Application of Machine Learning Algorithms in Coronary Heart Disease: A Systematic Literature Review and Meta-Analysis

Solomon Kutiame[1], Richard Millham[2], Adebayor Felix Adekoya[3]
Mark Tettey[4], Benjamin Asubam Weyori[5], Peter Appiahene[6]
University of Energy and Natural Resources, Ghana[1]
Durban University of Technology, South Africa[2]
National Cardiothoracic Center, Korle Bu Teaching Hospital, Accra – Ghana[4]
University of Energy and Natural Resources, Ghana[3, 5, 6]

***Abstract*—This systematic review relied on the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) statement and 37 relevant studies. The literature search used search engines including PubMed, Hindawi, SCOPUS, IEEE Xplore, Web of Science, Google Scholar, Wiley Online, Jstor, Taylor and Francis, Ebscohost, and ScienceDirect. This study focused on four aspects: Machine Learning Algorithms, datasets, best-performing algorithms, and software used in coronary heart disease (CHD) predictions. The empirical articles never mentioned 'Reinforcement Learning,' a promising aspect of Machine Learning. Ensemble algorithms showed reasonable accuracy rates but were not common, whereas deep neural networks were poorly represented. Only a few papers applied primary datasets (4 of 37). Logistic Regression (LR), Deep Neural Network (DNN), K-Means, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and boosting algorithms were the best performing algorithms. This systematic review will be valuable for researchers predicting coronary heart disease using machine learning techniques.**

Keywords—Coronary heart diseases; algorithms; datasets; ensembling algorithms; machine learning; artificial intelligence

I. Introduction

The rate at which people are losing their lives due to cardiovascular disease (CVD) is devastating, with the World Health Organization (WHO) estimating annually 17.9 million deaths globally. The heart and blood vessels [1] are the centers of CVD. There are several types of CVDs, such as coronary heart disease (CHD), abnormal heart rhythms (arrhythmias), and heart muscle disease (cardiomyopathy) with CHD being the primary type of CVD [2], contributing to approximately 64% of all CVDs [3]. Men are mainly affected by CHD compared to women.

In most cases, CHD begins in the fourth decade of life and progresses with age [4]. Recent statistics indicate that continentally, for CVD and CHD, Europe has (44%, 24%), North America (32%, 19%), Latin America and Caribbean (27%, 14%), Asia (35%, 16%), Africa (18%, 9%), and Australia (31%, 16%) respectively making a global percentage of 33% of deaths due to cardiovascular disease in general and 16% of deaths due to coronary heart disease [5]. The most precise method for diagnosing CHD is angiography, but it is invasive and expensive [6]. CVD is a global problem and a leading cause of death in Ghana [7]. Early detection and prognosis are critical for reducing the disease burden. Artificial Intelligence (AI) assists in identifying critical indicators of cardiac disease. AI helps determine disease history and supplies appropriate treatment [8]. AI has several benefits as identified in the literature. These include timely decision-making, helping surgeons perform complex surgeries and tasks, providing accurate cardiovascular imaging, reducing the risks of complex treatments, enhancing cardiology knowledge about patient behavior, and improving computer-aided diagnosis [8]. Artificial Intelligence and Machine Learning Algorithmic research focus on inexpensive, rapid, and non-invasive methods to accurately diagnose all CHD using high-level performance metrics, including Accuracy, precision, specificity, and sensitivity.

AI and machine learning allow computers to find, quantify, and interpret correlations between variables to improve patient care by algorithmically learning optimal data representations[9]. Machine-learning algorithms can sift through considerable amounts of cardiovascular data, making it easier to uncover predictive, diagnostic, and therapeutic options for various cardiovascular diseases. There are three types of machine learning algorithms: unsupervised, supervised, and reinforcement learning [10]. Supervised learning is a robust approach that uses machine language to classify and interpret labelled cardiovascular data [11]. For instance, in supervised learning, a physicist may seek to determine whether a particular electrocardiogram represents sinus rhythm or coronary artery, or ventricular fibrillation? Thus, supervised learning requires a dataset with predictor variables, also known as features in machine learning terminology, and labelled outcomes [9]. Unsupervised learning attempts to uncover the underlying structure or correlations between variables in a dataset [12]. This dataset was trained without explicit labels, and the algorithm clustered the data to identify the underlying patterns. Based on behavioural psychology, reinforcement learning employs a different strategy in which a software program operates in a pre-determined environment to maximise a reward. The main objective of this systematic literature review was to determine which supervised machine learning algorithms exhibit the best results for coronary heart disease prediction. In this systematic review: 1) the authors identified studies that employed machine learning techniques to diagnose coronary heart diseases in this systematic review; 2) to determine the most utilized supervised machine learning algorithms for coronary heart disease prediction; 3) to evaluate the performance of supervised machine learning algorithms relative to selected features such as Accuracy, specificity, sensitivity, and precision; and 4) to analyze the data sources for predicting coronary heart disease. The outcomes of this review will contribute to policy directions, practice, and further research on cardiovascular diseases, mainly coronary heart disease. The following sections are as follows: Section II, Materials and Methods, Section III, Literature Review, Section IV, Discussion, and the final part is Conclusion and future work.

II. Materials and Methods

This systematic review adopted the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) statement. PRISMA is a collection of elements used to report systematic reviews and meta-analyses [13]. This approach is intended to help report reviews and assess randomized trials [14], but may also be used as a foundation for reporting systematic reviews [15]. Following the PRISMA approach, this study considered research questions guiding the review, literature search criteria, and selection criteria.

A. Literature Search Criteria

Our selection criteria identified relevant published research and review papers by using keywords such as cardiology, cardiovascular disease, coronary heart disease, ischemic heart disease, coronary artery disease, machine learning algorithms, machine learning techniques, data mining, and artificial intelligence in cardiovascular disease. This study adopted composite literature search criteria which combine keywords using Boolean operators such as AND, --, ",", ~, and OR. The search engines used included PubMed, Hindawi, SCOPUS, IEEE Xplore, Web of Science, Google Scholar, Wiley Online, Jstor, Taylor and Francis, Ebscohost, and ScienceDirect. Furthermore, the search outcome produced over 563 papers, of which 37 were deemed appropriate based on the selection criteria of this study.

B. Selection Criteria

The study utilized peer-reviewed papers published in the English language only. Table I lists the exclusion and inclusion criteria used in this study. These peer-reviewed papers focused on the application of Machine Learning Techniques in investigating CHD. The selection criteria excluded non-peer-reviewed articles such as Dissertation, theses, books, chapters, etc. were excluded from the review.

TABLE I. Article Selection Criteria

Features	Inclusion Criteria	Exclusion Criteria
Language of Publication	English language	Not in the English Language
Research Type	Peer-reviewed	Thesis/Dissertation, case studies, books, reports, and magazines.
Research Focus	Related to ML techniques in CHD.	Not related to ML and CHD
Context	Global	Not Applicable (N/A)

III. Results

A literature search of the above databases identified 563 publications (Fig. 1). Screening by title indicated that 374 papers did not meet the selection criteria for this review. Additional screening by abstract revealed that 75 studies did not match the focus of the review. In contrast, full-text screening of the remaining papers revealed that 108 articles were related and did not apply machine learning techniques to CHD. Therefore, these papers were excluded from the final list. Finally, 37 articles on the implementation of ML algorithms in CHDs were included in this review.

