Supervised Learning Based Classification of Cardiovascular Diseases

Arif Hussain[1], Hassaan Malik[1, 2, *], Muhammad Umar Chaudhry[1, 3, 4]

[1]Department of Computer Science, National College of Business Administration and Economics, Multan, Pakistan

[2]Department of Computer Science, University of Management and Technology, Lahore, Pakistan

[3]Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Korea

[4]AiHawks, Multan, Pakistan

Abstract

Detecting cardiovascular disease (CVD) in the early stage is a difficult and crucial process. The objective of this study is to test the capability of machine learning (ML) methods for accurately diagnosing the CVD outcomes. For this study, the efficiency and effectiveness of four well renowned ML classifiers, i.e., support vector machine (SVM), logistics regression (LR), naive Bayes (NB), and decision tree (J48), are measured in terms of precision, sensitivity, specificity, accuracy, Matthews correlation coefficient (MCC), correctly and incorrectly classified instances, and model building time. These ML classifiers are applied on publically available CVD dataset. In accordance with the measured result, J48 performs better than its competitor classifiers, providing significant assistance to the cardiologists.

Keywords: cardiovascular disease, machine learning, artificial intelligence

1. Introduction

Cardiovascular disease (CVD), which is considered to be one of the most fatal diseases, includes a cluster of diseases such as aortic aneurysm, angina, rheumatic heart disease, coronary heart disease, congenital heart disease, stroke, peripheral arterial disease, etc. [1]. CVD affects heart muscles (myocardial infarction) and vasculatures (hypertension). The only two methods to reduce the occurrence of fatal events due to CVD are reducing cholesterol intake and increasing proper exercise.

The detection of CVD in the early stage is a difficult and crucial step. CVD are complex and heterogeneous, as it is caused by multiple factors including environmental and behavioral factors. Artificial intelligence (AI) has an ability to exploit the big data largely used in patient care. The big data refers to the large volume and variety of the data that are complex in nature. Conventional data processing methods and techniques are insufficient to effectively examine a large amount of data [2]. The big data generally displays high capacity, reliability, and various types of features. As a result, scalability and accuracy is improved by using the big data techniques as compared to the traditional methods such as regression and correlation-based models. The big data is based on four Vs [3]: volume, velocity, variety, and veracity. The techniques of AI and machine learning (ML) in the big data have achieved tremendous success that helps commerce and fabrications, in order to highlight the hidden patterns and to predict future prospects. Researchers use AI methods to track down the disease related contexts, e.g., controlling tuberculosis [4] and predicting extended-spectrum β-lactamases (ESBL) [5] and heart diseases [6].

* Corresponding author. E-mail address: f2019288004@umt.edu.pk

  Tel.: +92-333-3873355

ML methods are not only used for disease classification, but also widely used in industries. ML based framework are effectively used for evaluating the data collected from smart meters to analyze the original and fake data [7]. Several types of ML classifiers are also used for the identification of chatter vibration [8]. In addition, ML techniques are also utilized to distinguish cyber-attacks to perform the paradigm and authentication [9].

ML algorithm is not a panacea for prediction related tasks. Even a perfect ML model depends upon the quality and magnitude of the dataset from which it is trained. There are variant kinds of algorithms for classification and prediction of CVD outcomes. This study provides a comparative analysis among the performance matrices of ML classifiers, i.e., support vector machine (SVM), logistics regression (LR), naive Bayes (NB), and decision tree (J48), using the CVD dataset in diagnosing and making significant decisions to assist the clinical and health experts. The objective of this study is to evaluate and achieve the best accuracy with the lowest error rate in diagnosing the disease. Consequently, the efficiency and effectiveness of those approaches in terms of precision, sensitivity, specificity, accuracy, correctly and incorrectly classified instances, and model building time are evaluated.

In summary, the main contribution of this study is discussed below:

(1) The four ML classifiers applied for CVD diagnosis and the detailed state-of-the-art ML classifiers are discussed.

(2) The comparative analysis of ML classifiers is observed in terms of the performance evaluation matrices for CVD classification.

(3) J48 model achieves the highest classification accuracy.

2. Related Work

AI is a comprehension of mathematical designing that has the capability to improve health services by using new health care strategies and procedures. In ML, classification is considered to be the most important and essential part. A lot of research has been conducted by using ML to classify CVD from health care related datasets, proving that ML has good performance.

Haq et al. [6] measured the efficiency of seven algorithms for ML, i.e., K-nearest neighbors (KNN), LR, decision tree (DT), NB, artificial neural network (ANN), SVM, random forest (RF), and measured the efficiency of three algorithms for selecting features, i.e., relief feature selection, minimal-redundancy-maximal-relevance (mRMR), and least absolute shrinkage and selection operator (LASSO) to analyze the best classifier for heart diseases. LR is proven to be the most accurate classifier with the accuracy of 84% before applying feature selection. SVM achieves the accuracy of 80% after applying feature selection algorithms.

