Risk Factors Categorizations of Ischemic Heart Disease in South-Western Bangladesh

M. Raihan[1,4†], Sami Azam[2], Laboni Akter[3], Md. Mehedi Hassan[4†], Ryana Quadir[5], Asif Karim[2†], Saikat Mondal[4], Arun More[6]

[1]Department of Computer Science and Engineering, North Western University, Khulna 9100, Bangladesh
[2]Faculty of Science and Technology, Charles Darwin University, NT 0909, Australia
[3]Department of Biomedical Engineering, Khulna University of Engineering & Technology, Khulna, Bangladesh
[4]Computer Science and Engineering Discipline, Khulna University, Khulna 9208, Bangladesh
[5]Department of Computer Science and Engineering, Daffodil International University (DIU), Bangladesh
[6]Department of Cardiology, Ter Institute of Rural Health and Research, Murud-413510, India

Keywords: Ischemic heart disease; Machine learning; CVD; Data categorization; Medical data

Citation: Raihan M., Azam S., Akter L., et al.: Risk Factors Categorizations of Ischemic Heart Disease in South-Western Bangladesh. Data Intelligence 6(3), 834-868 (2024). doi: 10.3724/2096-7004.di.2024.0002
Submitted: September 25, 2023; Revised: February 19, 2024; Accepted: May 10, 2024

ABSTRACT

Ischemic heart disease (IHD) is one of the leading causes of death worldwide. However, different geographic regions show different variations of the risk factors of this disease based on the different lifestyles of people. This study examines the current IHD condition in southern Bangladesh, a Southeast Asian middle-income country. The main approach to this research is an AI-based proposal of a reduced set of the greatest impact clinical traits that may cause IHD. This approach attempts to reduce IHD morbidity and mortality by early detection of risk factors using the reduced set of clinical data. Demographic, diagnostic, and symptomatic features were considered for analysing this clinical data. Data pre-processing utilizes several machine learning techniques to select significant features and make meaningful interpretations. A proposed voting mechanism ranked the selected 138 features by their impact factor. In this regard, diverse patterns in correlations with variables, including age, sex, career, family history, obesity, etc., were calculated and explained in terms of voting scores. Among the 138 risk factors, three labels were categorized: high-risk, medium-risk, and low-risk features; 19 features were regarded as high, 25 were medium, and 94 were considered low impactful features. This research's technological

† Corresponding author: Asif Karim, M. Raihan and Md. Mehedi Hassan (E-mail: asif.karim@cdu.edu.au; ORCID: 0000-0001-8532-6816; E-mail: raihanbme@gmail.com; ORCID: 0000-0003-1072-3555; E-mail: mehedihassan@ieee.org; ORCID: 0000-0002-9890-0968).

methodology and practical goals provide an innovative and resilient framework for addressing IHD, especially in less developed cities and townships of Bangladesh, where the general population's socio-economic conditions are often unexpected. The data collection, pre-processing, and use of this study's complete and comprehensive IHD patient dataset is another innovative addition. We believe that other relevant research initiatives will benefit from this work.

1. INTRODUCTION

Cardiovascular diseases are the primary cause of death globally, and there are approximately 17.9 million deaths from heart-related diseases each year, estimated as 32% of all global deaths [1]. These heart diseases include Intracerebral Haemorrhage (ICH), Ischemic Heart Disease (IHD), Congenital Heart Disease (CHD), Uncertainty Interval (UI), and Years of Life Lost (YLLs) to premature mortality [2]. With improved socio-economic status, urbanization, and changes in dietary habits and lifestyle, Ischemic heart disease is increasing in developing countries. 'Ischemic' means that an organ, like the heart, is not getting enough blood and oxygen [4]. Ischemic heart disease, also called coronary heart disease or coronary artery disease, refers to heart problems caused by narrowed heart (coronary) arteries that supply blood to the heart muscle. IHD is a group of chronic diseases characterized by coronary ischemia of the heart and subsequent myocardial damage [5]. IHD rose from third to first (moderately contributing to the top leading) global years of life lost cause between 1990 and 2019. Over the past decade, the rate of Ischemic Heart Disease has increased significantly throughout the world [3]. From the current global situation, it can be predicted that in 2040, 8 of the top 10 causes of death worldwide will be non-communicable diseases (NCDs), with coronary heart diseases and stroke continuing to be in the first and second places, respectively. If we look at the global scenario, different geographic regions show different variations of the risk factors of heart disease. In the United States, where one person dies every 36 seconds due to cardiovascular disease, high blood pressure, high blood cholesterol, excessive alcohol use, unhealthy diet, overweight, obesity, and smoking are the leading causes of these heart diseases [6]. A similar situation can be found in the Australian continent, where hypertension, poor diet, smoking, excessive consumption of saturated fats and trans fats, and poor management of diabetes are primarily responsible. Nevertheless, over the last decade, the number of deaths due to cardiovascular disease has been declining in this region [7].

For European countries, excess alcohol use, smoking, suffering from long-term depression, air pollution (particulate matter), etc., are reportedly attributable to causing heart disease [8]. However, South Asians have the highest prevalence of heart disease compared to Europeans and Chinese, possibly due to its massive population size and early onset of the disease [9]. South Asia has the highest total burden of cardiovascular diseases, and there was a 73% increase in the years of life lost in this region in the last decade [16]. In Bangladesh, stroke and coronary heart disease are ranked as the two major causes of death to date, and like other heart diseases, the rate of Ischemic heart disease has risen during the last couple of decades [11]. According to the latest WHO data in 2021, coronary heart disease deaths in Bangladesh reached 118,287, or 15.23% of total deaths [11]. One of the primary reasons was the rapid urbanization in the last few decades [12]. The results of urbanization and fast economic growth increase the concern that a further rise in the disease burden may be seen due to habituation of a sedentary lifestyle, changing food habits, including consumption of processed food, irregular mealtimes, and reduced physical activities in the urban part [13]. Moreover, with the outburst of the COVID-19 pandemic in Bangladesh, the death toll from cardiac arrests and strokes has increased in 2020 (21.1%) compared to 2019 (17.9%) and 2018 (15.23%), according to the Bangladesh Bureau of Statistics (BBS) [14].

Like other South Asians, Bangladeshis are unduly prone to develop coronary artery diseases, often premature in onset, follow a rapidly progressive course, and are angiographically more severe [10]. Therefore, the primary goal is to detect the risk factors with clinical authenticity to ensure early or premature detection and stop the progression of the disease. We believe that if the risk factors can be categorized by their significance using the reduced set of clinical data, early detection is possible. There are different ways to identify the significant features or risk factors, but in this study, we employed 10 different Machine Learning techniques and a novel voting algorithm to rank these features into 3 categories.

In studying cardiovascular risk factors among the Bangladeshi population, the use of smokeless tobacco (chewing betel quid), high cholesterol, hypertension, and stressful life events have been found to be closely associated. During this study, we also found these references in our collected data. Unlike developed countries, some novel risk factors, including hypo-vitamins D, arsenic contamination in water and foodstuff, particulate matter, and air pollution, may play a unique role in Bangladesh [10]. Although high blood pressure and diabetic patients are the most vulnerable to developing coronary artery disease, most (76.7%) of the rural community have no or poor knowledge about this fact. While examining the lifestyle and cardiovascular risk factors in rural communities, it might be the ignorance about the risk factors of the disease and the unhealthy lifestyle pattern that are primarily responsible. A Chicago, IL, USA study reported that risk factors could motivate these rural people to change their risky behavior toward the disease [16]. Since no individual factor is responsible for any heart disease, assessing multiple risk factors associated with developing heart disease can bring more prudence and allow more cost-effective treatment and management. Besides, the association between lifestyles and cardiovascular risk factors in rural parts of the southwestern zone of Bangladesh was not focused separately on heart disease prevalence. This research helps us statistically analyze these risk factors and their impact on causing ischemic heart disease. Thus, we can summarize the contribution of this paper as follows:

a. The data were collected from one of the Heart Institutes in Khulna, Bangladesh, under the supervision of a Cardiologist which ensures the authenticity and regularisation of the proper clinical data
b. Feature engineering was done by employing various machine learning techniques which were selected from prior experiments done in the same field
c. A novel voting algorithm was developed based on the threshold values of the ML methods used in the feature selection step
d. Based on the voting algorithm, features were categorized into 3 different severity classes, which clearly indicates the risk factors of IHD.

2. LITERATURE

Quite a lot of research has been initiated and carried out in analyzing different heart diseases in recent years. Several clinical information and symptoms are often related to IHD, including age, blood pressure, total cholesterol, diabetes, hypertension, etc., which are frequently correlated with each other. With the increase of the massive amount of datasets available in recent years, the diagnosis of IHD can be performed using traditional statistical methods on each patient. However, it has been noticed that the diagnosis based on Machine Learning (ML) techniques demonstrated more precise identification of the risk factors than the traditional practices and has recently gained a lot of traction in several sub-disciplines of Health Informatics research [45-50]. Some work provided practical ways to condense the attribute list to a selective one. This section explores some relevant research work showing the contributing attributes or risk factors for developing heart disease. From this prior analysis, we would later make relevance with our experimented risk factors. A number of the studies did rely on statistical approaches, while some would exhibit the ML-based work relevant to this research.

R. Elosua, Carla Lluís-Ganella, et al. [12] evaluated the impact of risk factor confluence on ischemic heart disease (IHD) risk. They selected genetic variants associated with blood pressure, body mass index (BMI), waist circumference, triglycerides, type-2 diabetes mellitus, high-density lipoprotein, and low-density lipoprotein cholesterol and calculated Genetic Risk Scores (GRSs). The association between GRS and IHD was tested using the statistical method Logistic Regression (LR) model in the case-control study. Eventually, they figured out that the confluence of low-density lipoprotein cholesterol and triglycerides genetic risk load had an additive effect on IHD risk. The interaction between low-density lipoprotein cholesterol and IHD genetic load was more than multiplicative. This interaction had an extremely hazardous impact on atherosclerosis progression of inflammation and increased lipid levels. Their evaluation has relevance with some of our experimented factors, namely, BMI, and low-density lipoprotein, which gave us the confidence to bank on our analysis.