The reference lists of the 37 studies used in this review are listed in Table II. In the literature, it was found that these papers focused on three major areas: 1) CHD Prediction, 2) CHD detection, and 3) CHD diagnosis using ML techniques. A significant proportion of the studies aimed at predicting CHD were based on risk factors and classification methods. The retrieved papers covered 2014, 2016, 2017, 2018, 2019, 2020, and 2021. Fig. 2 depicts the rate of the collected publications on the phenomenon over the periods. The number of publications varied slowly over time except from 2018 to 2019, which saw a significant increase (from 3 to 12 papers). Our search results revealed no studies on CHD prediction using ML methods in 2015 and before 2014. This may be explained by the limited access to open datasets and the arguably emerging nature of ML methods before 2014. The focused nature of the topic and its accompanying paper selection process can also be attributed to the absence of studies in 2015. The number of publications on this phenomenon decreased from 12 in 2019 to 7 in 2020.

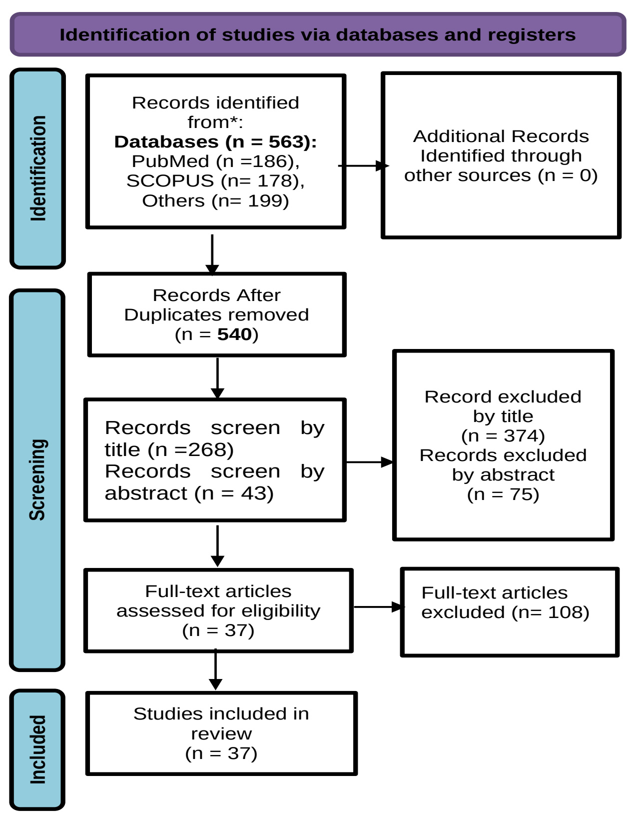

Fig. 1. PRISMA Flow Diagram.

TABLE II. THE LIST OF SELECTED STUDIES FOR THIS SYSTEMATIC LITERATURE REVIEW WITH THEIR FOCUS AREAS

#	Studies	Focus	#	Studies	Focus
1	[18]	Detection of CAD	**20**	[19]	detection and diagnosis of CHD
2	[20]	Improve the prediction CHD	**21**	[21]	CAD classification
3	[22]	CAD presence prediction	**22**	[23]	CHD Prediction
4	[24]	CAD presence prediction	**23**	[25]	Heart disease prediction
5	[26]	Coronary heart disease prediction	**24**	[27]	Heart disease prediction
6	[28]	CHD detection	**25**	[29]	CHD prediction
7	[30]	CAD prediction	**26**	[31]	CHD prediction
8	[32]	predict coronary heart disease	**27**	[33]	prediction of CHD
9	[16]	CHD Prediction based on risk factors	**28**	[34]	classification of coronary artery disease medical data sets
10	[1]	Accuracy of ML algorithms for predicting clinical events	**29**	[35]	Prediction of CHD
11	[17]	methodology of predicting CHD	**30**	[36]	CAD detection
12	[37]	CAD detection	**31**	[2]	CHD Prediction
13	[38]	prediction of heart diseases	**32**	[39]	Heart Disease Diagnosis
14	[40]	prediction of heart diseases	**33**	[41]	CHD prediction
15	[42]	CAD diagnosis	**34**	[43]	CHD prediction
16	[44]	Prediction of CHD	**35**	[45]	NN-based prediction of CHD
17	[46]	Diagnosing CHD	**36**	[47]	Prediction of CHD
18	[48]	prediction of heart disease	**37**	[49]	Prediction of CHD
19	[50]	CHD Diagnosis			

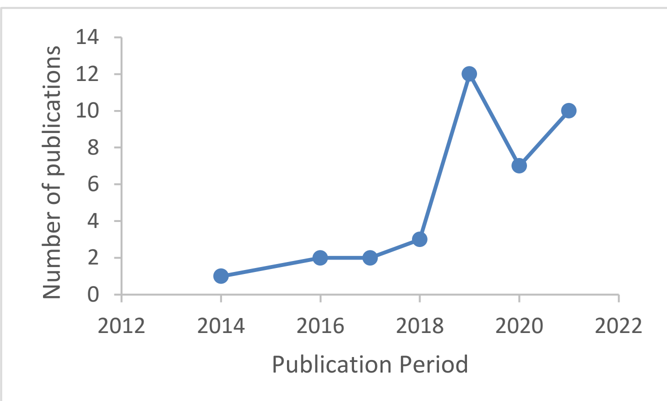

Fig. 2. Number of Publications on CHD using ML Techniques between 2014 and 2021.

A. *Algorithm*

This review considers all reported algorithms applied in previous studies. Table III and Fig. 3 show a group of algorithms applied in the 37 papers and the number of papers that adopted these methods. The review revealed that the most frequently used machine learning algorithms was Decision Tree (n=26, 24%), followed by support vector machines (n=24, 22%), Naïve Bayes (n=23, 21%), K-Nearest Neighbors (n= 19, 17%) and random forests (n=18, 16%). All studies employed multiple algorithms. Few employed five (5) (DT, KNN, SVM, RF, and NN) supervised learning ML algorithms and compared their predictive performance [18]. More than 95% of the studies employed supervised classification algorithms to process features. From Fig. 3, it can be observed that SVM, DT, and KNN retained their popularity. DT has attracted more attention in this field than the other algorithms. Over the periods (2014, 2016, 2017, 2018, 2019, 2020, and 2021), at least one article has been published using DT. SVM and Naïve Bayes have almost the same usage frequencies as DT. Another frequently used method is Logistic regression. Although Deep Neural Networks have demonstrated high predictive power, they have recorded few applications in the field of CHDs. 2017, 2019, and 2021 recorded an increasing usage of boosting and assembling techniques from 2 to 9 publications and a slight decrease to 7 in 2021. From Table II, it is noticeable that XGBoost, Adaboost, Boosted DT, and other ensemble techniques were applied. In addition, while no paper reported the application of reinforcement methods, only one article used an unsupervised algorithm (K-Means) to predict CHDs, although the algorithm performed well. Computational-intelligence-based methods, such as neural networks, have recorded the lowest number of applications.