Li et al. [10] applied different ML approaches to classify heart diseases. They proposed a diagnostic system that was accurate and efficient in producing the significant results. The heart disease dataset was pre-processed by removing the null values. They applied feature selection techniques to find the appropriate outputs. The result of experiment clarifies that the proposed feature selection algorithm (FCM/M) is suitable to diagnose the heart disease with SVM. As compared to previous methods, the proposed system (SRM-FCM/M) achieves better precision. Latha and Jeeva [11] presented a method of ensemble classification on Cleveland heart dataset to increase the accuracy of weak algorithms by combining multiple classifiers [12]. The study found that ensemble techniques such as bagging and boosting effectively improved the predicted accuracy of weak classifiers in identifying the risk of heart disease. A maximum increase of 7% accuracy was achieved with the help of ensemble classification.

Raza et al. [13] evaluated different ML techniques and combined them with different feature selection methods to find the best technique for classification [14]. They applied four heart disease datasets and different methods like principal component analysis, chi-squared testing, relief, and symmetrical uncertainty to create distinctive feature sets [15]. By doing this, they achieve the significant accuracy of 85% on all datasets. Alaa et al. [16] performed a comparison analysis of the ML techniques

based on AutoPrognosis framework, traditional techniques, and non-traditional techniques on the improvement of CVD and risk predictions. The research found that the ML based AutoPrognosis framework has the highest accuracy rate for CVD risk prediction as compared to traditional and non-traditional techniques. Soni [17] compared different methods for heart disease prediction. The author combined ML methods with data mining techniques to be used in clinical decision support system. NB, sequential minimal optimization (SMO), RF, decision table algorithms were used for the detection of heart diseases. Few parameters including electrocardiogram (ECG) results were also employed. Moreover, the author focused on multiparameter classification using WEKA tool. Different parameters are analyzed, including ECG and heart rate variability (HRV). By using these data, NB and DT achieved good diagnostics accuracy of both 85% in predicting heart diseases. Fitriyani et al. [18] proposed a method for heart disease diagnosis at the initial stage. They designed a module named clinical decision support systems (CDSS) for heart disease prediction. Moreover, they also conducted a comparative analysis with other medical modules and achieved the classification accuracy of 95.90% and 98.40% for Statlog and Cleveland datasets, respectively.

3. Materials and Methods

3.1. Dataset

For this study, CVD dataset [19] is used to develop a CVD diagnosis system based on ML. The sample size of the dataset is 70 thousand patients containing 11 features and one target value. The input features have three types: objective (realistic information), examination (result of medical checkup), and subjective (patient's information). There are two classes on the target output mark, which reflect whether a cardio patient or a regular subject is the individual entity being examined. Therefore, the extracted dataset consists of a matrix of 70k/12 attributes. The complete information and definition of 70k instances with 12 dataset characteristics is given in Table 1.

Table 1 Feature information of CVD dataset

Id	Parameters	Short code	Description
1	Age	Ag	Age in days
2	Height	Hgt	Height in cm
3	Weight	Wgt	Weight in kg
4	Gender	Gen	Male = 2; female = 1
5	Systolic blood pressure	sb_hi	Upper value (mmHg)
6	Diastolic blood pressure	db_lo	Lower value (mmHg)
7	Cholesterol	Chl	1: normal; 2: above normal; 3: well above normal
8	Glucose	Glu	1: normal; 2: above normal; 3: well above normal
9	Smoking	SMK	1 = yes; 0 = no
10	Alcohol	ALC	1 = yes; 0 = no
11	Physical activity	Act	1 = yes; 0 = no
12	Presence or absence of CVD	Cardio	1 = yes; 0 = no

3.2. The proposed methodology

The proposed framework is designed (Fig. 1) to classify cardio patients. For this study, the performance of variant kinds of ML predictive models for CVD diagnosis on full features of datasets is tested. The well renowned ML classifiers, i.e., SVM, LR, NB, and DT (J48), are applied in the system. The model's validation and performance evaluation matrices are computed. The methodology of the proposed system is divided into four phases. In the first phase, the pre-processing of CVD dataset is done by removing the duplicates and null values. Secondly, the dataset is divided with the ratio of 85% for training and 15% for testing process. Predictive ML classifiers are applied in the third phase. In the fourth and the final phase, the classifier's performance is calculated in terms of sensitivity, specificity, Matthews correlation coefficient (MCC), and F1-score.
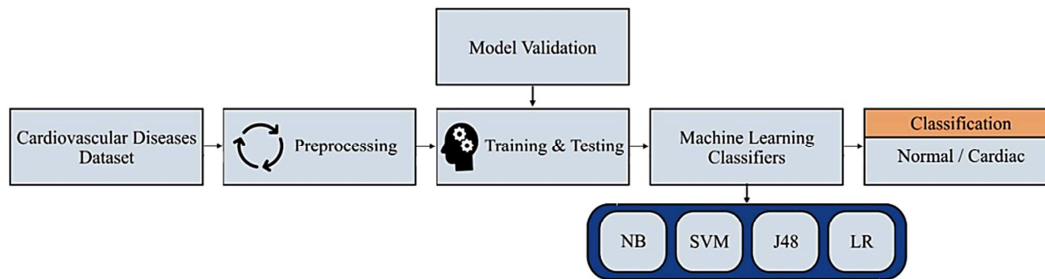
Fig. 1 Framework for CVD prediction

3.3. Dataset pre-processing

Dataset pre-processing is a crucial step for achieving the significant outputs of ML classifiers, which should be trained and tested in a powerful manner. Pre-processing techniques such as standard scalar and removal of missing values are enforced to the dataset for adequate use in the classifiers. All these data pre-processing techniques are used in this study.