Stacey E. Alexeeff, Noelle S. Liao et al. [13] very recently investigated the association between long-term exposure to delicate particulate matter due to air pollution and increased risk of IHD mortality, cerebrovascular mortality, and incident stroke. Their analysis showed that long-term exposure had a higher impact on cardiovascular mortality than short-term exposure. Their result concluded that this particulate matter-2.5 increase was associated with a 23% increased risk of IHD mortality and a 24% increased risk of cerebrovascular mortality. Meanwhile, our study found cardiogenic shock to be a major risk factor, but as we examined only the clinical features, particulate matter air pollution (particles suspended in the air that can be harmful to human health) was not included in the feature set. The authors combined the studies of the United States-Europe cases with some Asian, African, Canadian, and Australian case studies. Although there are limited studies on cardiovascular risks among the Bangladeshi population, Abu Abdullah Mohammad Hanif, Mehedi Hasan, MdShowkat Ali Khan, Md. Mokbul Hossain, Abu Ahmed Shamim, et al. found a nationally representative survey on 10-year cardiovascular risk using a non-laboratory-based World Health Organization risk chart [15]. It showed that insufficient physical activity, smoking tobacco, more sedentary time in an urban lifestyle, inadequate intake of fruits and vegetables, excess saturated and trans-fat, high salt intake, and self-reported diabetes were commonly found reasons that are positively associated with elevated cardiovascular disease risk among both sexes. They carried out uni-variable and multivariable Logistic Regression to identify factors associated with high cardiovascular risk. ShiuLun Au Yeung, Hugh Simon Hung San Lam, and C. Mary Schooling [17] assessed the ischemic heart disease (IHD) risk in 2017 by genetically predicting the Vascular Endothelial Growth Factor (VEGF). According to them, genetic variants are unlikely to be affected by lifestyle or socio-economic position factors. Instead, there was a possibility that some specific types of VEGF, for which genetic predictors have not yet been identified, might play a role in developing IHD.

Regarding the machine learning approaches in determining heart disease elements, several substantial works suggested practical ways of applying ML to analyze and construct data-driven diagnostic models. These ML methods had a great influence on our study for selecting some of the techniques. As proper feature selection plays an enormous role in medical diagnosis, choosing a good subset of features can reduce the treatment cost. JafarAbdollahi and Babak Nouri-Moghaddam et al. [14] proposed using the Wrapper method [18] for feature selection in diagnosing heart disease, which we employed in our experiment as well. They used the dataset of the University of California, Irvine (UCI) collected by David Aha [2]. The original dataset contained [13] numeric attributes like age, chest_pain, blood_pressure, cholesterol, blood_sugar, electrocardiographic, heart_rate, depression, slope, ca, thal_scan, vac, etc., and 14 facts indicating the patient's heart condition. According to their result, thallium scan and vascular occlusion were the essential features in diagnosing heart disease with a 97.57% accuracy. However, the computational cost of this method was very high, which was a major drawback of their work.

To determine optimal features that contribute more towards the diagnosis of heart disease, HoudaBenhar, Ali Idri, and Mohamed Hosni came up with an approach to assess the impact of threshold values for feature selection in heart disease classification [19]. They chose three feature ranking methods, and information gain (IG) was one of them that we selected as our primary technique. These methods needed a set of threshold values for determining the number of relevant features to be selected to decide on the final subset of features. The optimal threshold value with a combination of classifiers was later used for classification. We placed this concept of optimal threshold value on top of designing the voting algorithm. For the dataset, they employed four different datasets, including the Cleveland, Statlog, and Arrhythmia datasets, for training and testing combinations of different classifiers. As a feature selection method, IG used entropy to measure how much information a single feature gave them about the class at a time. It improved the selection of threshold values for the reduced feature set. The unprocessed dataset ultimately improved their experiment, which concludes that more feature ranking techniques are needed to construct ensemble feature ranking methods for better results. To enhance the performance of various heart disease prediction models made up of machine learning techniques, namely Naïve Bayes, SVM, and Random Forest, Ramesh, et al. [22] proposed the prediction models evaluated with an information gain feature selection algorithm.

To improve the heart disease classifier performance, T Desyani, A Saifudin, and Y Yulianti proposed a feature selection model to select relevant features and provide improved performance predictions [27].

Out of five feature selection techniques, SelectKBest gave the best performance (65% accuracy was the best performance), making it a part of their final proposed model. Although the model's accuracy was a mere 65% with SelectKBest, it outperformed their experiment's other feature selection methods. Senan et al. [28] suggested a machine-learning-based technique for distinguishing the essential associated variables in the electronic clinical records of cardiovascular disease patients. The SelectKBest function was used with a chi-squared statistical approach to find the most relevant characteristics in this investigation. The SelectKBest function delivers the Cleveland dataset's initial features with the highest scores. For this investigation, the exang oldpeak feature received the most incredible score of 652.85, while the fbs feature received the lowest score of 0.18. Ramalingam et al. [26] also discussed using machine learning to predict cardiovascular diseases or heart-related diseases using machine learning effectively. In this study, the feature selection technique Best first search was used. Huang et al. [29] developed multi-stage frameworks for developing a survival model for predicting the mortality of older people who underwent their first CABG. They used a fivefold cross-validation strategy to test the LGR, RF, CART, XGBoost, and MARS models, compare predicted performance with all variables, and evaluate classification results following feature selection per classification algorithm. Meanwhile, Sahin et al. [37] proposed a malware detection system based on machine learning to identify Android malware from benign applications. Linear regression eliminates needless feature selection during the feature selection step of the proposed malware detection system. Kakade et al. [31] analyzed feature selection using Logistic Regression in Case-Control DNA methylation data from Parkinson's disease. The methylation patterns of 1726 transcripts were collected from 66 samples of 450k, 43 of which were controls and 23 of which were sick. This work utilized Logistic Regression (LR) to reduce the number of features and create a classifier that outperformed all of the combined variables. To assess the significance of gully erosion susceptibility factors, Ahmadpour et al. [39] used the Boruta feature selection method. In this study, results and outputs of the Boruta algorithm were distance to the stream (33.5), land use (17.41), elevation (12.18), and lithology (7.34) were the most crucial gully erosion factors, followed by soil type, drainage density, LS factor, and slope angle.

In predicting mortality with heart failure caused by coronary heart disease, Ke Wang, Jing Tian, Chu Zheng, et al. [42] successfully deployed machine learning and Shapley Additive exPlanations (SHAP) methods. ML combined with SHAP provided an explicit explanation of individualized risk prediction for mortality and could give physicians an intuitive understanding of the influence of critical features in the model. According to their work, age, occupation, and nitrate drug use were impactful factors for both genders, and the best-performing model was the extreme gradient boosting (XGBoost) model with an AUROC of 0.8207. The SHAP value evaluated the importance of the output containing all combinations of features and provided consistent and locally accurate attribute values for each element in the prediction model. SHAP values and plots proved that the ML method could illustrate the key features and establish a high-accuracy mortality prediction model. Nevertheless, the information collected in this study was structured data, and this model was not well-validated and compared using a risk calculator. Still, the performance of SHAP and XGBoost algorithms were the key points of this research.

3. MATERIALS AND METHODS

The overview of the proposed approach is shown in Figure 1. In the beginning, the data on Ischemic Heart Disease (IHD) from the hospital was collected, and then we had the data cross-validated with the specialists in the medical field. Subsequently, we put into action several preprocessing strategies, such as providing solutions for missing data and doing data imputation. In the next steps, we put into action several machine learning strategies that are commonly utilized, including information gain, best first search, and SHAP analysis. Using the input of medical professionals, we were able to establish the appropriate threshold value. Afterward, we provided an algorithm that was designed to categorize the risk variables into three distinct groups: high, medium, and low category.

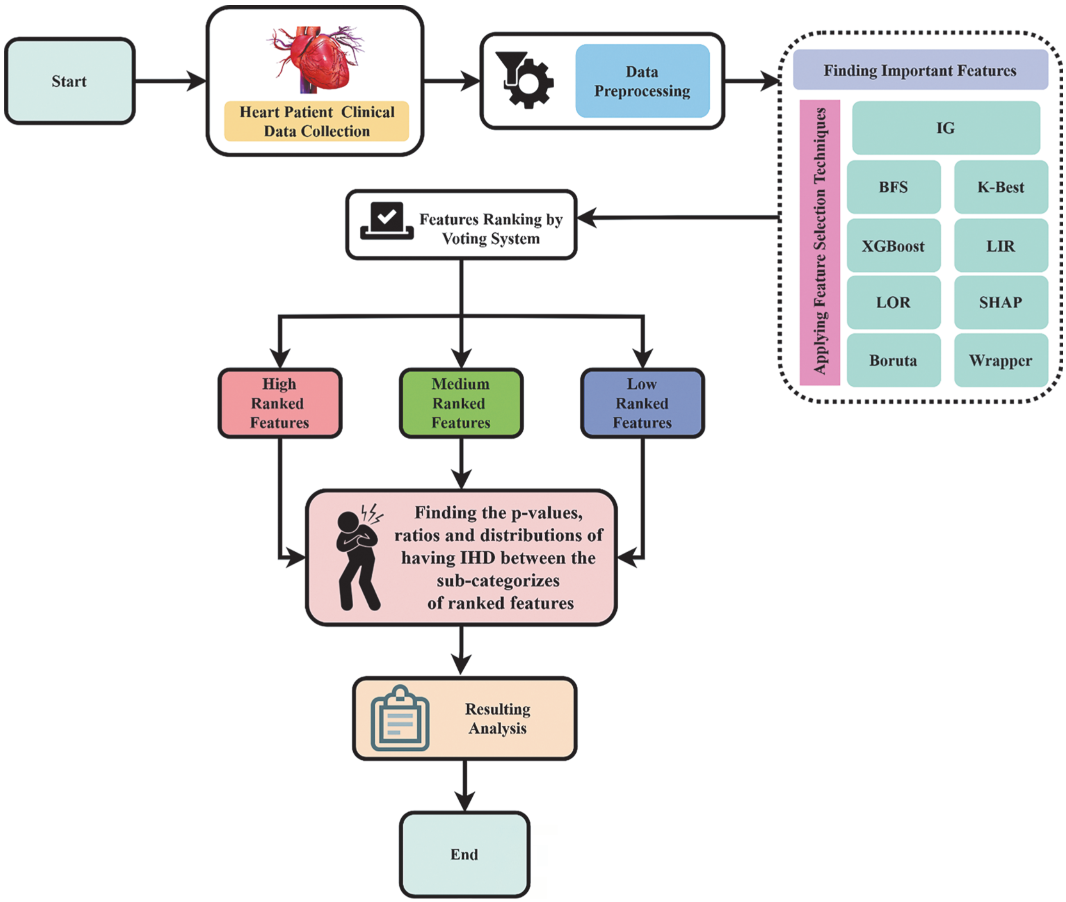


Figure 1. Overview of the proposed approach.

3.1 Dataset

Data was collected from the AFC Fortis Escort Heart Institute in Khulna, Bangladesh, under the supervision of Cardiologist Dr. Arun R More from 27th April to 7th May 2016. A total of 506 patients' medical record files (MRD) with 152 features, including outcomes, have been gathered and shown in Table 1. Of these 152 risk factors, only 138 have been selected for this research study. Four types of data have been collected. Some are diagnostic, some are about the patient's history, some are demographic, and some are about symptoms shown in supplementary materials (Table 2: Features Description with Reference Value). About nine types of heart attacks have been found in the record, including ACS, STEMI, NSTEMI, Unstable Angina, UMI, AMI, OMI, AWMI, and IWMI, which leads to IHD and the average Heart Rate of the patients is 82.85.