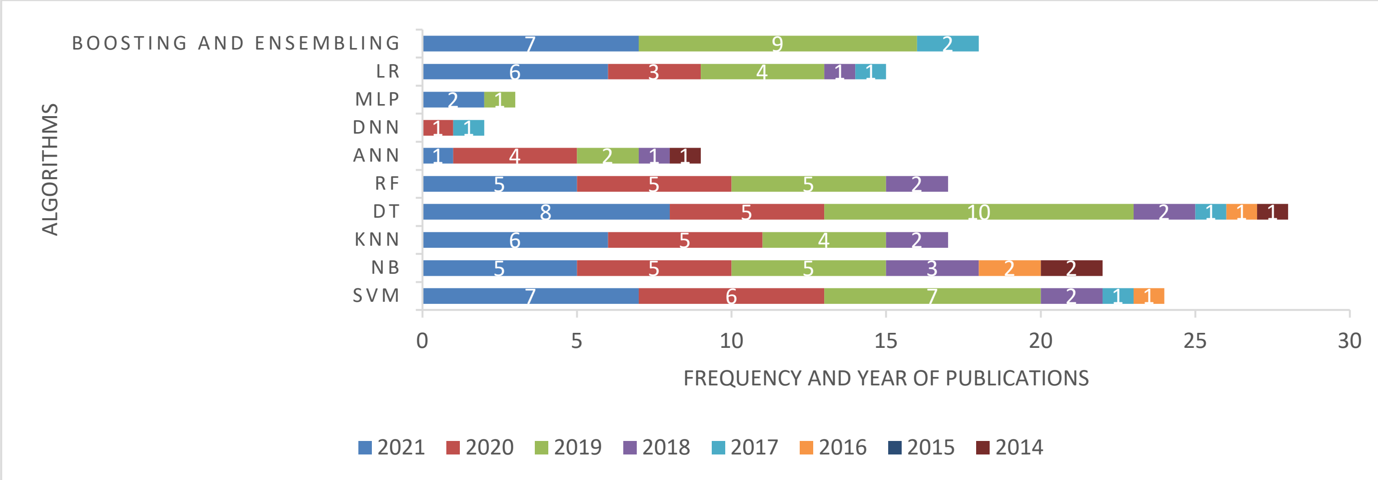

Fig. 3. Shows the Yearly usage Frequencies of Popular ML Methods for CHD Prediction.

TABLE III. ML ALGORITHMS APPLIED IN OBSERVED WORKS

Algorithms	**No. of Papers**
Support Vector Machine	24
Naïve Bayes	23
K-Nearest Neighbors	19
Decision Tree	26
Random Forest	18
Neural Network	14
Regression	17
Boosting & Ensembles	19

B. Algorithm Performance Metrics

One of the most crucial processes in the machine learning prediction process is model evaluation. A model's performance can be assessed using a variety of metrics. During the training process, measurements are frequently performed on unseen samples. The model performance measurement metrics used in the included studies were Accuracy, Precision, Recall/Sensitivity, Specificity, and F1 Score. The review revealed that the most used evaluation metrics in CHD prediction are Accuracy, followed by Recall/Sensitivity, F1 Score, Precision, and Specificity. Table IV shows that out of the 37 papers included, 36 reported the accuracy scores of the models used for the prediction. Supplementary Fig. 4 shows the usage percentages of the metrics in the investigated articles. This analysis focused on examining the performance of the algorithms applied in previous studies. However, because it is not advisable to "directly compare the efficiency of two algorithms or systems if they were evaluated on different datasets" [13], the evaluation of the best-performing algorithms is based on the datasets used. Drawing on the most used evaluation method (i.e. Accuracy), the best performing algorithms were determined based on the mean values of the accuracy scores of the models obtained from the 37 papers. As clinical data and study scope differ greatly among disease prediction studies, a comparison can only be made after a consistent benchmark on the dataset and scope are established. Therefore, only studies that implemented multiple machine learning methods were selected to compare the same data and disease prediction. The authors determined the best-performing algorithm for the phenomenon by comparing the mean accuracy scores of the models that utilized the same datasets. Table V lists the algorithms used on the different datasets and the computed mean scores in terms of specificity, Recall/sensitivity, precision, F1 score, and Accuracy. The current study draws on the computed average scores of the 36 studies to rank the most performing algorithms.

Thus, the higher the accuracy of an algorithm, the higher is the chance of making an accurate prediction. Fig. 5 indicates that among the common algorithms applied on the Z-Alizadeh Sani Heart Disease Dataset, LR (87.51%) had a higher predictive accuracy rate, followed by RF (86.07%), NN (85.74%), SVM (85.55%), KNN (75.72%), NB (68.34%), and DT (68.95%). For the Statlog Heart Disease Dataset, the DNN (98.15%) algorithm obtained the highest predictive accuracy score, as shown in Fig. 6. On this dataset, SVM, LR, NB, NN, and KNN obtained accuracy scores greater than 90%, indicating that these algorithms make significant contributions to disease predictions. Only four algorithms (DNN, DT, KNN, and SVM) have been applied to the long beach dataset so far per the findings of this review. In terms of the best performance, Fig. 7 shows that the DNN has a higher predictive accuracy score compared to the remaining algorithms for the long beach dataset. Fig. 8 and 11 show that the MLP-NN algorithm obtained the highest accuracy rates on the Framingham (73.4%) and South African Heart Disease datasets (73.4%).

TABLE IV. ALGORITHM PERFORMANCE METRICS APPLIED IN THE OBSERVED STUDIES

Metric	**Number of studies**
Accuracy	36
F1 Score	15
Precision	13
Recall/Sensitivity	20
Specificity	10

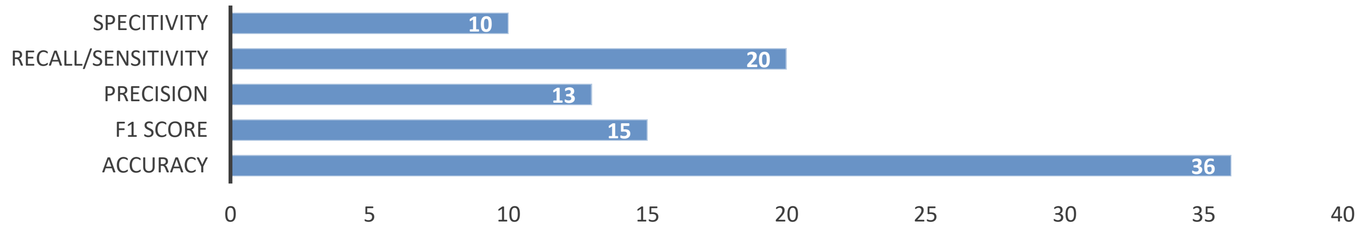

Fig. 4. Number of Algorithm Performance Metrics Applied in Observed Studies.

TABLE V. ALGORITHM PERFORMANCE BASED ON EVALUATION METRICS AND DATASETS

Dataset	Sample Size	Algorithms	ACC	F1	Recall/Sensitivity	Specificity
Z-Alizadeh Sani Heart Disease	303	DT	68.95	85.31	90.77	-
		SVM	85.55	88.85	86.23	88
		RF	86.07	92.80	93.85	-
		LR*	87.51	93.00	89.23	-
		NB	68.34	73.18	94.00	41
		NN	85.74	85.94	84.62	-
		KNN	75.72	90.76	90.77	-
Cleveland Dataset	303	DT	81.75	83.35	78.16	85.51
		SVM	84.72	82.65	81.69	84.15
		RF	83.02	85.33	88.84	91.32
		LR	85.24	81.28	77.58	81.17
		NB	82.74	77.73	81.82	84.34
		NN	83.03	89.84	76.9	82.04
		KNN	81.71	79.99	76.36	78.56
		SSA-N	86.7	-	60	100
		BO-SVM	93.3	-	80	100
		Adaboost	82.12	80.43	79.85	81.02
		DT+RF	88	-	---	-
		DT(SVC)	72.2	69.1	72.2	76.70
		DT(J48)	73	69.8	63.3	77.80
		MLP	-	72	73.9	73.90
		GB	-	92	92.84	90.32
		K-Means	94.06	-	-	-
Statlog Disease Dataset	270	DT	88.1	87.85	84.96	91.15
		SVM	91.97	91.26	87.46	94.64
		RF	89.05	91.93	86.50	90.42
		LR	91.97	91.93	89.34	91.41
		NB	91.38	92	90.86	90.71
		NN	93.03	89.60	93.8	
		KNN	90.35	89.60	86.02	93.42
		DNN	98.15	98	98	91.76
Framingham Heart Study" dataset	4240	DT	82.4	91.47	52.40	81.60
		SVM	68	-	68	68
		RF	83.68	96.61	84.74	80
		LR	66.83	-	67	66.5
		NB	60	-	31	
		NN	69.5	-	69.5	69.5
		KNN	85.45	90.76	86.21	98.67
		Boosted DT	73	-	36	73

