Machine Learning Architecture for Heart Disease Detection: A Case Study in Iraq

Basma Jumaa Saleh(✉), Rana Riad K. Al_Taie, Abbas Atwan Mhawes
Department of Computer Engineering, Mustansiriyah University, Baghdad, Iraq
eng.basmaj@uomustansiriyah.edu.iq

Abstract—In recent years, the amount of data has been increased dramatically, driven by many real-world fields such as marketing, learning, social media, multimedia, medicine…etc. Because of that, data mining algorithms have extensively used on these data to serve as one of the newest data modeling and analytical tools, by which, a knowledge-rich environment can be generated and decision-making can be improved. Data mining tools can be employed for reducing these tests and predicting future trends by valuable information-driven decisions. There are two categories of data mining algorithms: descriptive and predictive. The rules of clustering, association, summarization, and sequence discovery will be associated with descriptive type. On the other hand, predictive type will compromise classification, regression and time series analysis rules. In this paper, a study have been presented for helping specialists and physicians in Iraq to investigate heart problems via (Weka 3.8.3) software focusing on four data mining classification techniques (1BK, J48, Naïve Bayes and REPTREE). The predictive precision tests, the ROC curve, and the AUC value are calculated based on a compiled dataset that have been received from the hospital of Baghdad medical city and the hospital of Ibn al-Bitar. The performance of the J48 technique (94.5%) indicates optimum performance based on SMO no performance factor of Baghdad medical city.

Keywords—machine learning, data collection, heart problems, predicting, classifying

1   Introduction

Heart problems diseases remain the principal cause of death worldwide, but earlier detection will help in preventing heart attack and stroke from occurring. Ideally, some of symptoms have not been taken into account, which later cause people to lose their lives. Hence, physicians need to estimate the timing of the onset of the diseases before they occur in their patients. The healthcare industry produce voluminous of clinical data, unfortunately, are not well exploited to extract hidden valuable knowledge - making decision effectively [1]. Many of factors may become the reason of heart diseases for examples: high blood pressure, bad cholesterol level in the blood, unhealthy foods, lack of physical activities, smoking, noxious drug …etc, these causes

as risk factors for predicting heart disease [2]. In fact, there are two main elements for controlling heart disease: a healthy life style and timely diagnosis. Regular check-ups (such as chest X-rays, angiography, echocardiography...) plays a very important role in the diagnosis and early prevention of heart diseases. However, such tests will consume both time and money. Adopting data mining techniques will help in reducing tests and diagnosing heart diseases early and timely. Different data mining techniques can be applied for discovering relationships and hidden trends in the data. Furthermore, these data is predominantly will be used for decision support, forecasting and assessment.

The data miming is accurately systems started by extracting data, then transforming and loading it onto a data warehouse system. Afterword, the data will be stored and managed in a multidimensional database system, then, information technology professionals and business analysts for analyzing by application software will access the stored data. Finally, the result will be presented in a specific format like a graph or table. The efforts of data mining can be classified to either descriptive or predictive. With Descriptive model the general properties of the data in database will be characterized (the data will speak about itself). While with the predictive model, the stored data will be inferenced for making future predictions [3]. There are many of data mining techniques in predictive and descriptive models such as classification, association rules, clustering, regression summarizations etc. [4]. Both prediction and classification are types of data processing which can be used to generate concepts that define relevant data classes or predicting possible data patterns [3]. In this paper, we are working to identify the ideal data mining algorithms for forecasting cardiac disease that are both computationally efficient and accurate. Weka software was used to apply 1BK, J48, Nave Bayes, and REPTREE technologies. A slandered data set had been gotten from two of hospitals in Iraq named (Ibn Al-Bitar) and (Baghdad Medical City). The rest of the paper is organized, as follows. The paper will started by addressing the literature review and related works. Afterwards, the methodology experiment's approach will be described in details. Following that, the experimental results and comparisons will be denoted. Finally, the conclusions are given in the last section.

2 Literature review and related works

Many attempts have been made by researchers to look at the use of data mining algorithms to assist healthcare providers in diagnosing cardiac disease. These techniques can be machine learning [5], Markov Chain Models, Genetic Algorithms, Reinforcement Learning, Decision Trees, and so forth are all examples of statistics.