3.4. Machine learning classifiers

Classification is the process of predicting the class of given data points. ML classification algorithms are used to classify the data into their respective domains. In this study, four well-known predictive ML classifier algorithms with their hypothetical background are presented.

3.4.1. Logistic regression

LR algorithm is used to observe the discrete set of classes [20]. In LR, binary and multiclass classification methods are performed. In binary classification, the prediction is completed by the predictive variable y when $Y \in [0, 1]$, where 0 is known as negative class and 1 is positive class. In multiclass classification, the prediction is applied to the predictive value y when $Y \in [0, 1, 2, 3]$. In order to classify binary classes $[0, 1]$, the hypothesis $H\theta = \theta^T X$ is devised, and the threshold output value is $H\theta$ at 0.5. If the hypothesis $H\theta \geq 0.5$, it will predict $y = 1$; if the value is less than 0.5, it will predict $y = 0$. The hypothesis tends to limit the cost function between 0 and 1, e.g., $0 \leq H\theta \leq 1$.

The sigmoid function of LR can be written as:

$$H\theta(X) = \frac{1}{1+\ell^{-x}} f \tag{1}$$

LR is applied on the dataset to measure the probability of the occurrence of binary outcomes such as healthy (0) and cardiac (1) patient. Therefore, Eq. (1) of the sigmoid function is used to convert the outcomes into the categorical values.

Similarly, the cost function of LR can be written as:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{1} \text{cost}\left[H\theta(x^i, y^i) \right] \tag{2}$$

Eq. (2) shows that the cost function takes two parameters in the input. $H\theta(x^i)$ is hypothesis function, $H\theta(y^i)$ is the output, and m represents the number of samples. The cost function is used to measure how badly the LR model is performing for predicting the relationship between y and x. Before fitting the parameters to training the data with Eq. (2), a constraint is added to prevent the model from overfitting. However, gradient descent (GD) as the optimization algorithm is applied by minimizing $J(\theta)$ as a function of θ to find optimal parameters. It takes partial derivative of J with respect to θ, and updates θ via each iteration with a selected learning rate of 0.5 until GD is converged.

3.4.2. *Support vector machine*

SVM is considered to be the most powerful ML technique. SVM is a supervised learning algorithm which is mostly used for classification problems as support vector classification (SVC) [21]. This approach seeks to classify hyperplanes that are capable of unravelling datasets into elevated-dimensional spaces of features. The margin is called the distinction between datasets, and the SVM maximizes the margin [22]. When dealing with binary classification problems, the instances are alienated with the hyperplane $w^t \times x + b = 0$, where w is known as dimensional coefficient vector, b is the offset value, and x is the dataset value. Some of the popular kernel choices are linear $(x_i \times x_i)$, polynomial $[(x_i \times x_i + c)^d,\ c > 0,\ d \geq 2]$, and radial basis function $\exp(-y \parallel x_i - x_j \parallel^2)$ [23]. The purpose of using SVM is to identify a hyperplane in binary dimensional space that distinctly classifies the data points. Therefore, to separate the binary classes of data points, hyperplanes use the maximum distance between data points of both classes. The linear separable hyperplane is used for decision boundaries that assist in classifying the data points. The data points falling on either side of the hyperplane can be attributed to different classes. The size of the hyperplanes is dependent on the numbers of input features. When there is a misclassification done by the model on the prediction of the class of the data point, the loss is included along with the regularization parameter to perform gradient update. However, the required features are extracted from the dataset and split into the training and testing data. The learning rate of the model is set to 0.0001 and the regularization parameter λ is set to 1/epochs. Therefore, the regularizing value reduces the number of epochs increased.

3.4.3. *Naive Bayes*

NB is an algorithm for supervised learning based on the conditional probability theorem to evaluate the class function vector [24]. NB algorithm assumes that all the variables in the dataset are not correlated to each other. It can be written as:

$$P(c \mid x) = \frac{P(x \mid c) \times P(c)}{P(x)}$$
(3)

where $P(c|x)$ is the posterior probability of class c; $P(c)$ is the probability of class c; $P(x|c)$ is the likelihood probability of the predictor given class c; $P(x)$ is the prior probability of predictor.