		Adaboost	66.60	-	67	-
South African Heart Disease dataset	462	DT	70.07		50	
		SVM	72.75	55	50	88.4
		LR	72.7	56.3		84.44
		NB	71.6		62	
		KNN	73.2	50.4		91.1
		MLP-NN	73.4	55.3		87.1
Long Beach data set	200	DT	68.4			
		KNN	81.1			
		SVM	83.4			
		DNN	84			
the Korea National Health and Nutrition Examination Survey (KNHANES) 2007-2016	4146	LR	85.61		51.44	91.15
		SVM	77.87		77.40	77.81
		RF	76.06		76.44	76.06
		AdaBoost	90.12		52.88	90.36
		MLP	78.8		66.34	78.88
		NN-FRS	81.09			
		NN-FCA	23.87			

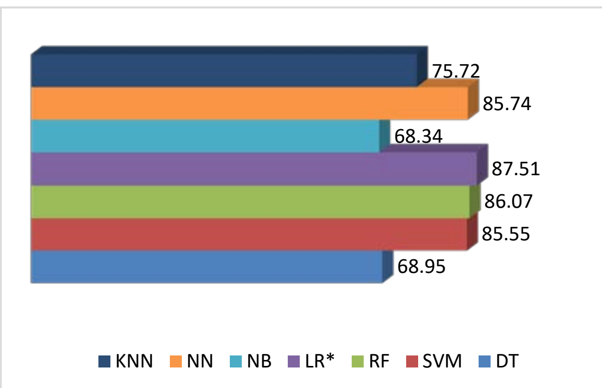

Fig. 5. Best Performing Algorithm - Z-Alizadeh Sani Heart Disease Dataset.

In the KHANNES and Cleveland heart disease dataset, it was found that the ensemble algorithms performed better than single algorithms. This is evident from Fig 7, Fig. 9, Fig 10, and Fig. 11, respectively. In Table VI, the primary dataset names and metrics and the corresponding machine learning algorithms used to predict them are discussed. This table also describes the best-performing algorithm for each model. From the clinical data obtained from primary sources, it was observed that the SVM algorithm is applied most frequently (in all four datasets), followed by the NB algorithm (in 3 datasets). Although AdaBoost has been considered the second least number of times, it showed the highest percentage (i.e. 90.12%) in revealing superior Accuracy followed by SVM.

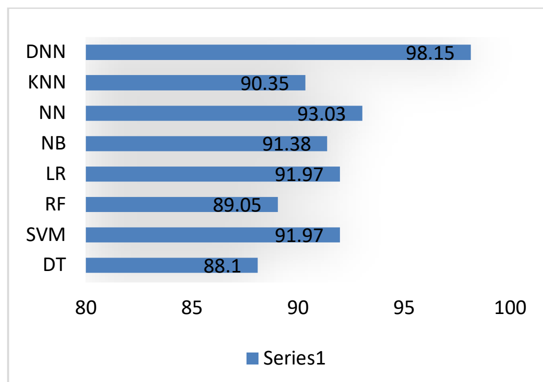

Fig. 6. Best Performing Algorithm - Statlog Heart Disease Dataset.

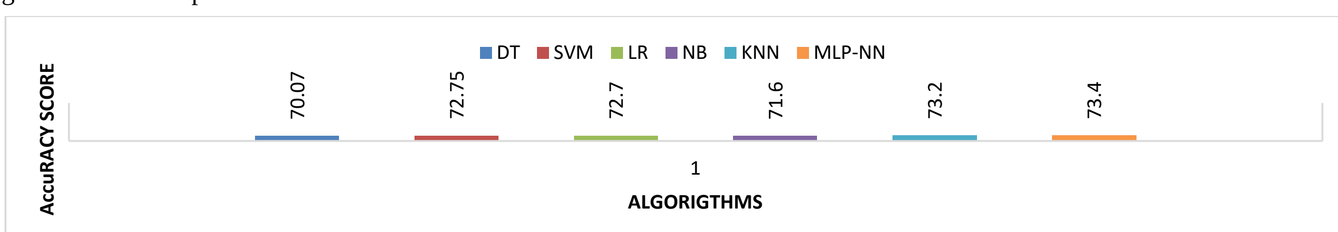

Fig. 7. Best Performing Algorithm - Framingham Heart Study Dataset.

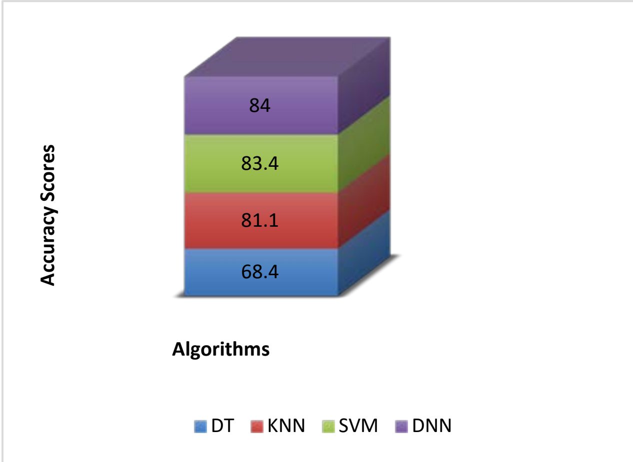

Fig. 8. Best Performing Algorithm – Long Beach Dataset.

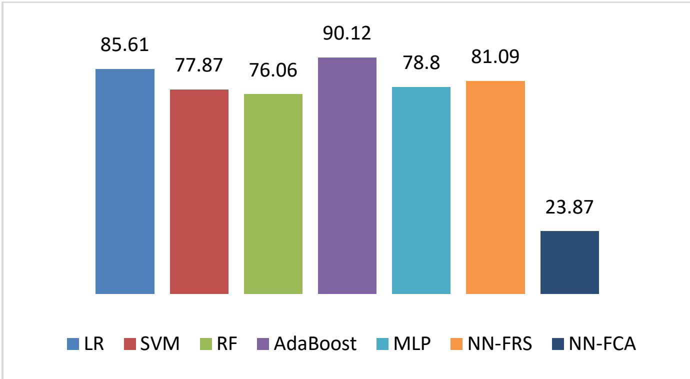

Fig. 9. Best Performing Algorithm – KNHANES Dataset.

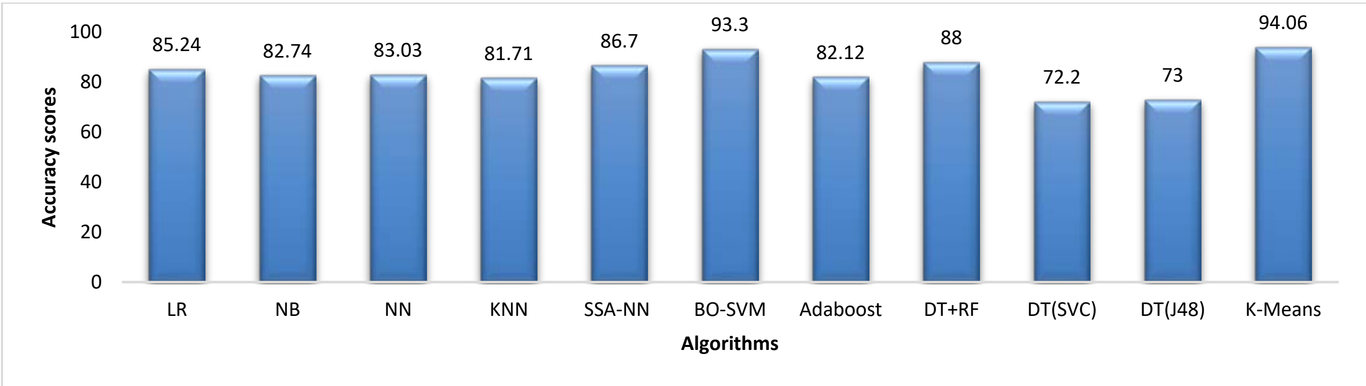

Fig. 10. Best Performing Algorithm: Cleveland Heart Disease Dataset.