Mai et al. [6] identified the gaps in researches between Cardiovascular diagnosis and curative. Also, they proposed a model to reduce those gaps systematically by applying single and hybrid data mining techniques to give appropriate heart disease medication, then, they applied the same techniques for diagnosing heart diseases. Generally, The proposed model aimed to answer some of questions such as: If these single and hybrid strategies assist healthcare practitioners in finding appropriate

therapy for individuals with heart disease? And, in terms of precision, which one is superior? While Vikas and Saurabh [7] worked on predicting the presence of heart disease accurately with reducing the number of attributes (originally were thirteen attributes). These attributes had been reduced to 11 attributes using different methods of classifier techniques (Naive Bayes, J48 Decision Tree and Bagging) for predicting the diagnosis of heart disease without losing efficacy and precision to. In this study, 10- fold cross validation method was used for measuring the unbiased estimate of mentioned prediction models. The results of this research show that bagging algorithm accuracy of 85.03%. However, the authors in [8] implemented hybrid system for heart disease using data from 50 individuals' risk variables. Two data approaches were used in the study: artificial neural and evolutionary computation. For initialization of neural network weights, the global optimization advantage of genetic algorithm was applied. In comparison to back-propagation algorithm, they found that the learning was more consistent, rapid, and precise. The final result indicated that the training accuracy was 96.2 percent, while the validation accuracy was 89 percent, as shown in Table IV. In [4], Vijayarani sought to determine whether of two classification algorithms (SVM) or Nave Bayes classifier) was the best based on two factors: classification accuracy and processing time performance. In comparison to the Naive Bayes classifier algorithm, the SVM classifier performed the best, according to the paper. However, the data had been classifies faster within Naïve Bayes classifier algorithm with minimum execution time. Umair et al. [9] they attempted to extract useful patterns from heart patients' information. As a basis, they used Weka machine learning software with three classification algorithm: Decision Tree, Naïve Bayes and Neural Network. 597 heart patients' data had been gotten online from the UCI repository. Furthermore, different metrics of performance were considered: Accuracy, TP rat, ROC curve value, precision, FP rate and, F-measure. Furthermore, for increasing accuracy and decreasing the time execution and complexity, two different scenarios were presented, 14 attributes had been satisfied with the first scenario, while 8 selected attributes were used in the second scenario, the data set was in ARFF format supported by Weka. They concluded that Naïve Bayes classification algorithms had the highest accuracy which approximately reached to 82.914%. For more than 1000 patients' records, [10] proposed a strategy for making diagnostic choices and determining the associated risks of clinical populations at a premature time. For identifying the risk level of patients, frequent item sets had been extracted depending on the symptoms indicated and the minimal support rating. The prediction results indicated that generation method of frequent item sets is better than existing methods. In addition, for diagnosing Chronic Obstructive Pulmonary Disease (COPD) in India, Shaila and Anupamma in [11] adopting DecisionTree technique for that purpose. The proposed model had been presented for Indian e-health system which contains patient's details. The "aadhaar" number had been used as a reference to unique "issued by the government of India as an upcoming identification proof which used to follow the treatments given to each patient in different hospitals".

The authors in [12], paper aimed to evaluate the accuracy of several data analysis categorisation algorithms. Cleveland data collection for cardiac disorders, which had