Table 1. List of Features.

Features	Sub-category	Count	Features	Sub-category	Count	Features	Sub-category	Count
age	Less than 20	1	profession	business	66	height	136-145	5
	20-29	6		housewife	53		146-155	28
	30-39	31		labor-work	14		156-165	64
	40-49	85		retired	21		166-175	36
	50-59	160		service	30		176-185	11
	60-above	221		student	2		186-195	2
sex	Female	381		teacher	11	weight	45-54	20
	Male	125		unemployed	5		55-64	51
BMI	Above-Obese	14	bp-systolic	120-129	83		65-74	58
	Normal	70		130-139	72		75-84	18
	OverWeight	58		140-180	185		85-94	8
	UnderWeight	4		Above 180	16		115-124	1
HR	100-170	5		Less than 120	137	ACS	yes	353
	75-128	51	bp-diastolic	80-89	168		no	153
	78-132	38		90-120	204	IWMI	yes	21
	80-136	41		Above 120	1		no	485
	83-140	54		Less than 80	118	PAD	yes	1
	85-145	37	UMI	yes	1		no	505
	88-149	31		no	504	stable-angina	yes	22

Table 1. *Continued.*

Features	Sub-category	Count	Features	Sub-category	Count	Features	Sub-category	Count
	90-153	27	CSA	yes	10		no	484
	93-157	9		no	496	unstable-angina	yes	61
	95-162	8	DCM	yes	5		no	445
	Less than 75	201		no	501	CHB	yes	16
non-stemi	yes	70	ICM	yes	6		no	490
	no	436		no	500	severe-mitral-regurgitation	yes	1
AMI	yes	46	pericardial-effusion	yes	1		no	505
	no	460		no	505	severe-multiple-sclerosis	yes	3
dissecting-aortic-aneurysm	yes	1	anterior-ischaemia	yes	1		no	503
	no	505		no	505	atrial-fibrillation	yes	5
CKD	yes	9	AKI	yes	10		no	501
	no	497		no	496	ALVF	yes	31
HTN	yes	349	PCI	yes	17		no	475
	no	154		no	489	SVT	yes	2
vertigo	yes	3	rhenumated-arthitis	yes	10		no	504
	no	503		no	496	RHD	yes	1
IDCM	yes	1	CCF	yes	1		no	505
	no	505		no	505	OMI	yes	6
synptonatic-bradycardia	yes	3	ESRD	yes	1		no	500
	no	503		no	505			
FVR	yes	4	hemorrhagic-shock	yes	1	cardiogenic-shock	yes	26
	no	502		no	505		no	480
RTI	yes	2	aortic-stenosis	yes	1	CABG	yes	15
	no	504		no	505		no	491
survivor-cardiac-arrest	yes	3	VHD	yes	1	RVI	yes	1
	no	503		no	505		no	505

Table 1. *Continued.*

Features	Sub-category	Count	Features	Sub-category	Count	Features	Sub-category	Count
Pneumonia	yes	1	atrioventricular-block	yes	1	sinus-rhythm	yes	1
	no	505		no	505		no	505
normal-epicardial-coronary-artery	yes	11	DVD	yes	53	thrombolysed-stk	yes	21
	no	495		no	453		no	485
surviver-cardiacarrest	yes	1	ventricular-tachycardia	yes	3	Hypothyroidism	yes	1
	no	505		no	503		no	505
ventricular-bigeminy	yes	1	RBBB	yes	1	valvular-cardiomyopathy	yes	2
	no	505		no	505		no	504
TMT	yes	27	CAD	yes	56	SVD	yes	54
	no	479		no	450		no	452
TVD	yes	3	DVT	yes	2	AWMI	yes	15
	no	503		no	504		no	491
PTCA	yes	88	LAD	yes	1	congenital-RCAA-orta-fistula	yes	1
	no	418		no	505		no	505
congestive-hepatitis	yes	2	dyslipoproteinemia	yes	66	renal-artesy-stenosis	yes	3
	no	504		no	440		no	503
psychosis	yes	1	hypothyrodism	yes	1	normal-coronasies	yes	1
	no	505		no	505		no	505
mild-pr	yes	1	urosepsis	yes	1	LBBB	yes	4
	no	505		no	505		no	502
PPM	yes	4	anaemic-heart-failure	yes	1	COPD	yes	3
	no	502		no	505		no	503
electrolyte-imbalance	yes	1	ischemetic-hepataitis	yes	1	moderate-ms	yes	1
	no	505		no	505		no	505
metabolic-encephalopathy	yes	1	severe-musculoskeletal-pain	yes	1	lt-vertebral-artery-stenosis	yes	1
	no	505		no	505		no	505

Table 1. *Continued.*

Features	Sub-category	Count	Features	Sub-category	Count	Features	Sub-category	Count
subacute-intestinal-obstruction	yes	1	DKA	yes	1	PAH	yes	2
	no	505		no	505		no	504
hypokalemia	yes	1	dyselectrolytemia	yes	4	symptotic-nodal-bradycardia-with-stpm	yes	1
	no	505		no	502		no	505
asymptotic-nodal-bradycardia-with-stpm	yes	1	marfan-syndrome	yes	1	cough-variant-asthma	yes	1
	no	505		no	505		no	505
POBA	yes	1	hyponatremia	yes	1	bronchial-asthma	yes	4
	no	505		no	505		no	502
severe-anaemia	yes	1	LRT	yes	1	anaemia	yes	1
	no	505		no	505		no	505
UTI	yes	2	pleural-effusion	yes	1	psoriasis	yes	1
	no	504		no	505		no	505
lms-disease	yes	1	hypoxic-ischemic-encephalopathy	yes	1	renal-impairment	yes	3
	no	505		no	505		no	503
HBsAG	yes	2	CVD	yes	2	BEP	yes	2
	no	504		no	504		no	504
dyslipidemia	yes	1	family-history	yes	152	smoking	ex	63
	no	505		no	347		yes	173
tobaco	yes	108	DM	yes	222		no	267
	no	385		no	283	vomiting	yes	105
physical-activity	yes	24	psychological-stress	yes	108		no	401
	no	473		no	382	radiation	yes	38
stroke	yes	129	drug-history	yes	363		no	420
	no	377		no	115	HB	Less than 13	104
dyspnea	yes	205	palpitation	yes	49		Less than 13.6	239
	no	299		no	452		13-16	11

Table 1. *Continued.*

Features	Sub-category	Count	Features	Sub-category	Count	Features	Sub-category	Count
syncope	yes	24	sweating	yes	178		13.6-17.7	120
	no	480		no	328	EF	Borderline	116
chest-pain	yes	389	creatinine	Abnormal	239		Hypertrophic Cardiomyopathy	7
	no	114		Normal	178		Normal	246
RBS	Diabetic	103	troponin	Normal	85		Reduced	84
	Normal	271		Probable Heart Attack	113	cag-findings-type	CAGNotDone	348
ECG	abnormal	376	patients_status	discharged	468		DVD	51
	normal	96		expired	38		Normal-critical-coronary-arteries	31
CAG	yes	234	IHD (outcome)	yes	353		SVD	49
	no	255		no	153		TVD	27

3.2 Data Pre-processing

The data pre-processing can be divided into two steps: data categorization and the imputation process.

3.2.1 Data Categorization

In this work, 138 risk factors and 506 instances were used. Of these 138 risk factors, there were three categories: Demographic features (6), Diagnostic features (8), and Symptoms features (124).

3.2.2 Imputation Process

Missing values in a dataset can be handled in several ways. The KNN imputer with k = 5 neighbors was used to eliminate missing data. Modifying the classifier's K value KNN classifiers can produce different results. Following previous research [23] on KNN imputation, k = 5 was chosen as the best value for this investigation.

3.3 Feature Selection Process

Feature Selection is selecting the features relevant to a model's output. Information gain, Best First Search, SelectKBest, XGBoost, Linear Regression, Logistic Regression, Shapley Additive explanations, Boruta, and Wrapper machine learning methods are all interconnected and play a crucial role in predicting Ischemic Heart Disease (IHD) by helping identify the most significant factors contributing to the disease.

By selecting the most informative features, these methods reduce model complexity, improve accuracy, and prevent overfitting (when the model performs well on training data but poorly on unseen data). Techniques like SHAP help understand how different features contribute to the model's predictions, aiding in better decision-making and knowledge discovery. Focusing on relevant features allows models to learn more effectively from the data, leading to better predictions of IHD risk. These methods generally apply to any machine learning problem where feature selection is crucial. However, the specific choice of methods depends on the characteristics of the data and the desired model complexity. These machine-learning methods work together to create a more accurate and interpretable model for predicting Ischemic Heart Disease. By selecting the most informative features, these techniques can improve model performance and provide valuable insights into the factors that contribute to the disease.

3.3.1 Information gain (IG)

Information gain (IG) reduces entropy or surprise by transforming a dataset. It is measured based on the entropy of a system, i.e., the degree of disorder of the system. The Ranker algorithm has made use of the IG class [25]. It calculated the Information Gain of the characteristics to determine the relevance of each attribute. It aided in selecting the most linked qualities by assessing the information gained. Mathematically, as in (1):

$$\text{Information gain }(\text{IG}) = \text{entropy parent } - \left[\text{weighted average}\right] \text{ entropy children} \tag{1}$$

3.3.2 Best First Search (BFS)

The results of BFS use the concept of a priority queue and heuristic search [26]. A heuristic search algorithm works on a specific rule, with the aim of reaching the goal from the initial state via the shortest path. This work used the CfsSubsetEval function for BFS in the WEKA tool [43].

3.3.3 SelectKBest

The SelectKBest method selects the features according to the k highest score [27]. It takes a score function as a parameter, which must apply to a pair (X, y). The score function must return an array of scores, one for each feature X[:, i] of X. SelectKBest then retains the first k features of X with the highest score [28].

3.3.4 XGBoost

XGBoost is the dominant technique for predictive modeling on regular data. The gradient boosting algorithm is the top technique on a wide range of predictive modelling problems, and XGBoost is the fastest implementation [29]. A gradient boosting library provides a parallel boosting trees algorithm to solve Machine Learning tasks.

3.3.5 Linear Regression

The results A Linear Regression model predicts the target as a weighted sum of the feature inputs [30]. The linearity of the learned relationship makes the interpretation easy. Linear models can model the dependence of a regression target y on some features x. The intellectual relationships are linear and can be written for a single instance as in (2):

$$y = \beta_0 + \beta_1 x_1 + \ldots\ldots\ldots\ldots + \beta_p x_p + \varepsilon \tag{2}$$

The first weight in the sum (β_0) is called the intercept and is not multiplied with a feature. The epsilon (ϵ) is the difference between the prediction and the actual outcome.

3.3.6 Logistic Regression (LOR)

The logistic regression (LOR) model allows the dataset to fit and retrieve the coeff_ property that contains the coefficients found for each input variable [31]. These coefficients can provide the basis for a crude feature importance score. This assumes that the input variables have the same scale or have been scaled before fitting a model.

