Association. Circulation 2023; 147: e93-e621.

[2] Chaurasia V. Early prediction of heart diseases using data mining techniques. Caribbean Journal of Science and Technology 2013; 1: 208-217.

[3] Hastie T, Tibshirani R, Friedman J and Franklin J. The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer 2005; 27: 83-85.

[4] Marsland S. Machine learning: an algorithmic perspective. Chapman and Hall/CRC, 2011.

[5] Shalev-Shwartz S and Ben-David S. Understanding machine learning: from theory to algorithms. Cambridge university press, 2014.

[6] Melillo P, De Luca N, Bracale M and Pecchia L. Classification tree for risk assessment in patients suffering from congestive heart failure via long-term heart rate variability. IEEE J Biomed Health Inform 2013; 17: 727-733.

[7] Al Rahhal MM, Bazi Y, AlHichri H, Alajlan N, Melgani F and Yager RR. Deep learning approach for active classification of electrocardiogram signals. Information Sciences 2016; 345: 340-354.

[8] Guidi G, Pettenati MC, Melillo P and Iadanza E. A machine learning system to improve heart failure patient assistance. IEEE J Biomed Health Inform 2014; 18: 1750-1756.

[9] Zhang R, Ma S, Shanahan L, Munroe J, Horn S and Speedie S. Automatic methods to extract new york heart association classification from clinical notes. Proceedings (IEEE Int Conf Bioinformatics Biomed) 2017; 2017: 1296-1299.

[10] Parthiban G and Srivatsa S. Applying machine learning methods in diagnosing heart disease for diabetic patients. International Journal of Applied Information Systems 2012; 3: 25-30.

[11] Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Blaha MJ, Dai S, Ford ES, Fox CS, Franco S, Fullerton HJ, Gillespie C, Hailpern

SM, Heit JA, Howard VJ, Huffman MD, Judd SE, Kissela BM, Kittner SJ, Lackland DT, Lichtman JH, Lisabeth LD, Mackey RH, Magid DJ, Marcus GM, Marelli A, Matchar DB, McGuire DK, Mohler ER 3rd, Moy CS, Mussolino ME, Neumar RW, Nichol G, Pandey DK, Paynter NP, Reeves MJ, Sorlie PD, Stein J, Towfighi A, Turan TN, Virani SS, Wong ND, Woo D and Turner MB; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics--2014 update: a report from the American Heart Association. Circulation 2014; 129: e28-e292.

[12] Krishnaiah V, Narsimha G and Subhash N. Heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review. International Journal of Computer Applications 2016; 136: 43-51.

[13] Rahi P and Kang SS. A novel paradigm in cardiovascular disease risk prediction through hybrid machine learning. Systems Engineering and Information Technology 2025; 49-66.

[14] Song Y, Ren S, Lu Y, Fu X and Wong KK. Deep learning-based automatic segmentation of images in cardiac radiography: a promising challenge. Comput Methods Programs Biomed 2022; 220: 106821.

[15] Raj S and Bayappu N. Multimodal deep learning in medical diagnostics: a comprehensive exploration of cardiovascular risk prediction. Prediction in Medicine: The Impact of Machine Learning on Healthcare 2024; 78-94.

[16] Khajehali N, Khajehali Z and Tarokh MJ. The prediction of mortality influential variables in an intensive care unit: a case study. Pers Ubiquitous Comput 2023; 27: 203-219.

[17] Kim YJ, Saqlian M and Lee JY. Deep learning-based prediction model of occurrences of major adverse cardiac events during 1-year follow-up after hospital discharge in patients with AMI using knowledge mining. Pers Ubiquitous Comput 2022; 26: 259-267.

[18] Olsen CR, Mentz RJ, Anstrom KJ, Page D and Patel PA. Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. Am Heart J 2020; 229: 1-17.

[19] Janosi A, Steinbrunn W, Pfisterer M and Detrano R. Heart Disease [dataset]. 1989. UCI Machine Learning Repository. Available from: https://doi.org/10.24432/C52P4X.

[20] Rodgers JL, Jones J, Bolleddu SI, Vanthenapalli S, Rodgers LE, Shah K, Karia K and Panguluri SK. Cardiovascular risks associated with gender and aging. J Cardiovasc Dev Dis 2019; 6: 19.

[21] Reue K and Wiese CB. Illuminating the mechanisms underlying sex differences in cardiovascular disease. Circ Res 2022; 130: 1747-1762.