303 occurrences, was used as the core database for testing and training the proposed system. ten - fold cross was used with the following classifiers to enhance the amount of data: Na¨ıve Bayes (NB), Decision Tree (DT), K-Nearest Neighbor (K-NN), Multilayer Perceptron (MLP), Radial Basis Function (RBF), Support Vector Machine (SVM) and Single Conjunctive Rule Learner (SCRL). Moreover, bagging, boosting and stacking used as ensemble prediction of classifiers. The results concluded that the SVM had the highest accuracy with 84.15%, while the SCRL had the lowest accuracy with 69.96%. Furthermore, after bagging technique was applied, the accuracy percentage stilled the best, while DT is the worst with 78.54%. However, with boosting technique, SVM won the highest percentage of accuracy 84.81%. Finally, stacking technique confirmed MLP and SVM were the most accurate, with an accuracy of 84.15 percent. In [13] application software had been produced for medical experts to predict the occurrence or recurrence of non-communicable diseases (NCDs) based on predictive data-mining model. For examining the proposed software application, patient's records obtained from a hospital named (Bahrain Defense Force). The proposed application was executed and tested in the mentioned hospital by the actual physicians. It showed that the prediction system can be adopted for predicting NCDs' diseases instantly and effectively. Furthermore, Sayali and Rashmi used both the Nave Bayes and the KNN algorithms to predict cardiac illnesses based on the datasets in [14]. The results explained that the NB accuracy was 82% is more than KNN. In addition, the authors proposed a method for illness risk prediction using a deep convolution network-based uni-model diabetes risk predictive model as an extension of their work (CNN-UDRP). With the help of structured, The CNN-UDRP algorithm has a predictive performance of above 65 percent. However, there are many of objectives addressed by the authors in [15], they aimed to analyze different data mining techniques that used for classification and different classification methods for imbalanced data. Furthermore, for prediction of heart diseases, a new classification method had been designed. In adiition, the proposed classification method had been analyzed with different data sets. Finally, the performance of the existing classification techniques had been compared with the proposed prediction method in terms of performance and AUC. While Pranav et al [16] used five machine learning algorithms to predict the possibility of having heart disease (Random Forest, , Support Vector Machine, Naïve Bayes, Logistic Model Tree (LMT), and Hoeffding Decision Tree). For training and testing the model, Cleveland dataset had been applied. Each algorithm had been analyzed and the result compared by means of accuracy. The authors fund that the random forest is the best among the other algorithm because it gave the maximum accuracy.

3 Methodology

The main goal of this research is to arrange a dataset for thepatient to effectively predict possible heart attacks. To identify the relation of cardiovascular disease in respect of specific attributes, a model was constructed using prediction techniques. Data mining is being used in the study to create models based on the selected class

prediction attributes. Because of its knowledge in pattern identification, analysis, and prediction [17], WEKA (Waikato Environment for Knowledge Analysis) was utilized to make predictions.The Waikato Environment for Knowledge Analysis (WEKA) is an open source software and machine learning toolkit developed by the University of Waikato , New Zealand, for knowledge research. The components of this program are data pre-processing methods, classification / regression algorithms, clustering algorithms, relationship rules-finding algorithms, and 15 attribute / subset evaluators + 10 functions selection search algorithms. While the key GUIs are "Explorer" (analysis of exploratory data), "Experimenter" (experimental environment) and "Information Flow" (interface influenced by the modern process model).

The standard dataset compiled in this analysis includes 200 cases, which are obtained under the supervision of the Ministry of National Health from Iraq hospitals. A significant number of different observations must be examined during the diagnosis to identify the heart attack with high precision, typically the doctors will rely on all the recorded symptoms, the answers that patient gives to questions, as well as the physical evaluation and laboratory testing. UCI statistics provide ample medical conditions to identify heart failure, thus data obtained from Baghdad hospitals based on these elements Medical aspects such as (endurance training, ST stress produced by training compared to other parts, maximum exercise gradient ST section, and major fluoroscopic-colored vessels number) will be taken into account. As a result, cardiologists use these parameters instead of medical factors (Heart rate, family history of health, smoking, echo perkiness detection, and prior angina attack). In addition to the recorded ultrasound, modified medical conditions will be called risk factors, family history, likelihood of previous angina will be applied to achieve sufficient medical factors; these results included four forms of angina, in addition to a standard degree. Table 1 demonstrates the distribution of the 5 types of heart disease; it consists of 13 scientific conditions essential to diagnose heart disease. These variables will be converted into numerical representation to construct a diagnostic system [18].

Table 1. Dataset that was collected

Age	(0-76) is Real			
Sex	"0" for Male		"1" for Female	
CP	"1" for typ_angina	"2" for Asympt	"3" for non_anginal	"4" for atyp_angina
BF	Real is blood flow			
Chol	"0" for Normal		"1" for Abnormal	
Fbs	"1" for >120 mg/del		"0" for <120 mg/del	
Rest ECG	Normal "0"	Abnormality "1"	left_vent_hyper "2"	
Th	Thalach Value is Real			
Exang	"0" for No		"1" for Yes	
FH	"0" for No		"1" for Yes	
SM	"0" for No		"1" for Yes	
HYP	"0" for No		"1" for Yes	
Pera.	"0" for No		"1" for Yes	