3.3.7 SHAP (Shapley Additive exPlanations)

The SHAP approach assigns the SHAP values, which are contribution values for a model's output for each feature of each data point [32]. These SHAP values encode the importance that a model gives to a feature. The mean of the columns of each matrix is calculated, and the vectors of mean SHAP values for each class are summed and ordered in a decreasing way. The first position of the resulting vector contains the essential feature, and the second position includes the second most important. Since SHAP can provide interpretability of a model's decisions by indicating the importance of the dataset features, a feature selection algorithm based on the essential attributes according to the absolute SHAP values would provide good results [44].

3.3.8 Boruta

Boruta algorithm is a wrapper built around the random forest classification algorithm implemented in the R package randomForest(). It is an ensemble method in which classification is performed by voting multiple unbiased weak classifiers—decision trees. These trees are independently developed on different bagging samples of the training set. The importance measure of an attribute is obtained as the loss of accuracy of classification caused by the random permutation of attribute values between objects [33]. It is computed separately for all trees in the forest which use a given attribute for classification. Then the average and standard deviation of the accuracy loss is calculated. Here, the extra randomness provides a clearer view of the essential features.

Consider the following parameters used in Boruta [45]:

1. maxRuns: The number of random forest runs that can be performed. Consider increasing this option if any provisional qualities remain. The default value is 100.
2. doTrace: This is a reference to the level of verbosity. 0 indicates that there will be no tracing. One shows that the attribute decision should be reported immediately upon completion. 2 entails everything in 1 plus reporting on each iteration. The default value is 0.
3. holdHistory: If set to TRUE, the complete history of crucial runs is retained (Default). When the plotImpHistory function is used, it returns a Classifier run vs. Importance plot.

3.3.9 Wrapper

The wrapper methods create several models with different subsets of input feature variables [34]. Later, the selected features result in the best-performing model according to the performance metric. The main idea behind a wrapper method is to detect the interaction between variables and search which set of features works best for a specific classifier, as shown in Figure 2.

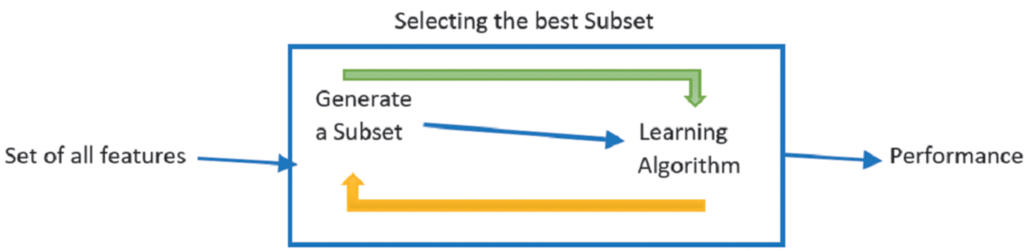


Figure 2. Wrapper Feature Selection Method Principle.

For the purpose of determining a threshold value, this study contained a wide variety of feature selection methods, all of which were performed. We have determined that the information gain threshold, the k-best threshold, and the wrapper approach threshold each have values of 0.01, 0.1, and 0.5, respectively. Negative correlation values, such as 61.14841 and 0.34807, are the only ones that are considered by the linear and logistic regression methods. The k-best algorithm concluded that the minimum threshold value should be 0.108375, while the highest threshold value should be 62.874272 on average. A correlation value of 0.7096016958870663 was selected using the wrapper technique. Twenty different characteristics have been selected using the SHAPly analysis. The F score computed by the XGBoost method is utilized to select the six attributes. The Boruta technique chooses attributes via a subset.

3.4 Feature Ranking

Selecting appropriate features to achieve the best result has been one of the most challenging topics. A combination of common and infrequent feature selection algorithms was used here, as mentioned above. From these ten different machine learning techniques, the threshold values were determined for each of them, and according to that value, a voting algorithm was proposed to rank the features. In this way, three feature categories were rated: High-ranked features, Medium-ranked features, and Low-ranked features.

3.5 Tools and Simulation Environment

Various statistical tools were utilized to compute the feature values. For example, Chi-square analysis was performed to determine the p-value using SPSS (Statistical Package for the Social Sciences) version 25.0 for Windows OS. WEKA tool version 3.8.4, the most recent stable version, was used to calculate Information Gain and BFS for feature selection. For feature selection using the SelectKBest class, the Python 3.9.5 version was utilized. The Scikit-learn API includes the SelectKBest class, extracting the best features from a given dataset. Python 3.9.5 was also used to implement the feature importance from the XGBoost model, Logistic Regression, and the Boruta method (Boruta packages).

Several different platforms were utilized in this article. Every instrument possesses a distinct collection of features and advantages that are exclusive to it. Python features a significant level of flexibility in terms of customization, and Weka is well-known for its capacity to implement a wide range of algorithms and visualization. For statistical analysis, SPSS is an excellent choice for statistical analysis because of its well-known user-friendly interface. If we choose the appropriate instrument for a particular technique, we may be able to attain increased efficiency. Despite being available on all platforms, there are a few feature selection strategies that are not inaccessible to everyone. A certain method can be the only one that is compatible with Weka or a Python script that has been modified in certain circumstances. It may be able to make use of pre-existing code or established procedures that have been utilized in a certain set of circumstances in the past in order to facilitate the adoption of these tools for new trials.

4. RESULTS AND DISCUSSION

The experimented analysis can be divided into feature orientations, ranking, and voting processes.

4.1 Features Orientation in Data

IHD is the term given to heart problems caused by narrowed heart arteries. When arteries are narrowed, less blood and oxygen reach the heart muscle, also called coronary artery disease and coronary heart disease. It can ultimately lead to a heart attack. In our analysis, we have collected 138 risk factors and 506 instances. Of these 138 risk factors, demographic features were 6; diagnostic features were 8, and symptoms features were 124. Of these 506 records, around 7.5% of patients expired, whereas about 89.5% died because of IHD. The detailed feature distributions can be shown in Table 1.

Table 1 lists the demographic features: age, sex, profession, height, weight, and BMI. This study has eight diagnostic features, shown in Table 1. The diagnostic features are HR, BP systolic, BP diastolic, HB, creatinine, RBS, troponin, and EF. The rest of the features are symptom features. In the feature table, most symptom features have two subcategories, such as the ACS group, which has two subcategories.

4.2 Outcome of before and after Imputation

Exploratory Data Analysis (EDA) is a data analysis technique used to identify and analyze overarching patterns in a given dataset. These patterns represent outliers and potentially unanticipated characteristics of the data. For our dataset, the EDA has been shown in Figure 3 and Figure 4. The rest of the plots are available in the supplementary section 2.

Figures 3 and 4 each illustrate the histogram, KDE plot, boxplot, and violin plot of HB (Hemoglobin) and EF (Ejection Fraction), respectively. KDE plots are also depicted in Figures 3 and 4. The graph shows a notable change before and after HB and EF imputation.

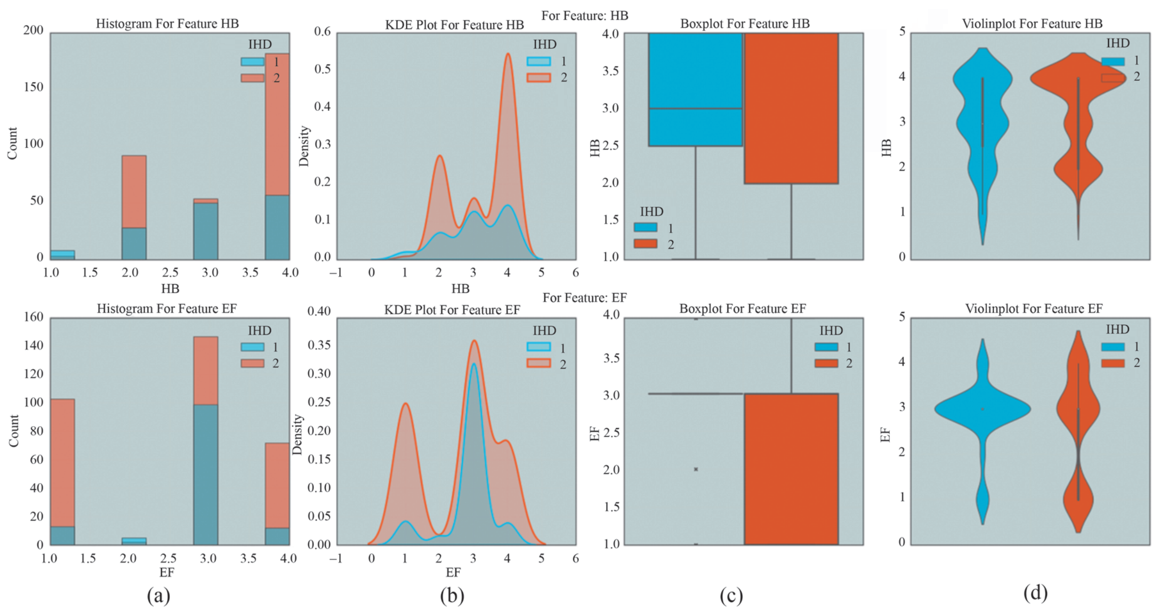
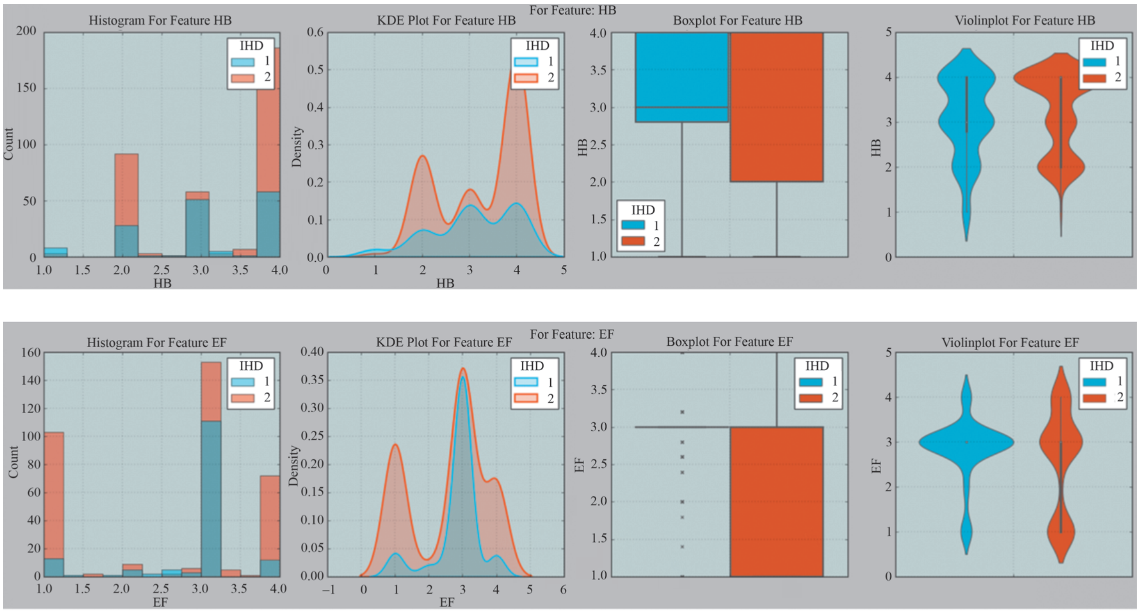


Figure 3. Before Imputation of HB and EF (a) Histogram (b) Kernel Density Estimate (KDE) (c) Boxplot and (d) Violin plot.