When working with the dataset, the values of the dataset associated with each class are distributed according to a normal distribution. The developing process of NB model is that the data is defined by a normal distribution with no co-variance between dimensions. This NB model is fit by finding the mean and standard deviation of the points within each label. The CVD database of various patients is based on 11 columns including gender, age, height, and weight (Table 1). The target is whether the patient is normal or having cardiac problems. If the targeted column $P(c)$ has value 1, then it means that the person has heart problems, and if the value is 0, then it means that the person is healthy.

3.4.4. *Decision tree classifier*

The DT (J48) supervised algorithm is based on tree shape structures where every node is a decision node [25]. Decisions are taken on the basis of an algorithmic method that identifies the ways to divide a dataset based on diverse conditions [26]. In DT, internal and external nodes are linked together, which make the decision done easily and efficiently.

4. Results and Discussion

This section contains the outcomes produced by the ML classifiers. The outcomes of ML classifiers are evaluated in terms of different performance evaluation matrices. The cross-validation (CV) approach is applied, and the performance matrices of each ML classifiers are also calculated. The details are given in the following subsections.

4.1. K-cross-validation

CV is probably considered to be the most important techniques used to validate the stability of the ML model on new data [27]. In *k*-fold CV, the dataset is separated into *k* equal size in which *k*-1 groups are used for training and the rest is for testing purposes. The performance of the classifiers is based on the *k* results. In this study, *k* =15 CV is applied, which means that 85% data is applied for training process and the rest is for testing the classifiers.

4.2. Performance evaluation matrices

Various performance evaluation matrices are used in this study. The binary confusion matrix 2×2 is used because two categories are included in the proposed output variables. In addition, it provides two types of correct and incorrect prediction. A confusion matrix is shown in Fig. 2.

Fig. 2 Confusion matrix

Classification accuracy shows the performance, and the classification error shows the incorrect classification of the model. They can be written as:

$$Classification\ Error = \frac{FP + FN}{TP + TN + FP + FN} \tag{4}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Sensitivity is also known as "true positive (TP) rate". It confirms that if a diagnostic test is positive, then the subject is a cardio patient. It can be written as:

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \tag{6}$$

Specificity is the "true negative (TN) rate" classifier for detecting negative cases. When the diagnostic test is negative, the person is healthy. It can be mathematically written as:

$$Specificity = \frac{TN}{TN + FP} \times 100\% \tag{7}$$

The mathematical expression of precision is written as:

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{8}$$

MCC represents the prediction ability of a classifier with the values which vary between $[-1, +1]$. If MCC classifier has the value +1, this indicates that the predictions of the classifier are ideal; if MCC classifier has the value -1, it indicates that the classifier produces incorrect predictions; the value near 0 implies that random predictions are created by the classifier. MCC's mathematical expression is written as:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \qquad (9)$$

Receiver operating characteristic curve (ROC) analysis compares the TP rate and false positive (FP) rate in the classification results while the area under the curve (AUC) characterizes the ROC of classifier. The larger the value of AUC is, the more effective the performance of the classifier is.

4.3. Results and discussion

This section contains the outcomes generated by the classification models in terms of variant kinds of performance evaluation matrices. Four different ML algorithms are tested for accuracy, i.e., NB, LR, J48, and SVM on full features of CVD dataset. All computations are performed in WEKA software on an Intel® Core™ i7-6700 CPU @ 3.40 GHz personal computer with installed memory (RAM) of 16 GB. The 15-fold CV test is applied, and the data is analyzed visually to figure out the distribution of values in terms of effectiveness and efficiency. The effectiveness of all four classifiers in terms of model building time, accuracy, correctly and incorrectly classified instances is calculated (Table 2). The results of the 15-fold performance assessment of four classifiers NB, LR, J48, and SVM on CVD dataset are shown in Table 2.

The processing time of each classifier with NB is 0.16 seconds in Fig. 3, which is very fast computationally compared to the LR, J48, and SVM. Furthermore, the processing time of SVM is 148.93 seconds which indicates that SVM is a lazy learner. From comparative graph (Fig. 4), it is easily noted that J48 has the highest value of correctly classified instances and the lower value of incorrectly classified instances while NB has the lowest value of correctly instances and the higher value of incorrectly classified instances.

In Table 3, NB shows poor performance that has 58.87% classification accuracy, 27.77% specificity, 58.9% sensitivity, and 22.6% MCC. The NB regression specificity value is 27.77%, indicating that the likelihood of diagnostic tests is negative, and the individual does not have CVD. In addition, 58.9% sensitivity shows the likelihood that the diagnostic test is positive for the person. For LR, the experiment is performed by using 15-fold CV. However, the performance of LR is good as shown in Table 3.