4   Key performance indicators

The Indicators has been utilized for this study are described as follows:

1. Precision

This is the proportion of relevant instances among the retrieved examples. The precision equation is as follows [19]:

$$Precision = \text{TP1}/(\text{TP1} + \text{FP1}) \tag{1}$$

2. Recall

Represent the minor proportion of the correct cases. The recall formula is as follows:

$$Recall = \text{TP1}/(\text{TP1} + \text{FN1}) \tag{2}$$

3. F-Measure

F-Measure equation is given in (3) from [20].

$$\mathcal{MCC} = \frac{\text{TP1} * \text{TN1} - \text{FP1} * \text{FN1}}{\sqrt{(\text{TP1}+\text{FP1})(\text{TP1}+\text{FN1})(\text{TN1}+\text{FP1})(\text{TN1}+\text{FN1})}} \tag{3}$$

4. ROC Área

For each feasible cut off for a study or study design, Roc Curves are extensively used for graphical display of the connection / trade off, including medical sensitivity and accuracy.

5. PRC Área

The list of patients who are not unwell but have bad test results will have no effect on precision-recall slopes When assessing and comparing to get the comprehensive picture, precision-recall graphs should be utilized as an addition to the frequently used ROC curves. The outcome classifier results [21], as indicated in Table 2.

Table 2. A two different concept has a variety of impacts

Specific Class	Class Predicted	
	ÝES	*No*
ÝES	Positive True (TP1)	Negative False (FN1)
No	Positive False (FP1)	Negative True (TN1)

- True positive (TP1): The patients were accurately identified as being positive
- True negative (TN1): The patients were accurately identified as being negative. The system becomes a model example when P and N are near to 100 percent.
- TN is True Negative: People who are healthy are accurately identified as such.

- FN is False Negative: Patients with cardiovascular disease were mistakenly classified as healthy[22].
- ℭℭ I: It denotes the proportion of patients who have been appropriately diagnosed both those who require and those who do not require medical tests. Accuracy is another name for precision [23]:

$$\acute{A}ccuracy = \frac{\breve{T}P1 + \breve{T}N1}{\breve{T}P1 + \breve{T}N1 + \breve{F}P1 + \breve{F}N1} \tag{4}$$

- MAE: A contrast of what was predicted and what happened in the end. 1-ACC can be used to calculate it. The mean absolute error of a successful basis is relatively modest [24].
- Kappa: The predictability consistency with the genuine class is measured by Kappa. It calculates the difference in forecasts based on the alignment detected vs the one expected by chance. The mathematical value of kappa is a value between one and ten (0-1). A value greater than 0 indicates that the classifier is better than chance [25].RMSE: The Root Mean Squared Error [26] is the difference between the estimated and actual value.

$$\c{R}MSE = \sqrt{\frac{1}{n} \sum_{j=1}^{n} \left(\frac{\rho_{i,j} - \tau_j}{\tau_j} \right)^2} \tag{5}$$

$\rho_{i,j}$ = predicted value.
i= fitness.
τ_j= fitness applicability target value for j.

5 Experimental results and performances comparisons

To prediction a heart disease, part of the dataset is used for training and remaining part is used for testing. The results of classifiers are listed in Table 3.

Figure 1 shows the overall of the classifier visualization. Using a variety of machine learning methods, assess the percentage split results. The model was developed with the use of 200 instances. Various assessment criteria such as classification accuracy, Recall, Precision, Kappa, MAE, RMSE, RAE, RRSE, F- Measure, PRC, and ROC results were measured and shown in Figures. Depending on the precision of the classification, Figure (a), J48 gives the best accuracy of (94.50 percent) and 1BK does the worst (85 percent) in predicting individuals with heart attacks. From Figure (b), REPTREE has the order to evaluate the accuracy measure (0.887), and recall is the greatest (0.885). In addition, it can be illustrated based on Figure (b), that J48 has the high accuracy (0.887), and recall is the maximum (0.885). Figure (c) shows J48 performs best, offering F- measure values (0.87). Figure (d) shows that although J48 MAE (0.748) is slightly larger than Naïve Bayes (0.0743) and 1BK with the highest value (0.0971) Naïve Bayes RMSE (0.219) is the lowest and J48 Kappa (0.82075) is the lowest. Figure (e) shows the least amount of J48 RRSE (27.48 percent) which leads to better predictive outcomes.

