Heart disease detection using machine learning

Mohammad Al-Habahbeh[1], Moath Alomari[1], Hebatullah Khattab[1], Raed Alazaidah[2], Nawaf Alshdaifat[3], Suhaila Abuowaida[4], Saleh Alqatan[1], Mohammad Arabiat[1]
[1]Department of Computer Science, Faculty of Information Technology, Zarqa University, Zarqa, Jordan
[2]Department of Data Science and Artificial Intelligence, Faculty of Information Technology, Zarqa University, Zarqa, Jordan
[3]Faculty of Information Technology, Applied Science Private University, Amman, Jordan
[4]Department of Computer Science, Faculty of Prince Al-Hussein Bin Abdallah II for IT, Al al-Bayt University, Mafraq, Jordan

Article Info

Article history:

Received Feb 14, 2024
Revised Oct 3, 2024
Accepted Oct 17, 2024

Keywords:

Classification
Feature selection
Heart disease
Learning strategies
Machine learning

ABSTRACT

Heart disease continues to be a major worldwide health issue, requiring accurate prediction models to improve early identification and treatment. This research aims to address two main objectives in light of the increasing prevalence of heart-related disorders. Firstly, it aims to determine the most efficient classifier for identifying heart disease among twenty-nine different classifiers that represent six distinct learning strategies. Furthermore, the research seeks to identify the most effective method for selecting features in heart disease datasets. The results show how well different classifiers and feature selection methods work by using two datasets with different features and judging performance using four important criteria. The evaluation results demonstrate that the RandomCommittee classifier outperforms in diagnosing heart illness, displaying strong skills across various learning strategies. This classifier exhibits favorable results in terms of accuracy, precision, recall, and F1-score metrics, hence confirming its appropriateness for predictive modeling in heart-related datasets. Moreover, the paper examines feature selection methods, specifically aiming to determine the most effective method for enhancing the predicted accuracy of heart disease models. The prediction models' overall performance is enhanced by their capacity to accurately identify and prioritize pertinent variables, thereby facilitating the early detection and management of heart-related problems.

Corresponding Author:

Raed Alazaidah
Department of Data Science and Artificial Intelligence, Faculty of Information Technology
Zarqa University
Zarqa, Jordan
Email: wichian.sit@mfu.ac.th

1. INTRODUCTION

Heart disease, including coronary artery disease, arrhythmias, and heart failure, is a serious health issue that can lead to complications and even death. Early detection is crucial for timely intervention, preventing severe complications, and improving patient outcomes and quality of life [1], [2]. Traditional methods for heart disease detection often involve manual assessments, clinical examinations, and basic diagnostic tests. However, modern approaches, especially those incorporating machine learning (ML), leverage advanced algorithms to analyze complex datasets from diverse sources, such as patient history, genetic information, and medical imaging, enabling more nuanced and accurate predictions [3]-[7]. According to the World Health Organization (WHO) [8], in 2019, 17.9 million people died due to cardiovascular diseases, accounting for 32% of global deaths. Heart attacks and strokes were the leading causes, accounting for 85% of these deaths.

ML is crucial in detecting heart disease by analyzing large datasets to identify patterns and markers [7]. It processes patient data like medical history, imaging, and genetic information, enabling precise diagnoses. ML algorithms help healthcare practitioners predict risks, enhance accuracy, and improve personalized treatment strategies, ultimately improving patient outcomes and reducing healthcare burdens. The motivation behind investigating heart disease detection using ML arises from the urgency to enhance early identification of cardiac conditions. Leveraging ML offers a promising avenue to develop efficient algorithms, aiming to improve patient outcomes by enabling timely diagnoses and interventions [7]. This research has two main objectives: the first is to determine the most efficient classifier for identifying heart disease among twenty-nine different classifiers such as: BayesNet, stochastic gradient descent (SGD), and bagging that represent six distinct learning strategies, and the second is to identify the most effective method for selecting features in heart disease datasets [9]-[12]. This research is organized as follows: section 2 discusses the related work; section 3 details the general research method with results; and section 4 covers the conclusion and future work.

2. RELATED WORK

In this section, we will discuss the previous research about heart disease detection using ML. Li *et al.* [13] created the fast conditional mutual information (FCMIM) system, which is based on ML and can detect heart failure. They used different learning methods, such as artificial neural network (ANN), support vector machine (SVM), decision tree (DT), Naive Bayes (NB), k-nearest neighbour (KNN), and linear regression (LR), to create the FCMIM-SVM model. The paper's strengths include thorough data pre-processing, introducing FCMIM feature selection, comparative analysis with other methods, diverse classifier evaluations, and achieving notably high accuracy (92.37%). The fact that SVM was chosen as the best classifier shows how well it worked in building the final model. It provides a strong system for quickly and accurately finding heart disease in their proposed FCMIM-SVM approach for future use in healthcare.