Table 2 Evaluation of classifiers in term of time, correctly instances, and incorrectly instances

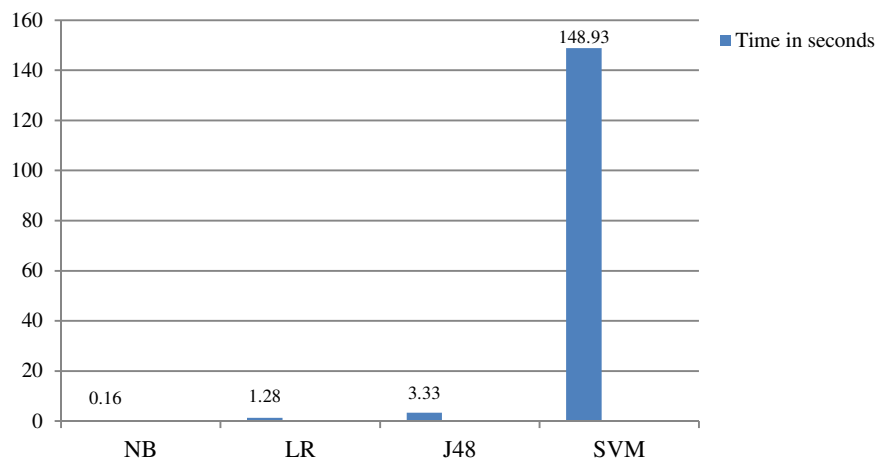| Evaluation criteria	Classifiers			
	NB	LR	J48	SVM
Time taken to build a model	0.1	1.28	3.33	148.93
Correctly classified instances	41209	50502	51159	45405
Incorrectly classified instances	28791	19498	18841	24595

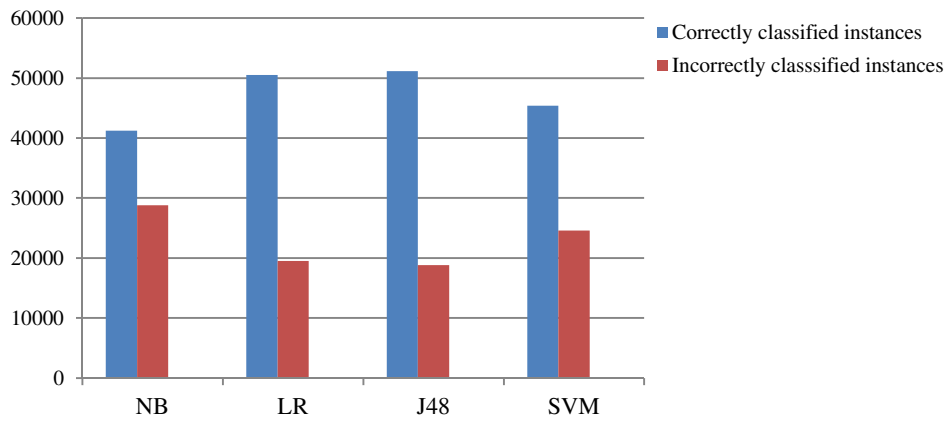Fig. 3 Classifier processing time in seconds

Fig. 4 Comparative graph of correctly and incorrectly classified instances

Table 3 Classifier 15-fold CV classification performance evaluation

Evaluation criteria	Classifiers			
	NB	LR	J48	SVM
Accuracy (%)	58.87	72.1457	73.0843	64.8643
Specificity (%)	27.77	67.79	70.19	59.87
Sensitivity (%)	58.9	72.1	73.1	64.9
MCC	0.226	0.445	0.462	0.299
Precision (%)	64.4	73.3	73.2	65

The LR classifier achieves 72.1457% accuracy, 67.79% specificity, 72.1% sensitivity, and 44.5% MCC. Similarly, SVM has the 64.86% classification accuracy, 59.87% specificity, 64.9% sensitivity, and 29.9% MCC. The performance of J48 is excellent which achieves the highest 73.0843% classification accuracy, the best 70.19% specificity, 73.1% sensitivity and 46.2% MCC value. The output of J48 outperforms the other three classifiers in terms of precision, sensitivity, specificity, and MCC, as shown in Fig. 5. J48 has 73.0843% predictive accuracy, 73.1%sensitivity, and 70.19% specificity. LB, which has 72.1547% classification precision, is the second significant classifier. The worst performance is observed in terms of precision, sensitivity, and specificity for NB classifiers.

In Fig. 6, the accuracy 73.0843% obtained by J48 is greater than the accuracy obtained by NB, LR, and SVM (which varies between 58.87% and 72.1457%). It is observed that J48 and LR have high MCC values, while NB and SVM have the lowest MCC values with 15-fold CV on the full feature's dataset, as depicted in Fig. 7. Therefore, it means that J48 and LR are the highest prediction ability classifiers as compared to NB and SVM. Simulation errors are found to test the better performance of classifiers. To do so, the measured efficacy of all four classifiers is in provisos of kappa statistic (KS), mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), and root relative squared error (RRSE). KS, MAE, and RMSE show numeric values while RAE and RRSE are in percentage. The detailed results are shown in Table 4. The graphical representation for simulation of errors is presented in Fig. 8.
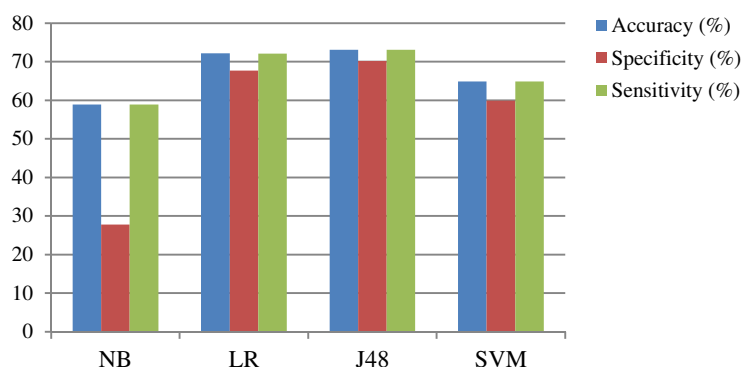