[22] Barton JC, Wozniak A, Scott C, Chatterjee A, Titterton GN, Corrigan AE, Kuri A, Shah V, Soh I and Kaski JC. Between-sex differences in risk factors for cardiovascular disease among patients with myocardial infarction-a systematic review. J Clin Med 2023; 12: 5163.

[23] Khoja A, Andraweera PH, Lassi ZS, Ali A, Zheng M, Pathirana MM, Aldridge E, Wittwer MR, Chaudhuri DD, Tavella R and Arstall MA. Risk factors for premature coronary heart disease in women compared to men: systematic review and meta-analysis. J Womens Health (Larchmt) 2023; 32: 908-920.

[24] Faucon A, Lambert O, De Pinho A, Ayav C, Combe C, Fouque D, Frimat L, Jacquelinet C, Laville M, Liabeuf S, Massy Z, Nicolas M and Stengel B. MO499: incidence of cause-specific cardiovascular events in Men and Women with CKD. Nephrol Dial Transplant 2022.

[25] Kazzi B, Shankar B, Elder-Odame P, Tokgözoğlu LS, Sierra-Galan LM and Michos ED. A woman's heart: improving uptake and awareness of cardiovascular screening for middle-aged populations. Int J Womens Health 2023; 15: 1171-1183.

[26] Nakas G, Bechlioulis A, Marini A, Vakalis K, Bougiakli M, Giannitsi S, Nikolaou K, Antoniadou EI, Kotsia A, Gartzonika K, Chasiotis G, Bairaktari E, Katsouras CS, Triantis G, Sionis D, Michalis LK and Naka KK. The importance of characteristics of angina symptoms for the prediction of coronary artery disease in a cohort of stable patients in the modern era. Hellenic J Cardiol 2019; 60: 241-246.

[27] Cosmi D, Tarquini B, Mariottoni B, Bennati M, Morganti M and Cosmi F. Typical and atypical chest pain: an equivocal and dangerous definition. European Heart Journal 2023; 44.

[28] Lee J, Oh O, Park DI, Nam G and Lee KS. Scoping review of measures of comorbidities in heart failure. J Cardiovasc Nurs 2024; 39: 5-17.

[29] Kaushika S. Atypical presentation of myocardial infarctions: more common above 60 years and females. Journal of Medical Science and Clinical Research 2018; 6.

[30] Raat W, Nees L and Vaes B. Diagnostic accuracy of signs and symptoms in acute coronary syndrome and acute myocardial infarction. Scand J Prim Health Care 2025; 43: 111-119.

[31] Boyle RSJ and Body R. The diagnostic accuracy of the Emergency Department Assessment of Chest Pain (EDACS) score: a systematic review and meta-analysis. Ann Emerg Med 2021; 77: 433-441.

[32] Thomas MR and Lip GY. Novel risk markers and risk assessments for cardiovascular disease. Circ Res 2017; 120: 133-149.

[33] Vargas JL, Pereira-Rodríguez J, Perez-Vazques D, Leyva-Valadez E, Lastra-Silva V, Florez DP,

Xavier-Santos G and Galdino G. Effect of exercise-based cardiac rehabilitation on the ischemic threshold in patients with high-risk ischemic heart disease. European Heart Journal 2022; 43.

[34] Samuel TJ, Beaudry R, Sarma S, Zaha V, Haykowsky MJ and Nelson MD. Diastolic stress testing along the heart failure continuum. Curr Heart Fail Rep 2018; 15: 332-339.

[35] Kawaji T, Hamatani Y, Kato M, Yokomatsu T, Miki S, Abe M and Akao M. Clinical significance of ST-segment depression during atrial fibrillation rhythm for subsequent heart failure events. Eur Heart J Open 2023; 3: oead060.

[36] Hakim SM, Elfawy DM, ElSerwi HB and Saad MK. Value of new ST-Segment/T-Wave changes for prediction of major adverse cardiac events after vascular surgery: a meta-analysis. Minerva Anestesiol 2020; 86: 652-661.

[37] Hadida Barzilai D, Cohen-Shelly M, Sorin V, Zimlichman E, Massalha E, Allison TG and Klang E. Machine learning in cardiac stress test interpretation: a systematic review. Eur Heart J Digit Health 2024; 5: 401-408.