Owida *et al.* [14] explored ML and deep learning methods for heart disease detection, reporting favorable outcomes, especially with deep learning achieving a high 94% accuracy via mobile device technology. Their comparative analysis of three methods highlighted the superiority of ML algorithms, particularly in smaller datasets, aligning with prior research suggestions. Evaluation methods like the confusion matrix, precision, specificity, sensitivity, and F1 score were employed. Notably, the k-neighbors classifier excelled in the ML approach after data preprocessing, specifically with a dataset featuring 13 features, showcasing its effectiveness in heart disease prediction. Jindal *et al.* [15] focus on predicting heart disease using patient medical data and developing a system using logistic regression and KNN algorithms. Their models, notably KNN and logistic regression, displayed strong accuracy in identifying heart disease evidence, relieving pressure by accurately pinpointing disease probabilities. The proposed method enhances medical care and cost reduction and is implemented in. Python format, providing valuable insights for patient prediction. The result of this paper confirms KNN's superior performance with over 88% accuracy in classifying patients with heart disease.

Salhi *et al.* [16] delved into heart disease prediction using data analytics, employing preprocessing and three techniques (neural networks, SVM, and KNN) on varied datasets to gauge accuracy and stability. Neural networks emerged as more manageable and achieved a robust 93% accuracy. Additionally, this paper highlighted the severity of heart diseases and emphasized early prediction for preventive measures. The authors underscored the role of ML in predicting heart diseases and advocating for healthy lifestyles to prevent such conditions and avoid potential health crises. Al-Batah *at el.* [17] emphasizes the significance of early identification and prediction of heart disease using ML techniques on patient data. By utilizing a public dataset featuring clinical and demographic information, various ML methods were compared, demonstrating their ability to accurately predict cardiac disease with high accuracy and area under the ROC (receiver operating characteristics) curve (AUC) values. The research extends its findings to Jordanian patients, highlighting satisfactory predictions. The results of this research hold substantial implications for healthcare, potentially improving patient outcomes and reducing healthcare costs. However, integrating ML models into Jordanian hospitals necessitates further research, including separate dataset evaluations for model reliability, addressing data quality, privacy concerns, interpretability issues, and designing user-friendly interfaces for healthcare professionals within existing systems.

3. METHOD

This section presents the datasets, method, analysis of 29 classifiers, and four feature selection approaches. Subsection 3.1 presents a comprehensive depiction of the four heart disease datasets employed in this paper. Subsection 3.2 presents the methods of research, while subsection 3.3 determines the optimal classifiers. Subsection 3.4 assesses and determines the most optimal feature selection approach out of four widely recognized methods. The main findings are discussed in subsection 3.5.

3.1. Description of datasets

The research work includes two datasets: dataset 1 and dataset 2. Dataset 1 consists of 606 instances and 14 features, while dataset 2 consists of 303 instances and 14 features. Table 1 lists the characteristics of the datasets.

Table 1. Characteristics of the datasets

Name	Instances	Features	Ref.
Dataset 1	606	14	[18]
Dataset 2	303	14	[19]

3.2. Method

The initial phase is collecting data, wherein two datasets pertaining to heart disease are acquired and downloaded from the UCI repository. Additional information about the datasets under consideration is available in the subsequent subsection. The second stage involves pre-processing the data, which involves doing various tasks such as data cleaning, handling missing data, data reduction, and other related operations. The goal is to prepare the data for analysis. The third phase of this paper focuses on determining the optimal classifier for heart disease datasets. The third phase incorporates various evaluation criteria, including accuracy (A), precision (P), recall (R), and F1-score (F1), as outlined in subsection 3.3. In addition, this stage considers 29 distinct classifiers that are part of six widely recognized learning techniques as shown in Table 2. Additional information on these classifiers can also be found in subsection 3.3.