Fig. 5 Performance in terms of precision, specificity, and sensitivity of classifiers
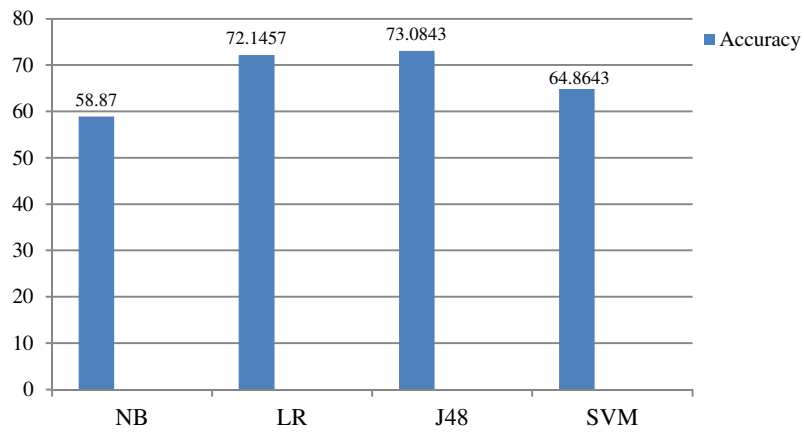
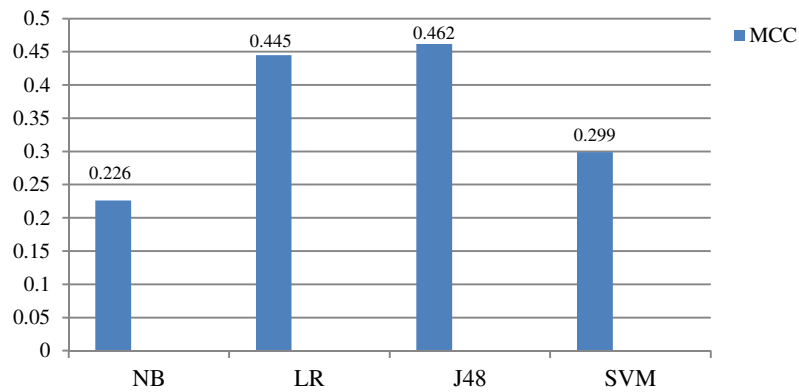Fig. 6 Classifier accuracy test with 15-fold CV on complete dataset features

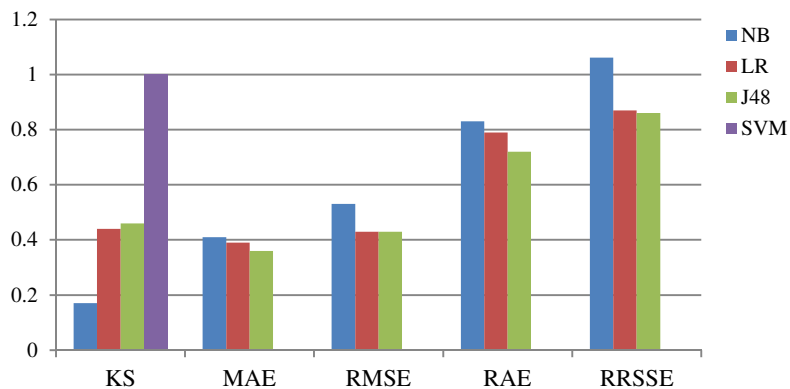Fig. 7 MCC of all four classifiers with 15-fold CV

Fig. 8 Compared simulation error with respect to KS, MAE, RMSE, RAE, and RRSE

Table 4 Summary of simulation errors

Evaluation criteria	Classifiers			
	NB	LR	J48	SVM
KS	0.17	0.44	0.4617	0.29
MAE	0.41	0.39	0.3611	0.35
RMSE	0.53	0.43	0.43	0.59
RAE (%)	83.85	79.09	72.21	70.27
RRSE (%)	106.10	87.60	86.91	118.55