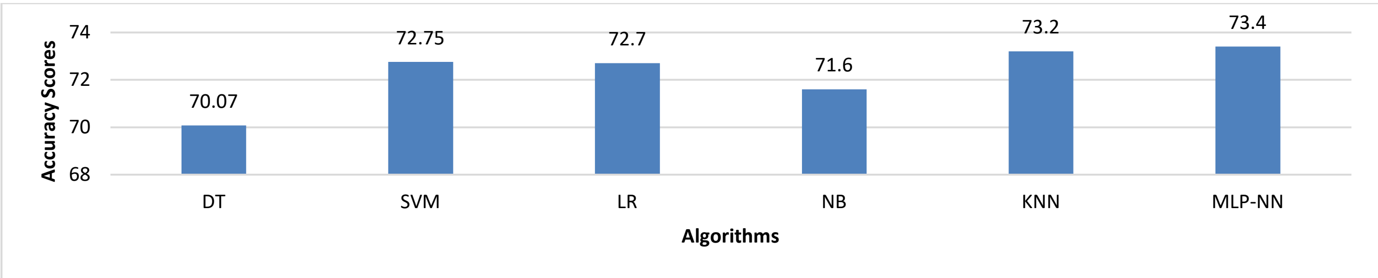

Fig. 11. Best Performing Algorithm: South African Heart Disease Dataset.

TABLE VI. PERFORMANCE METRICS FOR UNPUBLISHED PRIMARY DATASETS

Dataset	Sample Size	Algorithms	Accuracy Score	F1	Recall/ Sensitivity	Specificity
The sample was the medical records of the patients with coronary artery disease who were hospitalized in three hospitals affiliated with AJA University of Medical Sciences between March 2016 and March 2017	1324	ANN	87.52	70.6	88.01	73.64
		SVM	88.91	87.1	92.23	74.42
data from the Sylhet region of Bangladesh by physically going to door-to-door hospitals and healthcare industries	564	DT	82.45			
		SVM	86.03			
		LR	83.12			
		NB	83.14			
		KNN	81.12			
medical records data collected from Harapan Kita Hospital	450	DT	84.44	56		88
		SVM	75.5	75		88
		LR	73.3	60		92
		NB	76.6	38		86
		KNN	72.2	72.2		95
A dataset of 299 heart failure patients. Faisalabad Inst. Of cardiology and the Allied Hospital in Faisalabad.	299	DT	82.4	91.47	52.40	81.60
		SVM	68	-	68	68
		RF	83.68	96.61	84.74	80
		LR	66.83	-	67	66.5
		NB	60	-	31	
		NN	69.5	-	69.5	69.5
		KNN	85.45	90.76	86.21	98.67
		MLP-NN	73.4	55.3		87.1
		SVM	77.87		77.40	77.81
		RF	76.06		76.44	76.06
		AdaBoost	90.12		52.88	90.36
		MLP	78.8		66.34	78.88
		NN-FRS	81.09			
		NN-FCA	23.87			

C. Datasets and Data Source

In machine learning, datasets must be trained to test models for accurate predictions. In the studies reviewed in our research, two primary data sources were observed: (1) open access; and (2) unpublished primary datasets (Fig. 13). The open-access datasets employed in these studies include the Z-Alizadeh Sani, Cleveland, Statlog, Framingham Heart Study, South African Heart Disease, Long Beach, and KNHANES datasets. These datasets can be accessed from the UCI machine learning data repository. The open-access group comprises thirty-six studies and five databases (Table VII), but some datasets were used in more than one study. For example, [32][30] used the Cleveland Heart disease dataset obtained from the UCI database. The dataset contained 303 patients with heart diseases. According to the homepage of the dataset, most researchers allude to utilizing a subset of 14 of the 76 features. The subset includes sex, age, maximum heart rate achieved, chest pain type, serum cholesterol in mg/dl, fasting blood sugar, resting blood pressure, resting electrocardiographic results, exercise-induced angina, ST depression, number of major vessels and diagnosis of heart disease (the predictable attribute), and the slope of the peak exercise ST segment. Out of the 37 papers used in this study, ten (10) articles used the dataset to study the prediction of CHDs. The second most popular dataset used in previous CHD prediction studies was the Z-Alizadeh Sani heart disease dataset. The dataset contains features arranged into four groups: demographic, symptom and examination, ECG, and laboratory and echo features. It categorizes patients into CAD or Normal if their diameter narrowing is greater than or equal to 50%, and otherwise as Normal. The dataset was employed in six studies and is accessible from the UCI Repository. The Statlog heart disease database has 13 attributes that include age, sex, chest pain type (4 values), resting blood pressure, serum cholesterol in mg/dl, fasting blood sugar > 120 mg/dl, resting electrocardiographic results (values 0,1,2), maximum heart rate achieved, exercise-induced angina, old peak = ST depression induced by exercise relative to rest, the slope of the peak exercise ST segment, number of major vessels (0-3) colored by fluoroscopy, and thal. Two researches used the Framingham dataset, which is openly accessible on Kaggle.com. The data is from an active cardiac study on Massachusett and Framingham residents. The classification objective was to determine the patient's 10-year risk of CHD. The dataset contains more than 4,000 patient information and 15 features. Every feature is a prospective risk indicator. The risk indicators include behavioral, medical, and demographic factors. One study used a Dataset obtained from the Korean National Health and Nutrition Examination Survey VI (KNHANES-VI) to develop a CHD prediction model. These studies reported 25,340 records without previous myocardial infarction or angina from the KNHANES dataset.

D. Software/Tools used for the CHD Prediction

Machine learning-based approaches are commonly used to predict CHD. Several tools and programming techniques are used for developing the models for the predictions. The reviewed studies utilized different software for the analyses. These software/tools were categorized into programming and data mining software (Fig. 12). The programming software for machine learning data analysis reported by the studies includes Python programming environments such as Jupyter Notebook and the R programming environment RStudio. WEKA, MATLAB, and Rapid Miner are the commonly used data mining software for CHD prediction in these studies. Of the 27 papers reviewed, 63% (17 articles) reported on the software used. Out of these, 65% used data mining tools, while the remaining 35% used programming technologies. A comparison and evaluation of the different algorithms showed the highest prediction accuracy with the programming tools. Several algorithms, such as DT, RF, SVM, NN, and LR, have been tested on the Framingham Heart Disease dataset, for example, using Rapid Miner and R. The results showed that DT, RF, SVM, NN, and LR achieved accuracy rates of 84%, 78%, 68%, 71%, and 66% with R. Accuracy values obtained with Rapid Miner were as follows; DT (62%), RF (63%), SVM (68%), NN (68%), and LR (67%). The evaluation results show that the accuracy, specificity, and recall values of the various algorithms improved with R.