Table 2. Performance of the 29 classifiers using dataset 1

Learning strategy	Classifier	Dataset 1			
		A	P	R	F1
Bayes	BN	83.50	0.84	0.84	0.84
	NB	83.50	0.84	0.84	0.83
Average		83.50	0.84	0.84	0.83
Functions	SGD	85.31	0.86	0.85	0.85
	SL	84.49	0.85	0.85	0.84
	SMO	84.49	0.85	0.85	0.84
Average		84.76	0.85	0.85	0.85
Lazy	IBk	97.03	0.97	0.97	0.97
	KS	96.70	0.97	0.97	0.97
	LWL	80.69	0.81	0.81	0.81
Average		91.47	0.91	0.91	0.91
Meta	ABM1	83.33	0.83	0.83	0.83
	ASC	90.10	0.90	0.90	0.90
	BG	87.46	0.88	0.88	0.87
	CVR	88.28	0.88	0.88	0.88
	FC	87.62	0.88	0.88	0.88
	ICO	84.32	0.84	0.84	0.84
	LB	84.82	0.85	0.85	0.85
	MCC	84.16	0.84	0.84	0.84
	MCCU	85.31	0.86	0.85	0.85
	RC	98.68	0.99	0.99	0.99
	RFC	95.05	0.95	0.95	0.95
	RSS	89.27	0.89	0.89	0.89
Average		88.20	0.88	0.88	0.88
Rules	JRip	87.13	0.87	0.87	0.87
	OR	73.10	0.73	0.73	0.73
	PART	91.58	0.92	0.92	0.92
Average		83.94	0.84	0.84	0.84
Trees	DS	75.91	0.76	0.76	0.76
	HT	82.67	0.83	0.83	0.83
	J48	89.77	0.90	0.90	0.90
	RF	98.35	0.98	0.98	0.98
	RT	97.36	0.97	0.97	0.97
	REPT	86.47	0.87	0.87	0.86
Average		88.42	0.88	0.88	0.88

The subsequent phase seeks to determine the optimal feature selection technique to employ with datasets related to heart disease. This stage examines four widely recognized feature selection techniques, which have been compared and assessed using four evaluation metrics: accuracy, precision, recall, and F1-score. The ultimate stage entails engaging in a discussion.

3.3. Identifying the best classifier

A comprehensive evaluation and comparison were conducted on 29 classifiers from six different learning strategies to identify the most effective classifiers for handling the considered datasets. These classifiers have been utilized with their default implementations in Waikato evolutionary knowledge analysis (WEKA). This data mining studio encompasses a repertoire of techniques and tools employed in data analysis [20]. The evaluation step of the 29 classifiers considers four metrics: accuracy, precision, recall, and F1-score. Additional details about these measurements can be found in reference [21]. The metrics are calculated using (1)-(4):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Recall = \frac{2\ x\ Precisionx\ Recall}{Precision + Recall} \quad (4)$$

where TP is number of samples correctly predicted as positive, FP is number of samples wrongly predicted as positive, TN is number of samples correctly predicted as negative, and FN is number of samples wrongly predicted as negative.

Table 3 represents dataset 1, Table 4 represents dataset 2. Based on the tables, the following are the best classifiers for each dataset:

– Dataset 1: the best classifier is RandomCommittee in the learning strategy meta. It has the highest accuracy (98.68%), precision (0.99), recall (0.99), and F1-score (0.99) among all the classifiers in this dataset.
– Dataset 2: the best classifiers are SGD in learning strategy functions and MultiClassClassifierUpdateable in learning strategy meta. Both have the highest accuracy (83.83%), precision (0.84), recall (0.84), and F1-score (0.84) among all the classifiers in this dataset.

Table 3. Performance of the 29 classifiers using dataset 2

Learning strategy	Classifier	Dataset 2			
		A	P	R	F1
Bayes	BN	82.84	0.83	0.83	0.83
	NB	82.84	0.83	0.83	0.83
Average		82.84	0.83	0.83	0.83
Functions	SGD	**83.83**	**0.84**	**0.84**	**0.84**
	SL	82.51	0.83	0.83	0.82
	SMO	83.50	0.84	0.84	0.83
Average		83.28	0.84	0.83	0.83
Lazy	IBk	76.57	0.77	0.77	0.77
	KS	75.25	0.75	0.75	0.75
	LWL	75.25	0.75	0.75	0.75
Average		75.69	0.76	0.76	0.76
Meta	ABM1	75.25	0.75	0.75	0.75
	ASC	79.54	0.80	0.80	0.80
	BG	81.19	0.81	0.81	0.81
	CVR	79.54	0.80	0.80	0.79
	FC	79.54	0.80	0.80	0.80
	ICO	80.53	0.81	0.81	0.81
	LB	81.19	0.81	0.81	0.81
	MCC	82.18	0.82	0.82	0.82
	MCCU	**83.83**	**0.84**	**0.84**	**0.84**
	RC	79.21	0.79	0.79	0.79
	RFC	65.02	0.65	0.65	0.65
	RSS	82.18	0.83	0.82	0.82
Average		79.10	0.79	0.79	0.79
Rules	JRip	77.56	0.78	0.78	0.77
	OR	72.28	0.72	0.72	0.72
	PART	78.22	0.78	0.78	0.78
Average		76.02	0.76	0.76	0.76
Trees	DS	74.26	0.74	0.74	0.74
	HT	80.53	0.81	0.81	0.81
	J48	78.55	0.79	0.79	0.79
	RF	79.54	0.80	0.80	0.79
	RT	74.26	0.74	0.74	0.74
	REPT	78.88	0.79	0.79	0.79