[38] Díez-Villanueva P, Jiménez-Méndez C, Bonanad C, García-Blas S, Pérez-Rivera Á, Allo G, García-Pardo H, Formiga F, Camafort M, Martínez-Sellés M, Ariza-Solé A and Ayesta A. Risk factors and cardiovascular disease in the elderly. Rev Cardiovasc Med 2022; 23: 188.

[39] Sagris M, Antonopoulos AS, Theofilis P, Oikonomou E, Siasos G, Tsalamandris S, Antoniades C, Brilakis ES, Kaski JC and Tousoulis D. Risk factors profile of young and older patients with Myocardial Infarction. Cardiovasc Res 2022; 118: 2281-2292.

[40] Katarya R and Meena SK. Machine learning techniques for heart disease prediction: a comparative study and analysis. Health and Technology 2021; 11: 87-97.

[41] Wang M, Li Y, Li S and Lv J. Endothelial dysfunction and diabetic cardiomyopathy. Front Endocrinol (Lausanne) 2022; 13: 851941.

[42] Ajassa M, Gaglioti C, Longo F, Piga A, Ferrero G and Barbero U. P165 cardiovascular risk factors and hypogonadism influence on cardiac outcomes in an aging population of beta-Thalassemia patients: looking at the heart of the problem. Eur Heart J Suppl 2022; 24.

[43] Wood JC. Cardiac complications in thalassemia throughout the lifespan: victories and challenges. Ann N Y Acad Sci 2023; 1530: 64-73.

[44] Auger D and Pennell DJ. Cardiac complications in thalassemia major. Ann N Y Acad Sci 2016; 1368: 56-64.

[45] Yuan S, Burgess S, Laffan M, Mason AM, Dichgans M, Gill D and Larsson SC. Genetically proxied inhibition of coagulation factors and risk of cardiovascular disease: a mendelian randomization study. J Am Heart Assoc 2021; 10: e019644.

[46] Khourdifi Y and Baha M. Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization. International Journal of Intelligent Engineering and Systems 2019; 12.

[47] Sun W, Zhang P, Wang Z and Li DX. Machine learning-based prediction of cardiovascular diseases. IECE Transactions on Internet of Things 2024; 2: 50-54.

[48] Savo MT, De Amicis M, Cozac DA, Cordoni G, Corradin S, Cozza E, Amato F, Lassandro E, Da Pozzo S, Tansella D, Di Paolantonio D, Baroni MM, Di Stefano A, De Conti G, Motta R and Pergola V. Comparative prognostic value of coronary calcium score and perivascular fat attenuation index in coronary artery disease. J Clin Med 2024; 13: 5205.

[49] Nikopoulos S, Papafaklis MI, Tsompou P, Sakellarios A, Siogkas P, Sioros S, Fotiadis DI, Katsouras CS, Naka KK, Nikas D and Michalis L. Virtual hemodynamic assessment of coronary lesions: the advent of functional angiography and coronary imaging. J Clin Med 2024; 13: 2243.

[50] Camici P and Rimoldi O. Imaging of microvascular disease. 2021; 481-494.

[51] Kshirsagar J, McNulty J, Taji B, So D, Chong AY, Theriault-Lauzier P, Wisniewski A and Shrimohammadi S. Generative AI-assisted novel view synthesis of coronary arteries for angiography. 2024 IEEE International Symposium on Medical Measurements and Applications (MeMeA) 2024; 1-6.

[52] Kang W, Lee CA, Kang G, Paeng DG and Choi J. A novel method for angiographic contrast-based diagnosis of stenosis in coronary artery disease: in vivo and in vitro analyses. Diagnostics (Basel) 2024; 14: 1429.

[53] Vancheri F, Longo G, Vancheri S and Henein M. Coronary microvascular dysfunction. J Clin Med 2020; 9: 2880.

[54] Mileva N, Nagumo S, Mizukami T, Sonck J, Berry C, Gallinoro E, Monizzi G, Candreva A, Munhoz D, Vassilev D, Penicka M, Barbato E, De Bruyne B and Collet C. Prevalence of coronary microvascular disease and coronary vasospasm in patients with nonobstructive coronary artery disease: systematic review and meta-analysis. J Am Heart Assoc 2022; 11: e023207.

[55] Heringlake M. Cardiovascular hemodynamics: an introductory guide, 2nd ed. Anesthesia & Analgesia 2020; 131: e55-56.