Table 3. Results of the four classifiers algorithm

Algorithm Parameter	J48	Naïve Bayes	REPTREE	1BK
Correctly Classified Instances	87%	85.5%	88.5%	78.5%
Incorrectly Classified Instances	13%	14.5%	11.5%	21.5%
Statistic of Kappa	0.8075	0.7829	0.8276	0.6778
Absolute Mean Error	0.0748	0.0743	0.0775	0.0971
ṚMSE	0.2153	0.219	0.2044	0.2572
ṚAE	27.4759%	27.2793%	28.4464%	35.6616%
Ṛoot relative squared error	58.4666%	59.4838%	55.4971%	69.8451%
Áccuracy	94.5%	93%	93%	85%
Time of Construction	0.08 Sec	0.01 Sec	0.01 Sec	0 Sec

(a) Accuracy

(b) Precision Recall

(c) ROC PRC F-Measure

(d) RMSE MAE Kappa
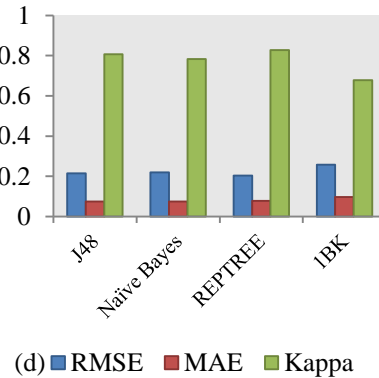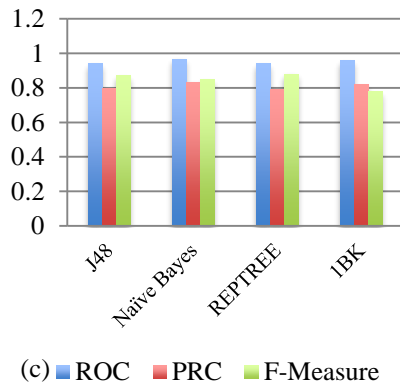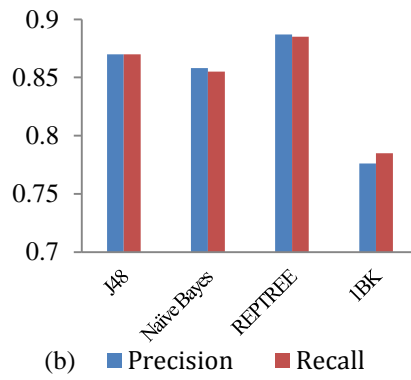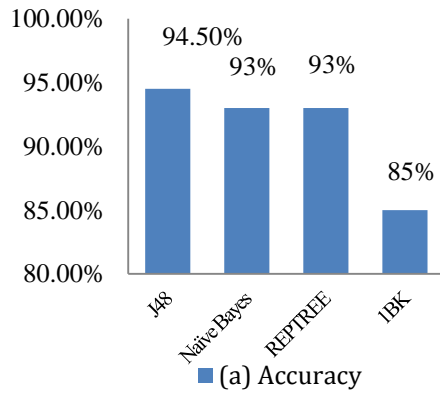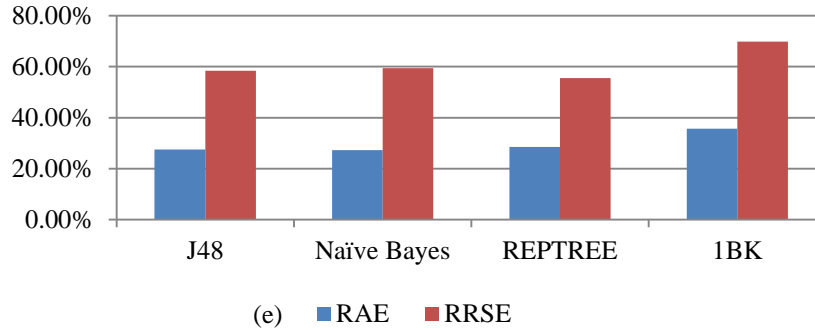
Fig. 1. (a) Comparison of accuracy for each classifier, (b) Ṗrecision and Ṛecall metrics, (c) ṚOC, ṖRC, and Ḟ-Ṃeasure metrics, (d) Ḳappa, ṂAE and ṚMSE metrics, (e) ṚAE and ṚRSE metrics