Table 4. Summarizing the best classifiers for each dataset

Dataset	Best classifier	A (%)	P	R	F1
Dataset 1	RandomCommittee	98.68	0.99	0.99	0.99
Dataset 2	SGD, MultiClassClassifierUpdateable	83.83	0.84	0.84	0.84

In order to obtain a comprehensive understanding, it has been determined that an evaluation will be conducted to find the most effective learning approach for managing the heart disease datasets. This evaluation will be based on the performance metrics of the classifiers that represent various learning strategies [22], [23].

3.4. The best feature selection method

Four distinct feature selection approaches have been selected to assess and compare their suitability for the heart disease domain. The evaluation will be based on four metrics: accuracy, precision, recall, and F1-score. The selection of these approaches is based on their widespread use in the field of ML. The chosen feature selection methods include ClassifierAttributeEval (ClassifierAE), CorrelationAttributeEval (CorrelationAE), GainRatioAttributeEval (GainRatio), and InfoGainAttributeEval (InfoGain). All these feature selection techniques have been utilized with their default configurations as they have been implemented in WEKA. The assessment results for the four feature selection methods discussed in this paper using 50% of the features of dataset 1, dataset 2 are listed in Table 5, (Tables 6 and 7 in Appendix). The assessment part of the feature selection methods involves considering the same 29 classifiers used in the previous section. Additionally, four metrics are under consideration. The metrics include accuracy, precision, recall, and F1-score.

Table 5. Evaluation results for the two feature selection methods for dataset 1

Classifier	DataSet 1							
	ClassifierAE				CorrelationAE			
	A	P	R	F1	A	P	R	F1
ABM1	77.06	0.77	0.77	0.77	77.06	0.77	0.77	0.77
ASC	77.39	0.79	0.77	0.77	77.39	0.79	0.77	0.77
BG	81.68	0.82	0.82	0.82	81.68	0.82	0.82	0.82
BN	78.38	0.79	0.78	0.78	78.38	0.79	0.78	0.78
CVR	79.37	0.80	0.79	0.79	79.37	0.80	0.79	0.79
DS	75.91	0.76	0.76	0.76	75.91	0.76	0.76	0.76
FC	79.21	0.80	0.79	0.79	79.21	0.80	0.79	0.79
HT	76.40	0.77	0.76	0.76	76.40	0.77	0.76	0.76
IBk	94.88	0.95	0.95	0.95	94.88	0.95	0.95	0.95
ICO	78.05	0.78	0.78	0.78	78.05	0.78	0.78	0.78
J48	83.00	0.83	0.83	0.83	83.00	0.83	0.83	0.83
JRip	78.05	0.78	0.78	0.78	78.05	0.78	0.78	0.78
KS	93.73	0.94	0.94	0.94	93.73	0.94	0.94	0.94
LB	78.05	0.78	0.78	0.78	78.05	0.78	0.78	0.78
LWL	77.72	0.78	0.78	0.78	77.72	0.78	0.78	0.78
MCC	76.73	0.77	0.77	0.77	76.73	0.77	0.77	0.77
MCCU	76.73	0.77	0.77	0.77	76.73	0.77	0.77	0.77
NB	78.88	0.79	0.79	0.79	78.88	0.79	0.79	0.79
OR	73.10	0.73	0.73	0.73	73.10	0.73	0.73	0.73
PART	79.04	0.79	0.79	0.79	79.04	0.79	0.79	0.79
RC	**95.54**	**0.96**	**0.96**	**0.96**	**95.54**	**0.96**	**0.96**	**0.96**
REPT	78.71	0.79	0.79	0.79	78.71	0.79	0.79	0.79
RF	94.88	0.95	0.95	0.95	94.88	0.95	0.95	0.95
RFC	**95.54**	**0.96**	**0.96**	**0.96**	**95.54**	**0.96**	**0.96**	**0.96**
RSS	82.34	0.82	0.82	0.82	82.34	0.82	0.82	0.82
RT	**95.54**	**0.96**	**0.96**	**0.96**	**95.54**	**0.96**	**0.96**	**0.96**
SGD	76.73	0.77	0.77	0.77	76.73	0.77	0.77	0.77
SL	75.91	0.76	0.76	0.76	75.91	0.76	0.76	0.76
SMO	77.56	0.78	0.78	0.77	77.56	0.78	0.78	0.77
Average	**81.59**	**0.82**	**0.82**	**0.82**	**81.59**	**0.82**	**0.82**	**0.82**