From Table 4, the outcomes represent the chance of having a best classification (0.4617) with the least warning error rate (0.3611) formed by J48. Furthermore, the second-best classification (0.4429) with error rate (0.3955) is produced by LR. NB and SVM have the highest value of error rate, which explain the large number of incorrectly classified instances for each algorithm. The value of MAE is observed to evaluate the quality of the classifiers. From Table 4, it is clearly identified that J48 and SVM have minimum MAE values as compared to other classifiers NB and LR (Fig. 9).
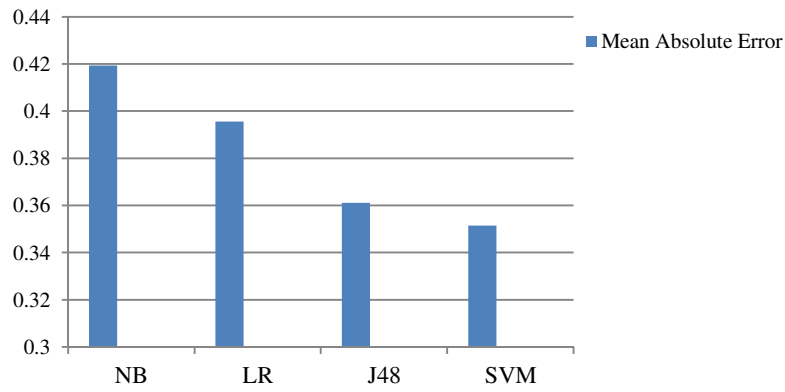
Fig. 9 MAE for NB, LR, J48, and SVM

J48 and SVM have the nearest predictive values to the target values as compared to LR and NB. The confusion matrix is used for evaluating the performance of a classification model. The confusion matrix of all four classifiers is presented in Fig. 10 below.
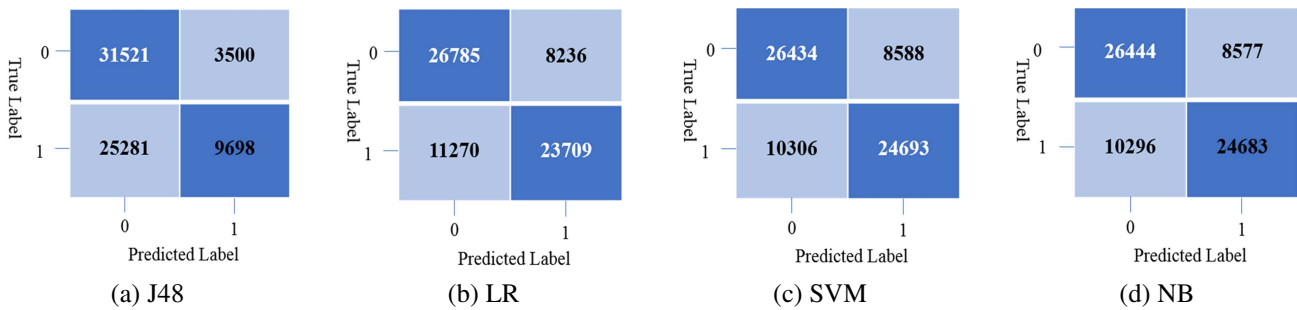

| (a) J48 | (b) LR | (c) SVM | (d) NB |

Fig. 10 Confusion matrix

5. Conclusions and Future Work

In this modern era, the amount of data collected from health surveillance has rapidly increased with the advancement of the medical technology. It is foreseeable that AI platform will enable the analysis of multiple infectious diseases including CVD, which is considered to be one of the main causes of death worldwide. In this study, variable generalization performance of supervised learning algorithms for CVD diagnosis was evaluated on CVD dataset by using four well-known supervised ML classifiers. In accordance with the obtained output, the higher probability of achieving an improved result can be inferred in terms of the accuracy of the CVD classification when J48 is applied. In the future work, the ensemble of pre-trained models would be applied for classifying CVD.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] J. Mackay, G. A. Mensah, S. Mendis, and K. Greenlund, The Atlas of Heart Disease and Stroke, Geneva: World Health Organization, 2004.
[2] Z. S. Wong, J. Zhou, and Q. Zhang, "Artificial Intelligence for Infectious Disease Big Data Analytics," Infection, Disease, and Health, vol. 24, no. 1, pp. 44-48, February 2019.
[3] S. Schneeweiss, "Learning from Big Health Care Data," The New England Journal of Medicine, vol. 370, no. 23, pp. 2161-2163, June 2014.
[4] A. Blumenthal, "Artificial Intelligence to Fight the Spread of Infectious Diseases," https://phys.org/news/2018-02-artificial-intelligence-infectious-diseases.html, February 20, 2018.

[5]   K. E. Goodman, J. Lessler, S. E. Cosgrove, A. D. Harris, E. Lautenbach, J. H. Han, et al., "A Clinical Decision Tree to Predict Whether a Bacteremic Patient is Infected with an Extended-Spectrum β-Lactamase-Producing Organism," Clinical Infectious Diseases, vol. 63, no. 7, pp. 896-903, October 2016.

[6]   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," Mobile Information Systems, vol. 2018, 3860146, December 2018.