(a) (b) (c) (d)

Figure 4. After Imputation of HB and EF (a) Histogram (b) Kernel Density Estimate (KDE) (c) Boxplot and (d) Violin plot.

4.3 Features Selection Outcome

Table 2 displays an overview of all feature selection strategies results and the threshold value. The supplementary sections contain the comprehensive results.

Around 20 features (ACS, Sex, troponin, CAG, chest pain, ECG, non-stemi, EF, Stroke, PTCA, unstable angina, cag findings type, height, HR, HB, AMI, DVD, Vomiting, palpitation, sweating) have been selected by the SHAP value shown in Figure 5.

4.4 Feature Ranking Process and Voting Algorithm

4.4.1 Feature Ranking

The ranking of the features for this dataset was accomplished through the utilization of the most generally utilized and acceptable methodologies. Each of the nine machine-learning algorithms had their threshold values computed, and then a voting mechanism was implemented in order to rank the features in accordance with those threshold values. For example, if the ACS features are presented in eight algorithms but not in one of the nine feature selection algorithms, then the vote value will be eight. The procedure for grading consisted of classifying the qualities into three separate groups: features that were ranked high, features that were ranked medium, and lowest.

Table 2. Selected Features in Different Selection Techniques.

Method	Selected Features	Feature Name	Significant Value
IG	138	ACS, cag-findings-type, non-stemi, PTCA, EF, DVD, chest-pain, AMI, CAD, SVD, sex, CAG, HB, ECG, rhenumated-arthitis, HR, sweating, bp-systolic, unstable-angina, smoking, troponin, stroke, IWMI, thrombolysed-stk, PCI, age, AWMI, vomiting, bronchial-asthma, patients_status, normal-epicardial-coronary-artery, cardiogenic-shock, radiation, severe-multiple-sclerosis, palpitation, syncope, bp-diastolic, DCM, atrial-fibrillation, CKD, PAH, RTI, valvular-cardiomyopathy, UTI, SVT, RBS, OMI, ICM, profession, tobacco, ALVF, FVR, PPM, RHD, asymptotic-nodal-bradycardia-with-stpm, severe-musculoskeletal-pain, marfan-syndrome, DKA, VHD, aortic-stenosis, metabolic-encephalopathy, cough-variant-asthma, subacute-intestinal-obstruction, PAD, hypoxic-ischemic-encephalopathy, pleural-effusion, severe-anaemia, dissecting-aortic-aneurysm, hyponatremia, pericardial-effusio, hypothyroidism, hemorrhagic-shock, hypothyroidism, urosepsis, electrolyte-imbalance, mild-pr, moderate-ms, normal-coronasies, anaemic-heart-failure, renal-artesy-stenosis, ventricular-tachycardia, renal-impairment, dyslipoproteinemia, TMT, AKI, synptonatic-bradycardia, DM, weight, CVD, BEP, HBsAG, CHB, BMI, HTN, drug-history, CABG, UMI, POBA, anterior-ischaemia, lt-vertebral-artery-stenosis, severe-mitral-regurgitation, LAD, ventricular-bigeminy, lms-disease, RBBB, psoriasis, anaemia, ischemetic-hepataitis, LRT, surviver-cardiacarrest, pneumonia, IDCM, atrioventricular-block, sinus-rhythm, ESRD, symptotic-nodal-bradycardia-with-stpm, RVI, hypokalemia, CCF, psychosis, dyslipidemia, congenital-RCAA-orta-fistula, height, LBBB, stable-angina, physical-activity, CSA, congestive-hepatitis, DVT, dyselectrolytemia, creatinine, vertigo, COPD, survivor-cardiac-arrest, TVD, family-history, psychological-stress, dyspnea.	highest IG value = 0.88417056 and lowest IG value = 0.00000169
BFS	18	Sex, bp systolic, ACS, IWMI, severe-multiple-sclerosis, non-stemi, AMI, PCI, rhenumated-arthitis, thrombolysed-stk, valvular-cardiomyopathy, AWMI, bronchial-asthma, stroke, chest-pain, EF, ECG, cag findings type	-
SelectKBest	44	ACS, cag-findings-type, troponin, CAG, PTCA, ECG, chest-pain, non-stemi, sex, sweating, height, EF, stroke, DVD, CAD, unstable-angina, SVD, profession, AMI, smoking, vomiting, HB, age, tobaco, palpitation, RBS, bp-systolic, patients_status, rhenumated-arthitis, radiation, BMI, bp diastolic, IWMI, thrombolysed-stk, cardiogenic-shock, syncope, DM, PCI, dyslipoproteinemia, normal-epicardial-coronary-artery, AWMI, ALVF, HR, HTN, drug-history	0.1
XGBoost	6	ACS, HB, CAG, Profession, Troponin, Height.	-
LIR	5	DCM, SVT, cardiogenic shock, DVT, psoriasis.	-

Table 2. *Continued.*

Method	Selected Features	Feature Name	Significant Value
LOR	61	height, BMI, bp-systolic, ACS, IWMI, CSA, DCM, non-stemi, ICM, AMI, CKD, AKI, RHD, ESRD, OMI, CABG, VHD, DVD, RBBB, CAD, SVD, TVD, DM, RBS, troponin, ECG, dissecting-aortic-aneurysm, atrial-fibrillation, rhenumated-arthitis, atrioventricular-block, thrombolysed-stk, ventricular-tachycardia, hypothyroidism, valvular-cardiomyopathy, DVT, AWMI, PTCA, LAD, congenital-RCAA-orta-fistula, congestive-hepatitis, normal-coronasies, mild-pr, LBBB, PPM, electrolyte-imbalance, metabolic-encephalopathy, severe-musculoskeletal-pain, lt-vertebral-artery-stenosis, PAH, hypokalemia, asymptotic-nodal-bradycardia-with-stpm, cough-variant-asthma, hypoxic-ischemic-encephalopathy, renal-impairment, dyslipidemia, smoking, psychological-stress, stroke, palpitation, syncope, sweating, radiation.	-
SHAP	20	ACS, Sex, troponin, CAG, chest-pain, ECG, non-stemi, EF, Stroke, PTCA, unstable angina, cag findings type, height, HR, HB, AMI, DVD, Vomiting, palpitation, sweating	-
Boruta	33	age, sex, profession, height, weight, BMI, HR, bp systolic, ACS, IWMI, unstable-angina, non-stemi, AMI, rhenumated-arthitis, DVD, thrombolysed-stk, CAD, SVD, PTCA, smoking, stroke, sweating, vomiting, radiation, chest-pain, HB, RBS, troponin, EF, ECG, CAG, cag findings type, patients' status	-
Wrapper	135	ACS, cag-findings-type, non-stemi, PTCA, EF, DVD, chest-pain, AMI, CAD, SVD, CAG, HB, ECG, rhenumated-arthitis, HR, sweating, bp-systolic, unstable-angina, smoking, troponin, stroke, IWMI, thrombolysed-stk, PCI, AWMI, vomiting, bronchial-asthma, patients_status, normal-epicardial-coronary-artery, cardiogenic-shock, radiation, severe-multiple-sclerosis, palpitation, syncope, bp-diastolic, DCM, atrial-fibrillation, CKD, PAH, RTI, valvular-cardiomyopathy, UTI, SVT, RBS, OMI, ICM, profession, tobacco, ALVF, FVR, PPM, RHD, asymptotic-nodal-bradycardia-with-stpm, severe-musculoskeletal-pain, marfan-syndrome, DKA, VHD, aortic-stenosis, metabolic-encephalopathy, cough-variant-asthma, subacute-intestinal-obstruction, PAD, hypoxic-ischemic-encephalopathy, pleural-effusion, severe-anaemia, dissecting-aortic-aneurysm, hyponatremia, pericardial-effusio, hypothyroidism, hemorrhagic-shock, hypothyroidism, urosepsis, electrolyte-imbalance, mild-pr, moderate-ms, normal-coronasies, anaemic-heart-failure, renal-artesy-stenosis, ventricular-tachycardia, renal-impairment, dyslipoproteinemia, TMT, AKI, synptonatic-bradycardia, DM, CVD, BEP, HBsAG, CHB, BMI, HTN, drug-history, CABG, UMI, POBA, anterior-ischaemia, lt-vertebral-artery-stenosis, severe-mitral-regurgitation, LAD, ventricular-bigeminy, lms-disease, RBBB, psoriasis, anaemia, ischemetic-hepataitis, LRT, surviver-cardiacarrest, pneumonia, IDCM, atrioventricular-block, sinus-rhythm, ESRD, symptotic-nodal-bradycardia-with-stpm, RVI, hypokalemia, CCF, psychosis, dyslipidemia, congenital-RCAA-orta-fistula, height, LBBB, stable-angina, physical-activity, CSA, congestive-hepatitis, DVT, dyselectrolytemia, creatinine, vertigo, COPD, survivor-cardiac-arrest, TVD, family-history, psychological-stress, dyspnea.	0.5

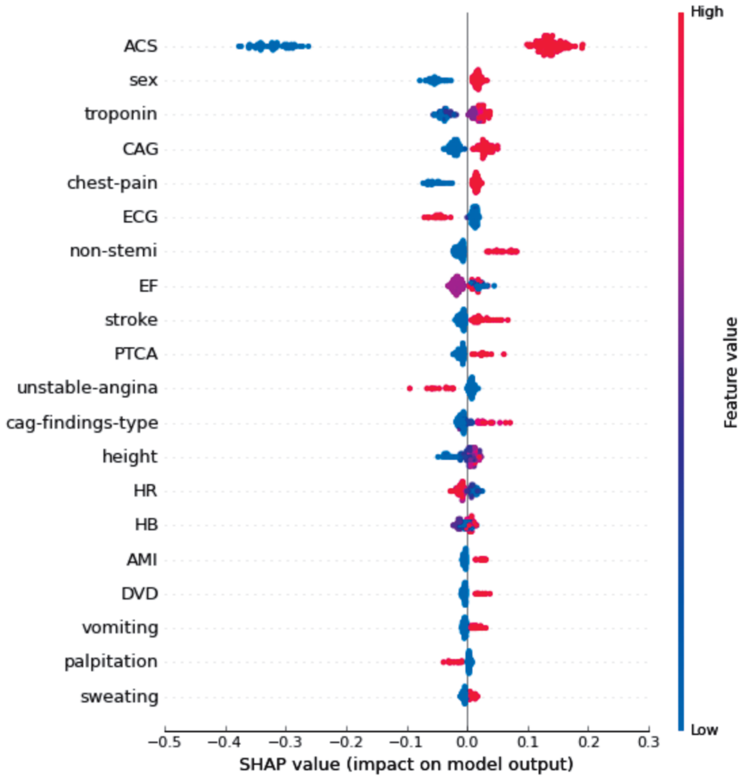


Figure 5. SHAP Analysis Outcome.