TABLE VII. SOFTWARE USED FOR ML DATA ANALYSES

Study	**Software/Tool**	**Algorithm**	**Accuracy**
[41]	WEKA	J48, BF Tree, REP Tree, NB Tree	55.77, 62.04, 60.06, 60.06
[18]	R	LR, DT, RF, SVM, NN, KNN	87.64, 79.78, 87.64, 86.52, 93.03, 84.27
[45]	R	SVM, ANN	NR
[49]	R	DT, Boosted DT, RF, SVM, NN, LR	84, 84, 78, 68, 71, 66
[49]	Rapid Miner	DT, Boosted DT, RF, SVM, NN, LR	62, 62, 63, 68, 68, 67
[22]	Python	DT, LR, KNN, NB, SVM	82.45, 83.12, 81.12, 83.14, 86.03
[19]	WEKA	DT, NB, SVM	70.70, 71.60, 71.00
[26]	WEKA	SVM, MLP NN, KNN, LR	73.80, 73.40, 73.20, 72.70
[21]	WEKA	RF, DT, KNN	96.71, 92.1, 91.49
[39]	WEKA	NB, DT, ANN	86.53, 89, 85.53
[42]	WEKA	KNN, DT, GB, RF, SVM, NB LR, ANN	85.55, 86.82, 91.34, 89.45 84.28, 82.33, 84.08, 85.07
[33]	WEKA	SVM, NB, KNN	0.83, 0.84, 0.80
[25]	WEKA	NB, DT, RF	78.56, 82.43, 85.78
[28]	MATLAB	KNN, NB, NN, SSA-NN, SVM, BO-SVM	80, 86.7, 80, 86.7, 80, 93.3
[32]	Python	DT, RF, (DT+RF)	79, 81, 88
[16]	Python	LR, KNN, DT(J48), DT (SVC), NB, NN (MLP)	71.4, 71.6, 73.0, 72.2, 71.4, 73.9
[27]	WEKA	RF, KNN, MLP, Bagging, C4.5, LR, NB, AdaBoost SVM	78.0%, 71.6%, 63.8%, 63.1%, 62.9%, 62.4%, 60.5%, 50.4%, 46%

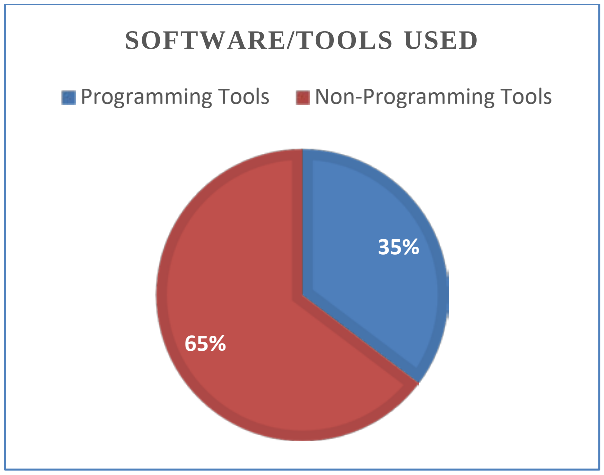

Fig. 12. Distribution of Analysis Tools.

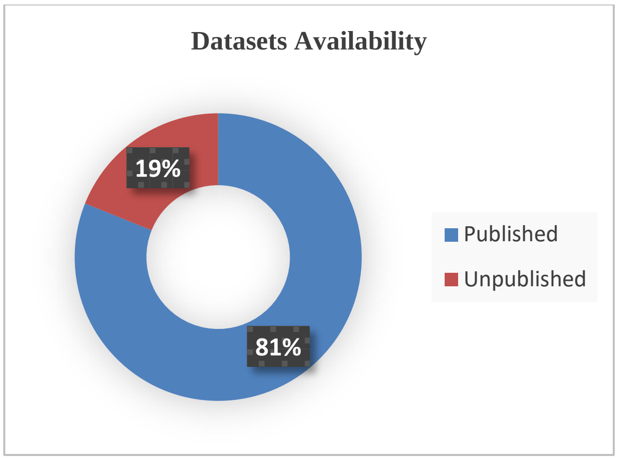

Fig. 13. Distribution of Dataset Types used.

IV. Discussion

This study systematically reviewed 37 studies that utilized machine learning algorithms for CHD prediction. Following the review questions, the study sought to identify the ML algorithms used for CHD predictions, evaluation techniques used and best-performing algorithms, the dataset used, and the software used for the analysis. The outcome of this review highlights the state-of-the-art ML Algorithms applied in predicting CHDs, their performance, and gaps future studies should pay attention to. The results show that previous studies have focused on supervised machine learning classification-based algorithms for observed algorithms. The review indicated that Decision trees, support vector machines, naïve Bayes, and K-Nearest Neighbors were the most utilized algorithms in CHD prediction studies, followed by Random Forest, Logistic Regression, Neural Network, and Bagging and Ensemble algorithms. There is insufficient application of innovative and sophisticated ensemble algorithms such as XGBoost despite gaining popularity as "an ensemble method that has empirically proven to be a highly effective approach by gaining the best results in numerous machine learning competitions" [13]. For CHD prediction using ML, only one study [20] implemented XGBoost. Adaboost, Boosted Decision Tree, and Bagging have received minimal attention despite their promising performance. Moreover, only one study employed the K-Means clustering algorithm. ML approaches can solve regression and clusterization problems, but such works were not observed in the reviewed articles. Many predictions were performed through binary options selection only. The review also revealed that in all the experimental works, single studies applied and compared multiple algorithms (e.g., five to ten algorithms in a single paper). This variety is considered to be more appropriate than a single algorithm for CHD prediction. Similarly, no study has reported on the use of the RL method. This could be due to the complexity of the RL algorithms, and the lack of relevant data, as pointed out by [13]. Nonetheless, the application of RL in medicine is rapidly gaining attention and requires a closer examination in CHD. State-of-the-art algorithms remain outside existing studies. Thus, the application of cutting-edge ML algorithms in CHD remains immature.

This systematic review also focused on investigating the most performing ML algorithms for CHD prediction based on reported evaluation techniques. The study identified five (5) performance measurement metrics for model evaluation. These are as follows precision, F1 Score, Accuracy, specificity, and recall/sensitivity. This number is considered sizeable for comparing and understanding the research results. Including other metrics such as AUC in future studies is imperative. AUC is regarded as a better measure of classifiers than Accuracy due to its unbiased nature on the test data. In terms of best-performing algorithms, the findings showed that boosting algorithms perform better than single algorithms for CHD prediction in terms of Accuracy comparison. This result contradicts the findings of [51] because the authors measured the performance of the algorithms using the AUC evaluation metric. In general, the overall analyses show that the Accuracy of ML algorithms are mostly between 0.8 and 0.9+ in CHD prediction. This indicates that the predictive ability of ML algorithms in CHD is promising, particularly with LR, DNN, K-Means, KNN, SVM, and boosting algorithms. However, there may be methodological barriers to matching clinician-level Accuracy. For example, there are insufficient cases for model training and testing. Therefore, further studies comparing ML models and human expertise are required. In addition, the optimal cut-off for Accuracy remains unclear in the examined studies. For instance, an AUC score of 0.95 or more is recommended, but this is not clear with Accuracy.

The review also indicated that out of the 37 studies, only five (5) employed primary clinical datasets. It is suggested that such data be used in future studies to increase the commitment to predicting real-life CHDs in local environments. This allows for comparing the results and seeing the real advantages or disadvantages of the proposed algorithms. The study also suggests that the publication of primary clinical datasets and research papers will positively impact future developments. The review found no standard guidelines for data partitioning. Most studies used a 10-fold cross-validation technique and a 70/30 or 80/20 splitting method for the training and validation sets. In addition, because the sample size of most datasets was relatively small, the pooled results could be biased. This systematic review shows that most studies employed data mining tools more than programming software, including R, Python (Jupyter Notebook, Google Collab, etc.). In general, the predictive performances in terms of the accuracy scores of the algorithms (i.e., SVM, DT, NN, RF, LR, and Boosted DT) obtained with the data mining software improved with R and Python on the same dataset. However, the runtime of a given algorithm is also crucial because if such a system is to be employed in intensive care units, a speedy decision needs to be made.