4. RESULTS AND DISCUSSION

The results discussed in subsections 3.3 and 3.4 provide insights into the most effective classifiers and feature selection methods for handling datasets related to heart disease [21], [24]-[26]. The chosen classifiers and feature selection methods have undergone comprehensive evaluation using important performance measures, including accuracy, precision, recall, and F1-score, across four separate datasets. When examining classifiers, it becomes clear that RandomCommittee consistently outperforms other methods on various datasets, demonstrating exceptional results in terms of accuracy, precision, recall, and F1-score. The strong performance of RandomCommittee, particularly in dataset 1, underscores its dependability in managing heart disease-related data. Furthermore, the paper demonstrates that SGD, when used as a learning strategy function, and MultiClassClassifierUpdateable, when used as a learning strategy meta, show impressive performance in dataset 2. The presence of a wide range of classifiers, such as KStar, RandomCommittee, RandomizableFilteredClassifier, RandomForest, and RandomTree, all achieving perfect scores in all metrics, highlights the significance of using multiple classifiers when working with intricate datasets.

fourRandomCommittee classifiers are highly suitable for managing heart disease datasets, according to the thorough assessment of classifiers. The paper looks at four common feature selection methods and rates them: ClassifierAttributeEval (ClassifierAE), CorrelationAttributeEval (CorrelationAE), GainRatioAttributeEval (GainRatio), and InfoGainAttributeEval (InfoGain). The results indicate that GainRatioAttributeEval is the most effective strategy in terms of all performance measures and datasets. The consistently high average scores in accuracy, precision, recall, and F1-score, as shown in Tables 2 to 7, confirm the usefulness of the feature selection method in identifying significant characteristics for heart disease-related data.

Selecting an effective feature selection strategy is essential for optimizing both the performance and interpretability of a model. As GainRatioAttributeEval is widely used, it shows that it can find relevant characteristics in datasets about heart disease. This makes the ML models used in this investigation more useful overall.

5. CONCLUSION

In this paper, the results from testing classifiers and choosing features show how important it is to use a complex approach when working with heart disease datasets. It is highly suggested that you use RandomCommittee classifiers along with GainRatioAttributeEval feature selection for dependable results in terms of F1-score, accuracy, precision, and recall. Researchers and practitioners in the field of heart disease prediction and management can utilize these findings by implementing the discovered optimal classifiers and feature selection methods to improve the dependability and efficiency of their ML models.

APPENDIX

Table 6. Evaluation results for the two feature selection methods for dataset 1

Classifier	Dataset 1							
	GainRatio				InfoGain			
	A	P	R	F1	A	P	R	F1
ABM1	84.16	0.84	0.84	0.84	84.16	0.84	0.84	0.84
ASC	86.63	0.87	0.87	0.87	86.63	0.87	0.87	0.87
BG	85.81	0.86	0.86	0.86	85.81	0.86	0.86	0.86
BN	84.82	0.85	0.85	0.85	84.82	0.85	0.85	0.85
CVR	85.31	0.85	0.85	0.85	85.31	0.85	0.85	0.85
DS	75.91	0.76	0.76	0.76	75.91	0.76	0.76	0.76
FC	85.81	0.86	0.86	0.86	85.81	0.86	0.86	0.86
HT	82.01	0.82	0.82	0.82	82.01	0.82	0.82	0.82
IBk	97.36	0.97	0.97	0.97	97.36	0.97	0.97	0.97
ICO	84.16	0.84	0.84	0.84	84.16	0.84	0.84	0.84
J48	87.46	0.88	0.88	0.87	87.46	0.88	0.88	0.87
JRip	85.31	0.85	0.85	0.85	85.31	0.85	0.85	0.85
KS	95.71	0.96	0.96	0.96	95.71	0.96	0.96	0.96
LB	84.65	0.85	0.85	0.85	84.65	0.85	0.85	0.85
LWL	78.05	0.78	0.78	0.78	78.05	0.78	0.78	0.78
MCC	83.99	0.84	0.84	0.84	83.99	0.84	0.84	0.84
MCCU	83.33	0.84	0.83	0.83	83.33	0.84	0.83	0.83
NB	83.33	0.84	0.83	0.83	83.33	0.84	0.83	0.83
OR	74.59	0.75	0.75	0.75	74.59	0.75	0.75	0.75
PART	87.79	0.88	0.88	0.88	87.79	0.88	0.88	0.88
RC	98.02	**0.98**	**0.98**	**0.98**	98.02	**0.98**	**0.98**	**0.98**
REPT	83.83	0.84	0.84	0.84	83.83	0.84	0.84	0.84
RF	**98.35**	**0.98**	**0.98**	**0.98**	**98.35**	**0.98**	**0.98**	**0.98**
RFC	97.19	0.97	0.97	0.97	97.19	0.97	0.97	0.97
RSS	87.29	0.88	0.87	0.87	87.29	0.88	0.87	0.87
RT	95.71	0.96	0.96	0.96	95.71	0.96	0.96	0.96
SGD	83.33	0.84	0.83	0.83	83.33	0.84	0.83	0.83
SL	83.83	0.84	0.84	0.84	83.83	0.84	0.84	0.84
SMO	82.84	0.83	0.83	0.83	82.84	0.83	0.83	0.83
Average	**86.43**	**0.87**	**0.86**	**0.86**	**86.43**	**0.87**	**0.86**	**0.86**