[7]   M. Elsisi, K. Mahmoud, M. Lehtonen, and M. M. Darwish, "Reliable Industry 4.0 Based on Machine Learning and IoT for Analyzing, Monitoring, and Securing Smart meters," Sensors, vol. 21, no. 2, 487, January 2021.

[8]   M. Q. Tran, M. K. Liu, and M. Elsisi, "Effective Multi-Sensor Data Fusion for Chatter Detection in Milling Process," The International Society of Automation Transactions, in press.

[9]   M. Elsisi, M. Q. Tran, K. Mahmoud, D. E. A. Mansour, M. Lehtonen, and M. M. Darwish, "Towards Secured Online Monitoring for Digitalized GIS Against Cyber-Attacks Based on IoT and Machine Learning," Institute of Electrical and Electronics Engineers Access, vol. 9, pp. 78415-78427, June 2021.

[10] 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," Institute of Electrical and Electronics Engineers Access, vol. 8, pp. 107562-107582, June 2020.

[11] C. B. C. Latha and S. C. Jeeva, "Improving the Accuracy of Prediction of Heart Disease Risk Based on Ensemble Classification Techniques," Informatics in Medicine Unlocked, vol. 16, 100203, July 2019.

[12] K. W. Johnson, J. Torres Soto, B. S. Glicksberg, K. Shameer, R. Miotto, M. Ali, et al., "Artificial Intelligence in Cardiology," Journal of the American College of Cardiology, vol. 71, no. 23, pp. 2668-2679, June 2018.

[13] A. Raza, A. Mehmood, S. Ullah, M. Ahmad, G. S. Choi, and B. W. On, "Heartbeat Sound Signal Classification Using Deep Learning," Sensors, vol. 19, no. 21, 4819, January 2019.

[14] R. Spencer, F. Thabtah, N. Abdelhamid, and M. Thompson, "Exploring Feature Selection and Classification Methods for Predicting Heart Disease," Digital Health, vol. 6, pp. 1-10, March 2020.

[15] K. Shameer, K. W. Johnson, B. S. Glicksberg, J. T. Dudley, and P. P. Sengupta, "Machine Learning in Cardiovascular Medicine: Are We There Yet?" Heart, vol. 104, no. 14, pp. 1156-1164, July 2018.

[16] A. M. Alaa, T. Bolton, E. Di Angelantonio, J. H. Rudd, and M. van der Schaar, "Cardiovascular Disease Risk Prediction Using Automated Machine Learning: A Prospective Study of 423,604 UK Biobank Participants," Public Library of Science One, vol. 14, no. 5, e0213653, May 2019.

[17] V. D. Soni, "Detection of Heart Disease Using Machine Learning Techniques," International Journal of Scientific and Technology Research, vol. 9, no. 8, pp. 285-288, August 2020.

[18] N. L. Fitriyani, M. Syafrudin, G. Alfian, and J. Rhee, "HDPM: An Effective Heart Disease Prediction Model for a Clinical Decision Support System," Institute of Electrical and Electronics Engineers Access, vol. 8, pp. 133034-133050, July 2020.

[19] H. Malik, M. S. Farooq, A. Khelifi, A. Abid, J. N. Qureshi, and M. Hussain, "A Comparison of Transfer Learning Performance Versus Health Experts in Disease Diagnosis from Medical Imaging," Institute of Electrical and Electronics Engineers Access, vol. 8, pp. 139367-139386, June 2020.

[20] B. Alić, L. Gurbeta, and A. Badnjević, "Machine Learning Techniques for Classification of Diabetes and Cardiovascular Diseases," 6th Mediterranean Conference on Embedded Computing, June 2017, pp. 1-4.

[21] X. Zhai and C. Tin, "Automated ECG Classification Using Dual Heartbeat Coupling Based on Convolutional Neural Network," Institute of Electrical and Electronics Engineers Access, vol. 6, pp. 27465-27472, May 2018.

[22] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Challenges and Opportunities of Deep Learning Models for Machinery Fault Detection and Diagnosis: A Review," Institute of Electrical and Electronics Engineers Access, vol. 7, pp. 122644-122662, August 2019.

[23] M. Mohammed, M. B. Khan, and E. B. M. Bashier, Machine Learning: Algorithms and Applications, Boca Raton: Chemical Rubber Company Press, 2017.

[24] R. Y. Goh and L. S. Lee, "Credit Scoring: A Review on Support Vector Machines and Metaheuristic Approaches," Advances in Operations Research, vol. 2019, 1974794, March 2019.

[25] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian Network Classifiers," Machine Learning, vol. 29, pp. 131-163, November 1997.

[26] M. Mohri, A. Rostamizadeh, and A. Talwalkar, Foundations of Machine Learning, Cambridge: Massachusetts Institute of Technology Press, 2018.

[27] R. R. Isnanto, A. F. Rochim, D. Eridani, and G. D. Cahyono, "Multi-Object Face Recognition Using Local Binary Pattern Histogram and Haar Cascade Classifier on Low-Resolution Images," International Journal of Engineering and Technology Innovation, vol. 11, no. 1, pp. 45-58, January 2021.