A. Gaps and Future Research Directions

This novel work represents the first systematic review of machine learning predictions in CHDs. Given that predicting diseases can help draw attention to avoidable interventions, it is imperative to know the state-of-the-art predictive models, their predictive performance, the nature of datasets, and the technologies for analysis. This review is important because it offers an opportunity to improve these models. Based on the findings of this study, future researchers should consider the following gaps:

1) A lot more studies employing XGBoost, deep neural networks are anticipated.

2) Limited studies focused on clustering and RL algorithms.

3) More studies employing ensemble algorithms, such as the ensemble of Logistic Regression (LR) and Support Vector Machine are suggested for improved prediction.

4) Nearly half of the included studies were conducted in the USA or China. Studies from Oceania, Africa, and the Americas (outside the USA) were limited. This may be partly due to the limited availability of traditional structured health data. Further studies from the perspective of developing countries are required.

5) A dominant reliance on small sample-sized datasets in the included studies. Considering that this may impact the performance of machine learning algorithms, studies with higher data sample sizes are required.

6) Included studies rarely assessed predictive performance in terms of AUC, which is posited to be the best accuracy measurement metric for classifiers. Future studies may focus on having AUC as a model performance metric measure.

V. Conclusion

Although CHD predictions using machine learning applications are being widely researched, many issues remain unaddressed. This study employed a systematic literature review technique to investigate the state-of-the-art ML algorithms used for CHD predictions, evaluation techniques used and best-performing algorithms, the dataset used, and software used for the analyses. The study revealed that a variety of algorithms can be applied to CHD predictions. However, all approaches belong to a class of supervised learning classification methods; most studies utilize published data, whereas fewer studies use primary clinical data. LR, DNN, K-Means, KNN, SVM, and boosting algorithms were found to be the best performing algorithms for CHD prediction; and programming data analysis techniques such as R, and Python were found to produce higher predictive scores than data mining tools such as Rapid miner, WEKA, and MATLAB. This study has some limitations. For instance, only papers from multidisciplinary peer-reviewed databases were, but we did not include articles found in the gray literature. Theses and book chapters are excluded. Considering that CHD is the third cause of total global deaths, understanding the most performing algorithms and software environment for predicting or diagnosing the disease will guide health practitioners and researchers in making proactive decisions to reduce the dangers. The study discovered that the DT algorithm was used the most (in 28 studies), followed by the SVM method (in 24 studies). However, the LR, DNN, K-Means, KNN, and SVM algorithms performed better in comparison. LR demonstrated the highest accuracy, 52 percent, in 8 of the 37 investigations where it was used. This was followed by DNN, which came out on top in 41% of the experiments analyzed.

References

[1] M. Al Mehedi Hasan, J. Shin, U. Das, and A. Yakin Srizon, “Identifying Prognostic Features for Predicting Heart Failure by Using Machine Learning Algorithm,” pp. 40–46, 2021, doi: 10.1145/3460238.3460245.

[2] R. Alizadehsani et al., “Coronary artery disease detection using computational intelligence methods,” Knowledge-Based Syst., vol. 109, pp. 187–197, 2016, doi: 10.1016/j.knosys.2016.07.004.

[3] J. C. Tardif, “Coronary artery disease in 2010,” Eur. Hear. Journal, Suppl., vol. 12, no. SUPPL. C, pp. 2–10, 2010, doi: 10.1093/eurheartj/suq014.

[4] M. A. Khan et al., “Global Epidemiology of Ischemic Heart Disease: Results from the Global Burden of Disease Study,” Cureus, vol. 12, no. 7, 2020, doi: 10.7759/cureus.9349.

[5] Statista, “Small business statistics ‘scary,’” Finweek, 2020. https://www.statista.com/statistics/880155/number-of-smes-in-united-kingdom-uk/.

[6] N. Kumar and D. Kumar, “Machine Learning based Heart Disease Diagnosis using Non-Invasive Methods: A Review,” J. Phys. Conf. Ser., vol. 1950, no. 1, 2021, doi: 10.1088/1742-6596/1950/1/012081.

[7] P. K. A. Abanilla et al., “Cardiovascular disease prevention in Ghana: Feasibility of a faith-based organizational approach,” Bull. World Health Organ., vol. 89, no. 9, pp. 648–656, 2011, doi: 10.2471/BLT.11.086777.

[8] A. Haleem, M. Javaid, R. P. Singh, and R. Suman, “Applications of Artificial Intelligence (AI) for cardiology during COVID-19 pandemic,” Sustain. Oper. Comput., vol. 2, no. February, pp. 71–78, 2021, doi: 10.1016/j.susoc.2021.04.003.

[9] K. W. Johnson et al., “Artificial Intelligence in Cardiology,” J. Am. Coll. Cardiol., vol. 71, no. 23, pp. 2668–2679, 2018, doi: 10.1016/j.jacc.2018.03.521.

[10] P. Constantinides and D. A. Fitzmaurice, “Artificial intelligence in cardiology: Applications, benefits and challenges,” Br. J. Cardiol., vol. 25, no. 3, pp. 86–87, 2018, doi: 10.5837/bjc.2018.024.

[11] M. R. M. Talabis, R. McPherson, I. Miyamoto, J. L. Martin, and D. Kaye, “Analytics Defined,” Inf. Secur. Anal., pp. 1–12, 2015, doi: 10.1016/b978-0-12-800207-0.00001-0.

[12] P. Sardar, J. D. Abbott, A. Kundu, H. D. Aronow, J. F. Granada, and J. Giri, “Impact of Artificial Intelligence on Interventional Cardiology: From Decision-Making Aid to Advanced Interventional Procedure Assistance,” JACC Cardiovasc. Interv., vol. 12, no. 14, pp. 1293–1303, 2019, doi: 10.1016/j.jcin.2019.04.048.

[13] A. Dudchenko, M. Ganzinger, and G. Kopanitsa, “Machine Learning Algorithms in Cardiology Domain: A Systematic Review,” Open Bioinforma. J., vol. 13, no. 1, pp. 25–40, 2020, doi: 10.2174/1875036202013010025.

[14] D. Moher et al., “Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement,” PLoS Med., vol. 6, no. 7, 2009, doi: 10.1371/journal.pmed.1000097.

[15] D. Moher et al., "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement," Syst. Rev., vol. 4, no. 1, pp. 1–9, 2015, doi: 10.1177/1755738018823800.

[16] V. Shorewala, "Early detection of coronary heart disease using ensemble techniques," Informatics Med. Unlocked, p. 100655, 2021, doi: 10.1016/j.imu.2021.100655.

[17] A. Akella and V. Kaushik, "Machine learning algorithms for predicting Coronary Artery Disease: Efforts toward an open source solution," bioRxiv, 2020, doi: 10.1101/2020.02.13.948414.

[18] A. Akella and S. Akella, "Machine learning algorithms for predicting coronary artery disease: Efforts toward an open source solution," Futur. Sci. OA, vol. 7, no. 6, 2021, doi: 10.2144/fsoa-2020-0206.

[19] A. H. Gonsalves, F. Thabtah, R. M. A. Mohammad, and G. Singh, "Prediction of coronary heart disease using machine learning: An experimental analysis," in ACM International Conference Proceeding Series, 2019, no. July, pp. 51–56, doi: 10.1145/3342999.3343015.

[20] L. Ashish, S. K. V, and S. Yeligeti, "Materials Today : Proceedings Ischemic heart disease detection using support vector Machine and extreme gradient boosting method," Mater. Today Proc., no. xxxx, pp. 1–5, 2021, doi: 10.1016/j.matpr.2021.01.715.

[21] D. Krishnani, A. Kumari, A. Dewangan, A. Singh, and N. S. Naik, "Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms," TENCON 2019 - 2019 IEEE Reg. 10 Conf., pp. 367–372, 2019.