The algorithms are applied to the data set using stratified 10-fold validation to test the effectiveness of classifying strategies for classifier and calculation precision. The accuracy, sensitivity, and selectivity measures are calculated using the resultant uncertain model. The matrix shows which samples are designated as true and which are marked as false. An evaluation of the uncertainty equation reveals that J48, REPTREE, Bayes Net, and Random Forest provide a predictive model of 200 cases where cardiovascular disease risk factor is favorable. The techniques clearly suggest different data mining techniques, which can determine a type of diagnosis.

We made a contrast between our findings and those who served in the same area, which means applying same methods of WEKA on the heart disease but with different parameters, to know the effectiveness of our study. We discovered that our suggested model is superior than another models alone in terms of predicting heart disease and detection, as well as classification accuracy. Table 4 and Figure 2 shows our suggested framework for classifying WEKA with other observations from the study.

Table 4. Comparison between results of previous experiments and proposed results

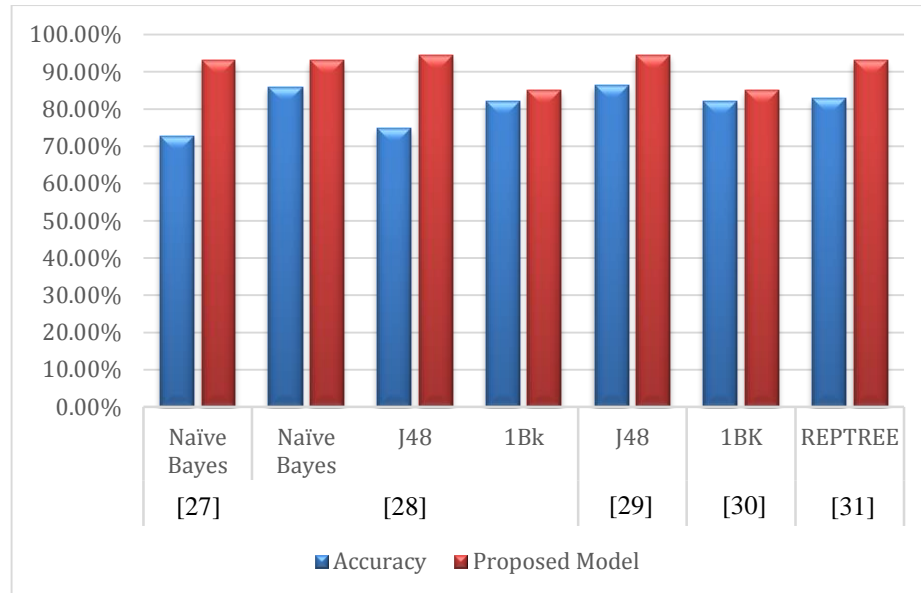| Another works' model	Methodologies Applied	Áccuracy	Suggested Model
[27] (2013)	Naïve Bayes	72.83%	93%
[28] (2020)	Naïve Bayes	85.93%	93%
[28] (2020)	J48	74.75%	94.5%
[28] (2020)	1Bk	82.02%	85%
[29] (2015)	J48	86.30%	94.5%
[30] (2016)	1BK	82.02%	85%
[31] (2017)	REPTREE	82.77%	93%

Fig. 2. Bar chart of comparison between previous experiments results and suggested results

6 Conclusion

In this research article, we have introduced anEfficient Heart Disease Prediction Program using data mining and we use a classifier ensemble to analyze the accuracy of heart disease prediction. The heart data set obtained from Ibn al-Bitar hospital and the machine-learning repository in Baghdad medical city was used for training and testing purposes. This device will help medical practitioners make effective decisions based on a given parameter. The experiment was successfully conducted with several data mining classification techniques (1BK, J48, Naïve Bayes and REPTREE) with 10 fold and it is found that the J48 algorithm gives better output with the accuracy of (94.5 per cent) over the supplied data set. It is believed that data mining will contribute significantly to cardiac disease research and ultimately improve quality. It may also be applied using multiple classification methods.