Table 7. Evaluation results for the two feature selection methods for dataset 2

Classifier	Dataset 2							
	ClassifierAE				CorrelationAE			
	A	P	R	F1	A	P	R	F1
ABM1	74.92	0.75	0.75	0.75	74.92	0.75	0.75	0.75
ASC	75.25	0.76	0.75	0.75	75.25	0.76	0.75	0.75
BG	76.24	0.76	0.76	0.76	76.24	0.76	0.76	0.76
BN	77.23	**0.78**	0.77	0.77	77.23	**0.78**	0.77	0.77
CVR	76.90	0.77	0.77	0.77	76.90	0.77	0.77	0.77
DS	74.26	0.74	0.74	0.74	74.26	0.74	0.74	0.74
FC	**77.89**	**0.78**	**0.78**	**0.78**	**77.89**	**0.78**	**0.78**	**0.78**
HT	76.90	0.77	0.77	0.77	76.90	0.77	0.77	0.77
IBk	69.64	0.70	0.70	0.70	69.64	0.70	0.70	0.70
ICO	77.23	0.77	0.77	0.77	77.23	0.77	0.77	0.77
J48	75.91	0.76	0.76	0.76	75.91	0.76	0.76	0.76
JRip	75.58	0.76	0.76	0.76	75.58	0.76	0.76	0.76
KS	72.61	0.73	0.73	0.72	72.61	0.73	0.73	0.72
LB	76.90	0.77	0.77	0.77	76.90	0.77	0.77	0.77
LWL	74.92	0.75	0.75	0.75	74.92	0.75	0.75	0.75
MCC	76.24	0.76	0.76	0.76	76.24	0.76	0.76	0.76
MCCU	75.91	0.76	0.76	0.76	75.91	0.76	0.76	0.76
NB	76.90	0.77	0.77	0.77	76.90	0.77	0.77	0.77
OR	72.28	0.72	0.72	0.72	72.28	0.72	0.72	0.72
PART	72.28	0.73	0.72	0.72	72.28	0.73	0.72	0.72
RC	70.63	0.71	0.71	0.71	70.63	0.71	0.71	0.71
REPT	73.93	0.74	0.74	0.74	73.93	0.74	0.74	0.74
RF	73.60	0.74	0.74	0.74	73.60	0.74	0.74	0.74
RFC	68.32	0.69	0.68	0.68	68.32	0.69	0.68	0.68
RSS	75.25	0.75	0.75	0.75	75.25	0.75	0.75	0.75
RT	66.01	0.66	0.66	0.66	66.01	0.66	0.66	0.66
SGD	75.91	0.76	0.76	0.76	75.91	0.76	0.76	0.76
SL	75.25	0.75	0.75	0.75	75.25	0.75	0.75	0.75
SMO	76.24	0.76	0.76	0.76	76.24	0.76	0.76	0.76
Average	**74.52**	**0.75**	**0.75**	**0.74**	**74.52**	**0.75**	**0.75**	**0.74**

REFERENCES

[1] E. Elbasi and A. I. Zreikat, "Heart Disease Classification for Early Diagnosis based on Adaptive Hoeffding Tree Algorithm in IoMT Data," *International Arab Journal of Information Technology*, vol. 20, no. 1, pp. 38–48, 2023, doi: 10.34028/iajit/20/1/5.

[2] A. S. C. Bose and V. Ramesh, "Highly Accurate Grey Neural Network Classifier for an Abdominal Aortic Aneurysm Classification Based on Image Processing Approach," *International Arab Journal of Information Technology*, vol. 20, no. 2, pp. 215–223, 2023, doi: 10.34028/iajit/20/2/8.