4.4.2 Voting Algorithm

We have proposed a voting algorithm, shown in Figure 6. After applying the voting algorithm, three categories are yielded: high, medium, and low.

Pseudocode of the Voting Algorithm:

```
FOR EACH POSITIVE INTEGER  VALUE OF THE VOTE
        DO
                IF THE VOTE VALUE IS GREATER THAN EQUAL TO 6 AND LESS THAN EQUAL
                TO 8 THEN
                        HIGH FEATURES
                IF THE VOTE VALUE IS GREATER THAN OR EQUAL TO 3 AND LESS THAN 5

                THEN
                        MEDIUM FEATURES
                ELSE
                        LOW FEATURES
END LOOP
```

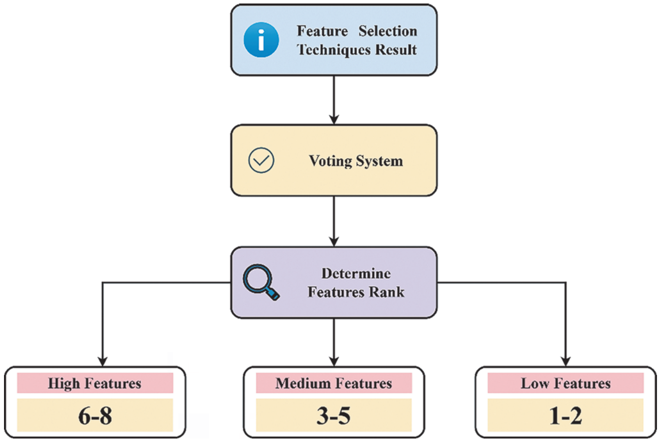


Figure 6. Voting Algorithm Process.

4.4.2.1 High Ranked Features

The 19 features were chosen as being of excellent quality. Table 3 lists the characteristics and the voting value. Each attribute in every ranking model has a different voting value and the worth of other machine learning techniques. The top-rated feature is blood pressure systolic, with a voting score of 6 out of 8. This voting value is considered IG, BFS, K-BEST, LOR, Boruta, and Wrapper, where all technique values are similar, and it is 1. However, XGBoost, LOR, and SHAP have not been voted for this feature because of the threshold value.

Table 3. Voting Value of High Ranked Features.

Feature	Vote	Feature	Vote
Height	6	Stroke	7
bp_systolic	6	Sweating	6
ACS	8	Chest_pain	6
IWMI	6	HB	6
Non_stemi	7	Troponin	7
AMI	7	EF	6
rheumatoid-arthritis	6	ECG	7
DVD	6	CAG	6
thrombolysed-stk	6	Cag-findings-type	6
PTCA	6	-	-

4.4.2.2 Medium Ranked Features

25 features were selected as medium features. The voting value of these features is shown in the supplementary material. The voting ranges of these features are from 3 to 5.

4.4.2.3 Low Ranked Features

There are 94 features selected as medium features. The low-ranking attribute with voting value has been provided in the supplementary material. The voting ranges of these features are from 1 to 2.

5. DISCUSSIONS

The features are rated using ten different machine learning algorithms, and the ultimate ranking is determined using the voting algorithm provided in this work. Three categories have been chosen for ranking. High-ranking features have 19 features, medium-ranking features have 25, and low-ranking features have 94. The voting system reveals features with high, medium, and low rankings. Table 4 presents the ratios between the high-ranked features subcategories and IHD.

Table 4. Distribution of High Ranked Features VS IHD.

Features			Class (IHD) Ratios (Probabability)	
Name	Sub Category		Class (NO)	Class (YES)
height	136-145	% within height	20.0%	80.0%
		% within IHD	0.7%	1.1%
	146-155	% within height	35.7%	64.3%
		% within IHD	6.5%	5.1%
	156-165	% within height	28.1%	71.9%
		% within IHD	11.8%	13.0%
	166-175	% within height	22.2%	77.8%
		% within IHD	5.2%	7.9%
	176-185	% within height	18.2%	81.1%
		% within IHD	1.3%	2.5%
	186-195	% within height	50.0%	50.0%
		% within IHD	0.7%	0.3%

Table 4. *Continued.*

Features			Class (IHD) Ratios (Probabability)	
Name	**Sub Category**		**Class (NO)**	**Class (YES)**
bp-systolic	120-129	% within bpsystolic	20.5%	79.5%
		% within IHD	11.1%	18.7%
	130-139	% within bpsystolic	36.1%	63.9%
		% within IHD	17.0%	13.0%
	140-180	% within bpsystolic	38.4%	61.6%
		% within IHD	46.4%	32.3%
	Above 180	% within bpsystolic	50.0%	50.0%
		% within IHD	5.2%	2.3%
	Less than 120	% within bpsystolic	20.4%	79.6%
		% within IHD	18.3%	30.9%
		% within IHD	19.4%	24.9%
ACS	no	% within ACS	100.0%	0.0%
		% within IHD	100.0%	0.0%
	yes	% within ACS	0.0%	100.0%
		% within IHD	0.0%	100.0%
IWMI	no	% within IWMI	31.5%	68.5%
		% within IHD	100.0%	94.1%
	yes	% within IWMI	0.0%	100.0%
		% within IHD	0.0%	5.9%
non-stemi	no	% within non-stemi	35.1%	64.9%
		% within IHD	100.0%	80.2%
	yes	% within non-stemi	0.0%	100.0%
		% within IHD	0.0%	19.8%
AMI	no	% within AMI	33.3%	66.7%
		% within IHD	100.0%	87.0%

Table 4. *Continued.*

Features			Class (IHD) Ratios (Probabability)	
Name	**Sub Category**		**Class (NO)**	**Class (YES)**
	yes	% within AMI	0.0%	100.0%
		% within IHD	0.0%	13.0%
rhenumated-arthitis	no	% within rhenumated-arthitis	28.8%	71.2%
		% within IHD	93.5%	100.0%
	yes	% within rhenumated-arthitis	100.0%	0.0%
		% within IHD	6.5%	0.0%
DVD	no	% within DVD	33.8%	66.2%
		% within IHD	100.0%	85.0%
	yes	% within DVD	0.0%	100.0%
		% within IHD	0.0%	15.0%
thrombolysed-stk	no	% within thrombolysed-stk	31.5%	68.5%
		% within IHD	100.0%	94.1%
	yes	% within thrombolysed-stk	0.0%	100.0%
		% within IHD	0.0%	5.9%
PTCA	no	% within PTCA	35.9%	64.1%
		% within IHD	98.0%	75.9%
	yes	% within PTCA	3.4%	96.6%
		% within IHD	2.0%	24.1%
stroke	no	% within stroke	35.0%	65.0%
		% within IHD	86.3%	69.4%
	yes	% within stroke	16.3%	83.7%
		% within IHD	13.7%	30.6%
sweating	no	% within sweating	36.9%	63.1%
		% within IHD	79.1%	58.6%

Table 4. *Continued.*

Features			Class (IHD) Ratios (Probabability)	
Name	**Sub Category**		**Class (NO)**	**Class (YES)**
	yes	% within sweating	18.0%	82.0%
		% within IHD	20.9%	41.4%
chest-pain	no	% within chest-pain	53.5%	46.5%
		% within IHD	39.9%	10.5%
	yes	% within chest-pain	23.4%	76.6%
		% within IHD	59.5%	84.4%
HB	13-16	% within HB	72.7%	27.3%
		% within IHD	5.2%	0.8%
	13.6-17.7	% within HB	23.3%	76.7%
		% within IHD	18.3%	26.1%
	Less than 13	% within HB	48.1%	51.9%
		% within IHD	32.7%	15.3%
	Less than 13.6	% within HB	23.8%	76.2%
		% within IHD	37.3%	51.6%
troponin	Normal	% within troponin	57.6%	42.4%
		% within IHD	32.0%	10.2%
	Probable Heart Attack	% within troponin	13.3%	86.7%
		% within IHD	9.8%	27.8%
EF	Borderline	% within EF	11.2%	88.8%
		% within IHD	8.5%	29.2%
	Hypertrophic Cardiomyopathy	% within EF	71.4%	28.6%
		% within IHD	3.3%	0.6%
	Normal	% within EF	40.2%	59.8%
		% within IHD	64.7%	41.6%
	Reduced	% within EF	14.3%	85.7%
		% within IHD	7.8%	20.4%

Table 4. *Continued.*

Features			Class (IHD) Ratios (Probabability)	
Name	**Sub Category**		**Class (NO)**	**Class (YES)**
ECG	abnormal	% within ECG	25.0%	75.0%
		% within IHD	61.4%	79.9%
	normal	% within ECG	54.2%	45.8%
		% within IHD	34.0%	12.5%
CAG	no	% within CAG	40.8%	59.2%
		% within IHD	68.0%	42.8%
	yes	% within CAG	17.9%	82.1%
		% within IHD	27.5%	54.4%
cag-findings-type	CA-NotDone	% within cag-findings-type	37.1%	62.9%
		% within IHD	84.3%	62.0%
	DVD	% within cag-findings-type	0.0%	100.0%
		% within IHD	0.0%	14.4%
	Normal-critical-coronary-arteries	% within cag-findings-type	74.2%	25.8%
		% within IHD	15.0%	2.3%
	SVD	% within cag-findings-type	0.0%	100.0%
		% within IHD	0.0%	13.9%
	TVD	% within cag-findings-type	3.7%	96.3%
		% within IHD	0.7%	7.4%

The voting algorithm has identified 19 significant features in this research. That is height ($p<0.779$), bp-systolic ($p<0.0001$), ACS ($p<0.0001$), IWMI ($p<0.002$), non-stemi ($p<0.0001$), AMI ($p<0.0001$), rhenumated-arthitis ($p<0.0001$), DVD ($p<0.0001$), thrombolysed-stk ($p<0.002$), PTCA ($p<0.0001$), stroke ($p<0.0001$), sweating ($p<0.0001$), chest-pain ($p<0.0001$), HB ($p<0.0001$), troponin ($p<0.0001$), EF ($p<0.0001$), ECG ($p<0.0001$), CAG ($p<0.0001$), cag-findings-type ($p<0.0001$).