7 Acknowledgment

8   References

[1] A.K. Sen, S.B. Patel, D.P. Shukla. "A data mining technique for prediction of coronary heart disease using neuro-fuzzy integrated approach two level." *International Journal of Engineering and Computer Science* 2.09 (2013).

[2] M. Sultana, A. Haider, M.S. Uddin . "Analysis of data mining techniques for heart disease prediction." 2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT). IEEE, 2016. https://doi.org/10.1109/CEEICT.2016.7873142

[3] N. Jain, V. Srivastava. "Data mining techniques: a survey paper." IJRET: International Journal of Research in Engineering and Technology 2.11 (2013): 2319-1163. http://dx.doi.org/10.15623/ijret.2013.0211019

[4] S. Vijayarani, S. Dhayanand. "Data mining classification algorithms for kidney disease prediction." International Journal on Cybernetics & Informatics (IJCI) Vol.4, No.4, 2015. https://doi.org/10.5121/ijci.2015.4402

[5] Qinghong Yang, Pengfei Feng, Zhichao Cheng, "Clothing Product Reviews Mining based on Machine Learning ", International Journal of online and Biomedical Engineering, iJOE-Volume 11, Issue 9, 2015. https://doi.org/10.3991/ijoe.v11i9.5069

[6] M. Shouman, T. Turner, R. Stocker. "Using data mining techniques in heart disease diagnosis and treatment." 2012 Japan-Egypt Conference on Electronics, Communications and Computers. IEEE, 2012. https://doi.org/10.1109/JEC-ECC.2012.6186978

[7] V. Chaurasia, S. Pal. "Data mining approach to detect heart diseases." International Journal of Advanced Computer Science and Information Technology (IJACSIT) Vol 2 (2014): 56-66.

[8] S.U. Amin, K. Agarwal, R. Beg. "Genetic neural network based data mining in prediction of heart disease using risk factors." 2013 IEEE Conference on Information & Communication Technologies. IEEE, 2013. https://doi.org/10.1109/CICT.2013.6558288

[9] U. Shafique, F. Majeed, H. Qaiser. "Data mining in healthcare for heart diseases." International Journal of Innovation and Applied Studies 10.4 (2015): 1312.

[10] M. Ilayaraja, T, Meyyappan. "Efficient data mining method to predict the risk of heart diseases through frequent itemsets." Procedia Computer Science 70 (2015): 586-592. https://doi.org/10.1016/j.procs.2015.10.040

[11] M.H. Tekieh, B. Raahemi. "Importance of data mining in healthcare: a survey." Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social. "Importance of data mining in healthcare: a survey." Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015. 2015. https://doi.org/10.1145/2808797.2809367

[12] S. Pouriyeh, S. Vahid, G. Sannino "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease." 2017 IEEE Symposium on Computers and Communications (ISCC). IEEE, 2017. https://doi.org/10.1109/ISCC.2017.8024530

[13] A. Aldallal, A.A.A. Al-Moosa. "Using Data Mining Techniques to Predict Diabetes and Heart Diseases." 2018 4th International Conference on Frontiers of Signal Processing (ICFSP). IEEE, 2018. https://doi.org/10.1109/ICFSP.2018.8552051

[14] S. Ambekar, R. Phalnikar. "Disease Risk Prediction by Using Convolutional Neural Network." 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE, 2018. https://doi.org/10.1109/ICCUBEA.2018.8697423

[15] K.V. Nagendra, M. Ussenaiah. "Design and Development of EGB Classification Model for predicting Heart Diseases." 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA). IEEE, 2020. https://doi.org/10.1109/ICIMIA48430.2020.9074880

[16] M. Gayathri, D.N. UmaMaheswari. "Cognitive Approach for Heart Disease Prediction using Machine Learning." 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE). IEEE, 2020. https://doi.org/10.1109/ic-ETITE47903.2020.242

[17] I.H. Witten, E. Frank. "Data mining: practical machine learning tools and techniques with Java implementations." Acm Sigmod Record 31.1 (2002): 76-77. https://doi.org/10.1145/507338.507355

[18] T. Tabreer . "Design and Implementation of Intelligent Algorithm for the Diagnosis of Heart Disease Using FPGA," MSc. Thesis, University of Technology, Iraq (2017).