[3] M. G. Sghaireen *et al.*, "Machine Learning Approach for Metabolic Syndrome Diagnosis Using Explainable Data-Augmentation-Based Classification," *Diagnostics*, vol. 12, no. 12, pp. 1–23, 2022, doi: 10.3390/diagnostics12123117.

[4] G. Samara, "Intelligent reputation system for safety messages in VANET," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 3, pp. 439-447, 2020, doi: 10.11591/ijai.v9.i3.pp439-447.

[5] H. A. Owida, N. L. Kuiper, and Y. Yang, "Maintenance and acceleration of pericellular matrix formation within 3D cartilage cell culture models," *Cartilage*, vol. 13(2_suppl), pp. 847S-861S, 2021, doi: 10.1177/1947603519870839.

[6] A. Al Smadi, A. Abugabah, S. F. Mahenge, and F. Shahid, "Information Systems in Medical Settings: A Covid-19 Detection System Using X-Ray Scans," *ICSEC 2022 - International Computer Science and Engineering Conference 2022*, pp. 72–77, 2022, doi: 10.1109/ICSEC56337.2022.10049310.

[7] S. Gunasekaran and S. Vivekasaran, "Disease Prognosis of Fetal Heart's Four-Chamber and Blood Vessels in Ultrasound Images Using CNN Incorporated VGG 16 and Enhanced DRNN," *The International Arab Journal of Information Technology (IAJIT)*, vol. 21, no. 6, pp. 1111-1127, November 2024, doi: 10.34028/iajit/21/6/13.

[8] "Cardiovascular diseases (CVDs)," WHO, 2021, Accessed: Jan. 02, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[9] A. Y. Alhusenat, H. A. Owida, H. A. Rababah, J. I. Al-Nabulsi, and S. Abuowaida, "A Secured Multi-Stages Authentication Protocol for IoT Devices," *Mathematical Modelling of Engineering Problems*, vol. 10, no. 4, pp. 1352–1358, 2023, doi: 10.18280/mmep.100429.

[10] M. Alzyoud *et al.*, "Diagnosing diabetes mellitus using machine learning techniques," *International Journal of Data and Network Science*, vol. 8, no. 1, pp. 179–188, 2024, doi: 10.5267/j.ijdns.2023.10.006.

[11] A. Al-qerem *et al.*, "Enhancing Stroke Prediction Using Generative Adversarial Networks for Intelligent Medical Care," *International Journal of Crowd Science*, pp. 1–13, 2024.

[12] H. A. Owida, O. S. M. Hemied, R. S. Alkhawaldeh, N. F. F. Alshdaifat, and S. F. A. Abuowaida, "Improved Deep Learning Approaches for Covid-19 Recognition in Ct Images," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 13, pp. 4925–4931, 2022.

[13] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," *IEEE Access*, vol. 8, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[14] H. A. Owida, N. M. Turab, and J. Al-Nabulsi, "Carbon nanomaterials advancements for biomedical applications," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 2, pp. 891-901, 2023, doi: 10.11591/eei.v12i2.4310.

[15] H. Jindal, S. Agrawal, R. Khera, R. Jain, and P. Nagrath, "Heart disease prediction using machine learning algorithms," in *IOP*