The high-ranked feature's distribution with IHD (outcome) is displayed in Table 4. The BMI is associated with the prominent features related to height and weight. Conversely, the electrocardiogram (ECG) results presented in Table 4 encompass two distinct aspects: normal and abnormal. An ECG abnormal subgroup is considered effective for heart disease when it falls within the range of 61.4% (IHD ratio class "no") and

79.9% (IHD ratio class “yes”). In accordance with the diagnostic outcome of troponin characteristics, troponin may be classified into two subcategories: normal and possibly heart attack. If the troponin ratio for those with an IHD ratio class of “no” is 9.8% and for those with an IHD ratio class of “yes” is 27.8%, then this feature is considered crucial. It is important to emphasize that stroke is a binary condition. The stroke has a ratio of 16.3% (no) and 83.7% (yes). Similarly, Table 3 presents the distribution of other highly ranked traits with IHD. Please refer to the supplemental material for the remaining values of P, IG, BFS, K-Best, XGB, LIR, LOR, SHAP, Boruta, and Wrapper.

6. LIMITATIONS

In this study, the data were primarily collected from AFC Fortis Escorts Heart Institute of the southeast region- Khulna, instead of covering the majority of the hospitals of the southwest zone. Due to the permission issue from an authority, it was missing test records for tests that were done in other places, the lack of computerized records or under-represented medical data, and presumably, the underdeveloped systems for data reporting and collecting in those hospitals. Although this one institute represented the overall IHD situation of the southern region of Bangladesh, the unavailability of other hospital data might limit some of our conclusions about some other groups and reduce our coverage. Consequently, there was enormous uncertainty as only one sample was taken instead of taking a count across the entire population of the southern region. This might also be the reason for having a seemingly small sample size. Besides, some missing values were in the sample data since real-world data was presented.

7. CONCLUSIONS

The American Heart Association reports a rise in the occurrences of IHD during the past twenty years. It continues to be a prominent factor contributing to mortality in both industrialized and developing nations. Nevertheless, there is a significant worry over the high mortality rates of IHD in several nations, particularly those falling within the lower and middle-income categories. This study aimed to assess the risk factors associated with the development of IHD and categorize them based on their severity level in the southern region of Bangladesh, a middle-income nation. Several machine learning feature selection approaches were employed to identify the notable characteristics or risk factors of IHD, which were then classified into three distinct categories. Among these 138 risk factors, we categorized them as High, Medium, and Low based on their respective implications. As implied by their nomenclature, the 19 high-risk factors are undeniably crucial and necessitate prompt attention. Identifying these high-risk characteristics might potentially reduce the number of clinical tests required for patients. The 25 moderate risk factors should be classified as ‘urgent’ as they have the potential to escalate if not promptly addressed. Despite their modest risk, routine clinical check-ups are necessary for the remaining 94 risk factors. The application of machine learning and statistical analysis in this diagnostic can greatly benefit the patient. It decreases the expense of medical examinations for a reduced number of required tests, affecting IHD. Therefore, it is important to provide due consideration to this matter in therapeutic treatments, not only for this particular location but also from a worldwide standpoint, and therefore, prompt attention should be given. In order to enhance the efficacy of the model, it is imperative to augment the number of instances inside the dataset. Data preparation techniques such as Regression Analysis, Bayesian Belief Network, and Naive Bayes are commonly employed in many domains. The implementation will involve the utilization of many sophisticated Ensemble Learning techniques, including Radial Basis Neural Network (RBNN), Time Domain Neural Network (TDNN), Convolution Neural Network (CNN), and Support Vector Machine (SVM) with multi-class and binary Kernel. Ultimately, an expert system will be constructed utilizing the most optimal model derived from the aforementioned research investigation.

8. DATA AVAILABILITY STATEMENT

AFC Fortis Escorts Heart Institute has provided the data for this research and some other research projects as mentioned in the MoU agreement. The dataset has been completely de-identified. Requests to access the dataset should be directed to *https://github.com/mraihan/IHD_Research_Paper_Supplementary_Materials/blob/main/SelectedFeatures.csv* for authors' approval.

9. ETHICS STATEMENT

This research study is ethically approved under a memorandum of understanding (MoU) agreement (*https://github.com/mraihan/IHD_Research_Paper_Supplementary_Materials/blob/main/agreement.pdf*). Per the MoU, the dataset can be used in any project as long as collaborators from Khulna University and AFC Fortis Institute consent or are involved.

10. AUTHOR CONTRIBUTIONS

M. Raihan: conceptualization

Sami Azam: methodology

Laboni Akter: writing original draft

Md. Mehedi Hassan: conceptualization

Ryana Quadir: writing review and editing

Asif Karim: writing review and editing

Saikat Mondal: data curation

Arun More: data curation

11. FUNDING

There is no funding for this research.

12. CONFLICTS OF INTEREST

The authors have no conflict of interest.

13. ACKNOWLEDGMENTS

We want to be grateful to Computer Science and Engineering (CSE) Discipline, Khulna University, Khulna, Bangladesh, AFC Fortis Escorts Heart Institute, Khulna, Bangladesh, and Rural Health Progress Trust (RHPT), Murud, Latur, India, for providing us technological and clinical support. During data collection, (the founder of RHPT, Cardiologist) Dr. Arun More, and (faculty member of CSE, Khulna University) Saikat Mondal provided their valuable knowledge and guidelines. They helped us evaluate the clinical report efficiently, which aided us in this research study.

14. SUPPLEMENTARY MATERIALS

A glossary of terms, feature selection results, feature reference value table, voting value, and features sub-category distribution tables have been provided. Each feature's plot of the before and after imputation and XGBoost outcomes is provided. The supplementary file can be downloaded from https://www.scidb.cn/en/s/NFJje2.

REFERENCES

[1] Who international report: Cardiovascular diseases (CVDs). Available at: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (2022) Accessed 9 May 2022

[2] Virani S., Alonso A., Aparicio H., Benjamin E., Bittencourt M., Callaway C. et al.: Heart Disease and Stroke Statistics-2021 Update. Circulation 143(8), (2021)

[3] Criteria I: Ischemic Heart Disease. Ncbi.nlm.nih.gov Available at: https://www.ncbi.nlm.nih.gov/books/NBK209964/ (2022) Accessed 9 May 2022

[4] Wang F., Yu Y., Mubarik S., Zhang Y., Liu X., Cheng Y. et al.: Global Burden of Ischemic Heart Disease and Attributable Risk Factors, 1990-2017: A Secondary Analysis Based on the Global Burden of Disease Study 2017. Clinical Epidemiology, Volume 13: 859–870 (2021)

[5] Centers for Disease Control and Prevention: Heart Disease in the United States, 1999-2018. CDC WONDER Online Database. Atlanta, GA: Centers for Disease Control and Prevention, (2018), Accessed March 12 2021

[6] Australian Bureau of Statistics 2020: Causes of Death 2019, cat. no. 3303.0, October (2020)

[7] Australian Institute of Health and Welfare 2020: Key Statistics: Cardiovascular Disease, National Hospital Morbidity Database (NHMD), (2020)

[8] British Heart Foundation: Heart & Circulatory Disease Statistics 2021, published in April (2021), Accessed December 2021

[9] Moran A., Vedanthan R.: Cardiovascular Disease Prevention in South Asia: Gathering the Evidence. Global Heart 8(2): 139 (2013)

[10] Islam A., Majumder A.: Coronary artery disease in Bangladesh: A review. Indian Heart Journal, 65(4): 424–435 (2013)

[11] Coronary Heart Disease in Bangladesh: World Life Expectancy. Available at: https://www.worldlifeexpectancy.com/bangladesh-coronary-heart-disease (2022) Accessed 9 May 2022

[12] Saquib N., Saquib J., Ahmed T., Khanam M., Cullen M.: Cardiovascular diseases and Type 2 Diabetes in Bangladesh: A systematic review and meta-analysis of studies between 1995 and 2010. BMC Public Health, 12(1), (2012)

[13] Misra A., Misra R., Wijesuriya M., Banerjee D.: The metabolic syndrome in South Asians: Continuing escalation & possible solutions. Indian Journal of Medical Research, 125(3): 345–54 Mar 1 (2007)

[14] Wang H.: Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet (2020)

[15] Hanif A.A., Hasan M., Khan M.S., Hossain M.M., Shamim A.A., Hossaine M., Ullah M.A., Sarker S.K., Rahman S.M., Bulbul M.M., Mitra D.K.: Ten-years cardiovascular risk among Bangladeshi population using non-laboratory-based risk chart of the World Health Organization: Findings from a nationally representative survey. PLoS ONE, 16(5): e0251967- May 26 (2021)

[16] Elosúa, R., Lluís-Ganella, C., Subirana, I., Havulinna, A.S., Läll, K., Lucas, G.J., Sayols-Baixeras, S., Pietilä, A., Alver, M., Cabrera de León, A., Sentí, M., Siscovick, D.S., Mellander, O., Fischer, K., Salomaa, V.V., & Marrugat, J.: Cardiovascular risk factors and ischemic heart disease: is the confluence of risk factors greater than the parts? A genetic approach. Circulation: Cardiovascular Genetics. 9(3): 279-86 June (2016)

[17] Alexeeff S.E., Liao N.S., Liu X., Van Den Eeden S.K., Sidney S.: Long, Äêterm PM2. 5 exposure and risks of ischemic heart disease and stroke events: Review and Meta, Äêanalysis. Journal of the American Heart Association, 10(1): e016890 Jan 5 (2021)

[18] Benhar H., Idri A., Hosni M.: Impact of Threshold Values for Filter-based Univariate Feature Selection in Heart Disease Classification. In HEALTH INF, pp. 391–398 (2020)

[19] Abdollahi J., Nouri-Moghaddam B.: Feature selection for medical diagnosis: Evaluation for using a hybrid Stacked-Genetic approach in the diagnosis of heart disease. arXiv preprint arXiv:2103.08175 Mar 15 (2021)

[20] Pouriyeh S., Vahid S., Sannino G., De Pietro G., Arabnia H., Gutierrez J.: A comprehensive investigation and comparison of machine learning techniques in the domain of heart disease. In 2017 IEEE symposium on computers and communications (ISCC) pp. 204–207 IEEE July 3 (2017)

[21] Nouri-Moghaddam B., Ghazanfari M., Fathian M.: A novel multi-objective forest optimization algorithm for wrapper feature selection. Expert Systems with Applications, 175: 114737 Aug 1 (2021)

[22] Ramesh G., Madhavi K., Dileep Kumar Reddy P., Somasekar J., Tan J.: Improving the accuracy of heart attack risk prediction based on information gain feature selection technique Materials Today: Proceedings (2021)

[23] Khalid N.E., Ibrahim S., Haniff P.N.: MRI Brain Abnormalities Segmentation using K-Nearest Neighbors (k-NN). International Journal on Computer Science and Engineering, 3(2): 980–90 Feb (2011)

[24] Elgamal Z.M., Yasin N.B., Tubishat M., Alswaitti M., Mirjalili S.: An improved harris hawks optimization algorithm with simulated annealing for feature selection in the medical field. IEEE Access, 8: 186638–52 Oct 8 (2020)