[19] B. Jumaa Saleh, A. Yousif Falih Saedi, A. Talib Qasim al-Aqbi, L. Abdalhasan Salman "A REVIEW PAPER: ANALYSIS OF WEKA DATA MINING TECHNIQUES FOR HEART DISEASE PREDICTION SYSTEM", Library Philosophy and Practice, (2020) 4032. https://digitalcommons.unl.edu/libphilprac/4032

[20] R. Riad K. AL-Taie, B. Jumaa Saleh, A. Yousif Falih Saedi, L. Abdalhasan Salman " Analysis of WEKA data mining algorithms Bayes net, random forest, MLP and SMO for heart disease prediction system: A case study in Iraq", International Journal of Electrical and Computer Engineering (IJECE), Vol. 11, No. 6, (2021): 5229~5239. http://doi.org/10.11591/ijece.v11i6.pp5229-5239

[21] S. Sakr, R. Elshawi, A. Ahmed, W.T. Qureshi, C. Brawner "Using machine learning on cardiorespiratory fitness data for predicting hypertension: The Henry Ford ExercIse Testing (FIT) Project." PLoS One 13.4 (2018). https://doi.org/10.1371/journal.pone.0195344

[22] S. Hari Ganesh, M. Gajenthiran "Comparative study of data mining approaches for prediction heart diseases." IOSR Journal of Engineering 4 (2014): 36-39. https://doi.org/10.9790/3021-04733639

[23] Halima El Hamdaoui, Saïd Boujraf, Nour El Houda Chaoui, Badreddine Alami, Mustapha Maarouf, " Improving Heart Disease Prediction Using Random Forest and AdaBoost Algorithms ", International Journal of online and Biomedical Engineering, iJOE- Volume 17, Issue 11, 2021, https://doi.org/10.3991/ijoe.v17i11.24781

[24] N. Gupta, N. Ahuja, S. Malhotra, A. Bala, G. Kaur. "Intelligent heart disease prediction in cloud environment through ensembling." Expert Systems 34.3 (2017): e12207. https://doi.org/10.1111/exsy.12207

[25] H. Kutrani, S. Eltalhi "Cardiac Catheterization Procedure Prediction Using Machine Learning and Data Mining Techniques". IOSR Journal of Computer Engineering (IOSR-JCE)1, Volume 21, Issue 1, Ver. I (Jan - Feb 2019).

[26] Vinutha M.R., Chandrika J., " Prediction of Liver Disease using Regression Tree ", International Journal of online and Biomedical Engineering, iJOE- Volume 17, Issue 2, 2021. https://doi.org/10.3991/ijoe.v17i02.19287

[27] R. Alizadehsani, J. Habibi, M.J. Hosseini"A data mining approach for diagnosis of coronary artery disease." Computer methods and programs in biomedicine 111.1 (2013): 52-61. https://doi.org/10.1016/j.cmpb.2013.03.004

[28] A.K. Ramotra, A. Mahajan, R. Kumar. "Comparative Analysis of Data Mining Classification Techniques for Prediction of Heart Disease Using the Weka and SPSS Modeler Tools." Smart Trends in Computing and Communications. Springer, Singapore, 2020. 89-97. https://doi.org/10.1007/978-981-15-0077-0_10

[29] A.J. Aljaaf, D. Al-Jumeily, A.J. Hussain. "Predicting the likelihood of heart failure with a multi level risk assessment using decision tree." 2015 Third International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE). IEEE, 2015. https://doi.org/10.1109/TAEECE.2015.7113608

[30] Z. Masetic, A. Subasi -. "Congestive heart failure detection using random forest classifier." Computer methods and programs in biomedicine 130 (2016): 54-64. https://doi.org/10.1016/j.cmpb.2016.03.020

[31] M. Mirmozaffari, A Alinezhad. "Data Mining Apriori Algorithm for Heart Disease Prediction." Int'l Journal of Computing, Communications & Instrumentation Engg 4.1 (2017): 20-23. https://doi.org/10.15242/ijccie.dir1116010

9 Authors

Basma Jumaa Saleh is with Department of Computer Engineering, Mustansiriyah University, Baghdad, Iraq.

Rana Riad K. Al_Taie is with Department of Computer Engineering, Mustansiriyah University, Baghdad, Iraq.

Abbas Atwan Mhawes is with Ministry of Education, Baghdad, Iraq.