conference series: materials science and engineering, 2021, pp. 1–11, doi: 10.1088/1757-899X/1022/1/012072.
[16] D. E. Salhi, A. Tari, and M. T. Kechadi, "Using Machine Learning for Heart Disease Prediction," in *Lecture Notes in Networks and Systems*, 2021, pp. 70–81. doi: 10.1007/978-3-030-69418-0_7.
[17] M. S. Al-Batah, M. S. Alzboon, and R. Alazaidah, "Intelligent Heart Disease Prediction System with Applications in Jordanian Hospitals," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 9, pp. 508–517, 2023, doi: 10.14569/IJACSA.2023.0140954.
[18] "Heart-Disease-Detection," Github, 2023, Accessed: Jan. 02, 2023. [Online]. Available: https://github.com/mrinmoyxb/Heart-Disease-Detection/blob/main/heart_disease_data.csv.
[19] "Heart Disease Dataset," Kaggle, 2022, [Online]. Available: https://www.kaggle.com/datasets/yasserh/heart-disease-dataset. Accessed: Jan. 02, 2023.
[20] "Heart Failure Prediction Dataset," Kaggle, 2022, Accessed: Jan. 02, 2023, [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction.
[21] M. Lu *et al.*, "Persistence of severe global inequalities in the burden of Hypertension Heart Disease from 1990 to 2019: findings from the global burden of disease study 2019," *BMC Public Health*, vol. 24, no. 1, pp. 1–10, 2024, doi: 10.1186/s12889-023-17573-9.
[22] A. Al-Momani, M. N. Al-Refai, S. Abuowaida, M. Arabiat, N. Alshdaifat, and M. N. A. Rahman, "The effect of technological context on smart home adoption in Jordan," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 2, pp. 1186–1195, 2024, doi: 10.11591/ijeecs.v33.i2.pp1186-1195.
[23] M. Alluwaici, A. K. Junoh, and R. Alazaidah, "New Problem Transformation Method Based on the Local Positive Pairwise Dependencies among Labels," *Journal of Information and Knowledge Management*, vol. 19, no. 1, 2020, doi: 10.1142/S0219649220400171.
[24] K. Brown *et al.*, "Using Artificial Intelligence for Rheumatic Heart Disease Detection by Echocardiography: Focus on Mitral Regurgitation," *Journal of the American Heart Association*, vol. 13, no. 2, 2024, doi: 10.1161/JAHA.123.031257.
[25] A. A. L. Smadi, A. Abugabah, A. Mehmood, and S. Yang, "Brain Image Fusion Approach based on Side Window Filtering," *Procedia Computer Science*, vol. 198, pp. 295–300, 2021, doi: 10.1016/j.procs.2021.12.243.
[26] H. A. Owida, "Recent Biomimetic Approaches for Articular Cartilage Tissue Engineering and Their Clinical Applications: Narrative Review of the Literature," *Advances in Orthopedics*, vol. 1, 2022, doi: 10.1155/2022/8670174.

BIOGRAPHIES OF AUTHORS

Mohammad Al-Hababhbeh received the B.Sc. degrees in computer science from Zarqa University. He is currently pursuing the master degree with the School of Computer Sciences from Zarqa University. His research interests in machine learning. He can be contacted at email: 20229537@zu.edu.jo.

Moath Alomari an accomplished Oracle database administrator with over 12 years of experience, completed bachelor's degree in computer science at Mu'tah University from 2008 to 2012. Currently pursuing a master's degree in computer science at Zarqa University, he is deeply passionate about data science, artificial intelligence, machine learning, and data analytics. His extensive professional background and academic pursuits reflect his commitment to pushing the boundaries of technology and exploring innovative solutions in the field of computer science. He can be contacted at email: 20229722@zu.edu.jo.

Hebatullah Khattab received a bachelor, master, and Ph.D. from the University of Jordan in 2001, 2006, and 2020, respectively. All degrees are in computer science. Currently, she is the head of the Department of Computer Science at Zarqa University. Her research interests include machine learning, image processing, and parallel processing. She can be contacted at email: h.awwad@zu.edu.jo.

Raed Alazaidah received the B.Sc. degrees in computer information system, M.Sc. degrees in computer science from Al al-Bayt University, Jordan, in 2012 and 2015, respectively, and the Ph.D. degree in computer science from Universiti Sains Malaysia, Malaysia, in 2023. He can be contacted at email: razaidah@zu.edu.jo.

Nawaf Alshdaifat received the B.Sc. degrees in computer science from Al al-Bayt University and M.Sc. degrees in computer science from the University of Jordan, Jordan, in 2002 and 2011, respectively, and the Ph.D. degree in computer science from Universiti Sains Malaysia, Malaysia, in 2023. His research interests include deep learning and machine learning. He can be contacted at email: n_alshdaifat@asu.edu.jo.

Suhaila Abuowaida received the B.Sc. degrees in computer information system, M.Sc. degrees in computer science from Al al-Bayt University, Jordan, in 2012 and 2015, respectively, and the Ph.D. degree in computer science from Universiti Sains Malaysia, Malaysia, in 2023. Her research interests include deep learning, depth estimation, point cloud, and computer vision. She can be contacted at email: sabuoweuda@zu.edu.jo.

Saleh Alqatan recived his Ph.D. in computer science from the Malaysia University Terengganu, Malaysia. He received his M.Sc. in information technology from the University Utara Malaysia (UUM) in 2010, Malaysia. He is now working as an assistant professor in the Department of computer science at the Zarqa University, Jordan. His research interest: software engineering, mobile learning, and system quality. He can be contacted at email: salqatan@zu.edu.jo.

Mohammad Arabiat received the B.Sc. degrees in computer science from National Technical University Kharkov Poly Technical Institute, M.Sc. degrees in computer science from Kharkov National University of Economics, Ukraine, in 2002 and 2003, respectively and the Ph.D. degree in information technology from V.N. Karazin, Ukraine in 2014. His research interests include computer information system. He can be contacted at email: marabiat@zu.edu.jo.