[25] Ramesh G., Madhavi K., Reddy P.D., Somasekar J., Tan J.: Improving the accuracy of heart attack risk prediction based on information gain feature selection technique. Materials Today: Proceedings, Feb 19 (2021)
[26] Ramalingam V.V., Dandapath A., Raja M.K.: Heart disease prediction using machine learning techniques: a survey. International Journal of Engineering & Technology. 7(2.8): 684–7 Oct (2018)
[27] Desyani T., Saifudin A., Yulianti Y.: Feature Selection Based on Naive Bayes for Caesarean Section Prediction. InIOP Conference Series: Materials Science and Engineering (Vol. 879, No. 1, p. 012091). IOP Publishing Jul 1 (2020)
[28] Senan E.M., Abunadi I., Jadhav M.E., Fati S.M.: Score and Correlation Coefficient-Based Feature Selection for Predicting Heart Failure Diagnosis by Using Machine Learning Algorithms. Computational and Mathematical Methods in Medicine, Dec 20 (2021)
[29] Huang Y.C., Li S.J., Chen M., Lee T.S., Chien Y.N.: Machine-Learning Techniques for Feature Selection and Prediction of Mortality in Elderly CABG Patients. InHealthcare (Vol. 9, No. 5, p. 547). Multidisciplinary Digital Publishing Institute, May (2021)
[30] Singh K.P., Basant N., Gupta S.: Support vector machines in water quality management. Analytica chimica acta. 703(2): 152–62 Oct 10 (2011)
[31] Kakade A., Kumari B., Dholaniya P.S.: Feature selection using logistic regression in case-control DNA methylation data of Parkinson's disease: A comparative study. Journal of Theoretical Biology, 457: 14–8 Nov 14 (2018)
[32] Marcílio, W. E., Eler D.M.: From explanations to feature selection: assessing SHAP values as feature selection mechanism. In 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) (pp. 340-347) IEEE Nov 7 (2020)
[33] Kursa M.B., Rudnicki W.R.: Feature selection with the Boruta package. J Stat Softw 36: 1–13 (2010) doi:10.18637/jss.v036.i11
[34] Romalt A.A., Kumar R.M.: An Analysis on Feature Selection Methods, Clustering and Classification used in Heart Disease Prediction-A Machine Learning Approach. J. Crit. Rev. 7(6): 138–42 (2020)
[35] Hu B., Li Y., Wang G., Zhang Y.: The Blood Gene Expression Signature for Kawasaki Disease in Children Identified with Advanced Feature Selection Methods. BioMed research international. June 28 (2020)
[36] Morrison M.L., Sands A.J., McCusker C.G., McKeown P.P., McMahon M., Gordon J., Grant B., Craig B.G., Casey F.A.: Exercise training improves activity in adolescents with congenital heart disease. Heart. 99(15): 1122–8 Aug 1 (2013)
[37] Şahin, D.Ö., Kural, O.E., Akleylek, S. et al. A novel permission-based Android malware detection system using feature selection based on linear regression. Neural Comput & Applic 35, 4903–4918 (2023)
[38] Awal M.A., Masud M., Hossain M.S., Bulbul A.A., Mahmud S.H., Bairagi A.K.: A novel Bayesian optimization-based machine learning framework for COVID-19 detection from inpatient facility data. IEEE Access. 9: 10263–81 Jan 11 (2021)
[39] Ahmadpour H., Bazrafshan O., Rafiei-Sardooi E., Zamani H., Panagopoulos T.: Gully Erosion Susceptibility Assessment in the Kondoran Watershed Using Machine Learning Algorithms and the Boruta Feature Selection. Sustainability. 13(18): 10110 Jan (2021)
[40] Mafarja M., Mirjalili S.: Whale optimization approaches for wrapper feature selection. Applied Soft Computing. 62: 441–53, Jan 1 (2018)
[41] Wang K., Tian J., Zheng C., Yang H., Ren J., Liu Y., Han Q., Zhang Y.: Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. Computers in Biology and Medicine 137: 104813 Oct 1 (2021)

[42] Witten I., Frank E., Hall M., and Pal C.: Data Mining, 4th ed. San Diego: Elsevier Science & Technology Books (2017)

[43] Wang D, Thunéll S, Lindberg U, Jiang L, Trygg J, Tysklind M. : Towards better process management in wastewater treatment plants: Process analytics based on SHAP values for tree-based machine learning methods. Journal of Environmental Management 301:113941 Jan 1 (2022)

[44] Qu J., Ren K., Shi X.: Binary grey wolf optimization-regularized extreme learning machine wrapper coupled with the boruta algorithm for monthly streamflow forecasting. Water Resources Management 35(3): 1029–45 Feb (2021)

[45] Ghosh, P., Karim, A., Atik, S. T., Afrin, S., and Saifuzzaman, M.: Expert cancer model using supervised algorithms with a lasso selection approach. International Journal of Electrical and Computer Engineering (IJECE) 11, 2631 (2021) doi:10.11591/ijece.v11i3.pp2631-2639

[46] Rajendran, P., Haw, S.-C., and Naveen, P.: Classification of heart disease using Machine Learning Techniques. 5th International Conference on Digital Technology in Education. (2021) doi:10.1145/3488466.3488482

[47] Junayed, M. S., Jeny, A. A., Atik, S. T., Neehal, N., Karim, A., Azam, S., Shanmugam, B.: AcneNet - a deep CNN based classification approach for Acne Classes. 12th International Conference on Information & Communication Technology and System (ICTS) (2019) doi:10.1109/icts.2019.8850935

[48] Krittanawong, C., Virk, H. U., Bangalore, S., Wang, Z., Johnson, K. W., Pinotti, R., et al.: Machine learning prediction in cardiovascular diseases: A meta-analysis. Scientific Reports 10 (2020) doi:10.1038/s41598-020-72685-1

[49] Ghosh, P., Azam, S., Hasib, K. M., Karim, A., Jonkman, M., and Anwar, A.: A performance based study on deep learning algorithms in the effective prediction of breast cancer. International Joint Conference on Neural Networks (IJCNN) (2021) doi:10.1109/ijcnn52387.2021.9534293

[50] Bergamini, M., Iora, P. H., Rocha, T. A., Tchuisseu, Y. P., Dutra, A. de, Scheidt, J. F., et al.: Mapping risk of ischemic heart disease using machine learning in a Brazilian state. PLOS ONE 15 (2020) doi:10.1371/journal.pone.0243558

AUTHOR BIOGRAPHY

M. Raihan is serving as an Assistant Professor at North Western University, located in Khulna, Bangladesh. He has completed his graduation from Khulna University and post-graduation from Khulna University of Engineering & Technology (KUET). Currently he is pursuing his Ph.D. (Doctor of Philosophy) in Computer Science and Engineering from Khulna University, Khulna, Bangladesh. He has more than 75 publications till now and out of that 68 publications are already published and available in different online portals. He has worked as a guest reviewer in several international conferences and journals like Heliyon (Q1), IEEE Access (Q1), Information Science (Q1), Gene Reports, Informatics in Medicine Unlocked etc. Health informatics, machine learning, data science, biomedical engineering, cognitive science, predictive modeling, and mobile computing are the domains of his research.
E-mail: raihanbme@gmail.com; ORCID: 0000-0003-1072-3555

SAMI AZAM (Member, IEEE) is currently a Leading Researcher and a Professor with the Faculty of Science and Technology, Charles Darwin University, Casuarina, NT, Australia. He has several publications in peer-reviewed journals and international conference proceedings. His research interests include computer vision, signal processing, artificial intelligence, and biomedical engineering.
E-mail: sami.azam@cdu.edu.au; ORCID: 0000-0001-7572-9750

Laboni Akter has M.Sc. degree in Biomedical Engineering from Khulna University of Engineering & Technology and B.Sc. degree in Applied Physics, Electronics, and Communication Engineering from Bangabandhu Sheikh Mujibur Rahman Science and Technology University. She worked as an adjunct lecturer in the Department of Computer Science at North Western University. She has worked as a reviewer for several journals and conferences: Frontiers, Heliyon, Journal of Imaging Informatics in Medicine, and PLOS ONE. Her research interests include machine learning, deep learning, data science, and biomedical engineering.
E-mail: laboni.kuet.bme@gmail.com; ORCID: 0000-0002-0063-3188

Md. Mehedi Hassan is a dedicated young researcher, holding a B.Sc. Engineering degree in computer science and engineering from 2022 and currently pursuing his M.Sc. Engineering degree at Khulna University, Bangladesh. Mehedi’s research interests encompass a broad spectrum, ranging from human brain imaging, neuroscience, machine learning, and artificial intelligence to software engineering.
E-mail: mehedihassan@ieee.org; ORCID: 0000-0002-9890-0968

Ryana Quadir received the B.Sc. degree in Computer Science and Engineering from North South University (NSU), and the master's degree from Daffodil International University (DIU), Bangladesh. Currently, she's pursuing her master's degree at the University of Michigan-Dearborn. She was a Software QA Automation Engineer at Stibo DX for 8 years. Her research interests include deep learning, fuzzy logic, and AI in medical diagnostics.
E-mail: ryana25-866@diu.edu.bd

ASIF KARIM (Member, IEEE) is currently a Research Active Lecturer with Charles Darwin University, Australia. Alongside being an Active Researcher, he has considerable industry experience in IT, primarily in the field of software engineering. His research interests include machine intelligence, health informatics, and smart contracts.
E-mail: asif.karim@cdu.edu.au; ORCID: 0000-0001-8532-6816

Saikat Mondal is a doctoral researcher in the Software Research Lab, Department of Computer Science, University of Saskatchewan, Canada. He is working as a faculty member at Khulna University, Bangladesh. His research interests include Empirical Software Engineering, Data Mining and Analytics, Explainable Machine Learning, and GenAI. To date, Mr. Mondal has published more than 20 research papers. He also received many prestigious awards, including the University of Saskatchewan (USASK) Graduate Thesis Award, Research Excellence Award, Keith Geddes Award, and Best Faculty Award (Bangladesh). He has been regularly serving as a reviewer/co-reviewer in several top journals (e.g., EMSE, IST, JSS, TOSEM) and conferences (e.g., MSR, ASE, SANER). Saikat is very passionate about community service. He served as the president of the USASK Computer Science Graduate Council and webmaster of the IEEE North Saskatchewan Chapter.
E-mail: saikatcsebd@gmail.com; ORCID: 0000-0003-1767-6392

Arun More is a cardiologist. He established the Rural Health Progress Trust in India and serves as its president. He is now the Chief Executive Officer and Founder of the Mobile Hypertension Clinic (MH CLinic). The elegant idea of Comprehensive Management of Hypertension was his brainchild. Millions of people, particularly in remote places, will be able to get the life-saving treatment they need at his chain of hypertension clinics. He has previously served as the head of operations at the Khulna City-based Fortis Escorts Heart Institute.
E-mail: arunmoregsa@gmail.com