[22] M. Nowshad, R. Chowdhury, E. Ahmed, A. D. Siddik, and A. U. Zaman, "Heart Disease Prognosis Using Machine Learning Classification Techniques," pp. 1–6, 2021.

[23] N. Louridi, M. Amar, and B. El Ouahidi, "Identification of Cardiovascular Diseases Using Machine Learning," 7th Mediterr. Congr. Telecommun. 2019, C. 2019, pp. 1–6, 2019, doi: 10.1109/CMT.2019.8931411.

[24] D. Hemalatha and S. Poorani, "Machine learning techniques for heart disease prediction," J. Cardiovasc. Dis. Res., vol. 12, no. 1, pp. 93–96, 2021, doi: 10.31838/jcdr.2021.12.01.05.

[25] A. K. Pal, P. Rawal, R. Ruwala, and V. Patel, "Generic Disease Prediction using Symptoms with Supervised Machine Learning," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., pp. 1082–1086, 2019, doi: 10.32628/cseit1952297.

[26] H. Khdair and N. M. Dasari, "Exploring Machine Learning Techniques for Coronary Heart Disease Prediction," vol. 12, no. 5, 2021.

[27] I. K. A. Enriko, "Comparative study of heart disease diagnosis using top ten data mining classification algorithms," ACM Int. Conf. Proceeding Ser., pp. 159–164, 2019, doi: 10.1145/3338188.3338220.

[28] S. P. Patro, G. S. Nayak, and N. Padhy, "Heart disease prediction by using novel optimization algorithm: A supervised learning prospective," Informatics Med. Unlocked, vol. 26, p. 100696, 2021, doi: 10.1016/j.imu.2021.100696.

[29] S. Dhar, K. Roy, T. Dey, P. Datta, and A. Biswas, "A hybrid machine learning approach for prediction of heart diseases," 2018 4th Int. Conf. Comput. Commun. Autom. ICCCA 2018, pp. 1–6, 2018, doi: 10.1109/CCAA.2018.8777531.

[30] P. Rani, R. Kumar, N. M. O. Sid, and A. Anurag, "A decision support system for heart disease prediction based upon machine learning," J. Reliab. Intell. Environ., no. 0123456789, 2021, doi: 10.1007/s40860-021-00133-6.

[31] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, "A Hybrid Intelligent System Framework for the Prediction of Heart Disease Using Machine Learning Algorithms," vol. 2018, 2018.

[32] G. Renugadevi, G. Asha Priya, B. Dhivyaa Sankari, and R. Gowthamani, "Predicting heart disease using hybrid machine learning model," J. Phys. Conf. Ser., vol. 1916, no. 1, 2021, doi: 10.1088/1742-6596/1916/1/012208.

[33] A. B. Nassif, O. Mahdi, Q. Nasir, M. A. Talib, and M. Azzeh, "Machine Learning Classifications of Coronary Artery Disease," 2018 Int. Jt. Symp. Artif. Intell. Nat. Lang. Process. iSAI-NLP 2018 - Proc., pp. 1–6, 2018, doi: 10.1109/iSAI-NLP.2018.8692942.

[34] J. K. Kim and S. Kang, "Neural Network-Based Coronary Heart Disease Risk Prediction Using Feature Correlation Analysis," vol. 2017, 2017.

[35] A. Caliskan and M. E. Yuksel, "Classification of coronary artery disease data sets by using a deep neural network," vol. 1, no. 4, pp. 271–277, 2017, doi: 10.24190/ISSN2564-615X/2017/04.03.

[36] K. Miao, J. Miao, and G. Miao, "Diagnosing Coronary Heart Disease using Ensemble Machine Learning," Int. J. Adv. Comput. Sci. Appl., vol. 7, no. 10, pp. 30–39, 2016, doi: 10.14569/ijacsa.2016.071004.

[37] S. I. Ayon, M. M. Islam, and M. R. Hossain, "Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques," IETE J. Res., vol. 0, no. 0, pp. 1–20, 2020, doi: 10.1080/03772063.2020.1713916.

[38] U. Das, A. Y. Srizon, and M. A. M. Hasan, "Accurate Recognition of Coronary Artery Disease by Applying Machine Learning Classifiers," ICCIT 2020 - 23rd Int. Conf. Comput. Inf. Technol. Proc., pp. 19–21, 2020, doi: 10.1109/ICCIT51783.2020.9392732.

[39] M. E. Lizabeth, B. Rickner, and A. L. Ange, "C ONGENITAL H EART D ISEASE IN A DULTS First of Two Parts," Rev. Artic. Med. Prog., vol. 256, pp. 53–59, 2014.

[40] M. M. Ghiasi, S. Zendehboudi, and A. A. Mohsenipour, "Decision tree-based diagnosis of coronary artery disease: CART model," Comput. Methods Programs Biomed., vol. 192, p. 105400, 2020, doi: 10.1016/j.cmpb.2020.105400.

[41] M. Abdar, E. Nasarian, X. Zhou, G. Bargshady, V. N. Wijayaningrum, and S. Hussain, "Performance improvement of decision trees for diagnosis of coronary artery disease using multi filtering approach," 2019 IEEE 4th Int. Conf. Comput. Commun. Syst. ICCCS 2019, no. Dm, pp. 26–30, 2019, doi: 10.1109/CCOMS.2019.8821633.

[42] M. Yaqoob, F. Iqbal, and S. Zahir, "Comparing predictive performance of k-nearest neighbors and support vector machine for predicting ischemic heart disease," vol. 1, no. 2, pp. 49–60, 2020.

[43] R. Alizadehsani et al., "Machine learning-based coronary artery disease diagnosis: A comprehensive review," Comput. Biol. Med., vol. 111, p. 103346, 2019, doi: 10.1016/j.compbiomed.2019.103346.

[44] V. Sharma, S. Yadav, and M. Gupta, "Heart Disease Prediction using Machine Learning Techniques," Proc. - IEEE 2020 2nd Int. Conf. Adv. Comput. Commun. Control Networking, ICACCCN 2020, vol. 1, no. 6, pp. 177–181, 2020, doi: 10.1109/ICACCCN51052.2020.9362842.

[45] H. Ayatollahi and L. Gholamhosseini, "Predicting coronary artery disease : a comparison between two data mining algorithms," pp. 1–9, 2019.

[46] M. T. Islam, S. R. Rafa, and M. G. Kibria, "Early Prediction of Heart Disease Using PCA and Hybrid Genetic Algorithm with k-Means," ICCIT 2020 - 23rd Int. Conf. Comput. Inf. Technol. Proc., 2020, doi: 10.1109/ICCIT51783.2020.9392655.

[47] N. Basha, S. P. Ashok Kumar, C. Gopal Krishna, and P. Venkatesh, "Early Detection of Heart Syndrome Using Machine Learning Technique," 4th Int. Conf. Electr. Electron. Commun. Comput. Technol. Optim. Tech. ICEECCOT 2019, vol. 90, pp. 387–391, 2019, doi: 10.1109/ICEECCOT46775.2019.9114651.

[48] O. Terrada, S. Hamida, B. Cherradi, A. Raihani, and O. Bouattane, "Supervised machine learning based medical diagnosis support system for prediction of patients with heart disease," Adv. Sci. Technol. Eng. Syst., vol. 5, no. 5, pp. 269–277, 2020, doi: 10.25046/AJ050533.

[49] J. Beunza et al., "Comparison of machine learning algorithms for clinical event prediction (risk of coronary heart disease)," J. Biomed. Inform., p. 103257, 2019, doi: 10.1016/j.jbi.2019.103257.

[50] A. Dutta, T. Batabyal, M. Basu, and S. T. Acton, "An Efficient Convolutional Neural Network for Coronary Heart Disease Prediction."