VOLUME - 11, ISSUE - 05, MAY - 2022 • PRINT ISSN No. 2277 - 8160 • DOI : 10.36106/gjra

Original Research Paper

Computer Science

PREDICTIVE THE HEART DISEASE USING THE WEIGHTED GAIN DECISION TREE ALGORTHIM

Deepak Saraswat*

Ph.D. full time Research Scholar, Department of Physics and Computer Science, Dayalbagh Educational Institute (Deemed University), Agra-282005, U.P. *Corresponding Author

Prof. Preetvanti Singh

Professor, Department of Physics and Computer Science, Dayalbagh Educational Institute (Deemed University), Agra-282005, U.P.

ABSTRACT Data mining techniques have been mostly used in medical area for prediction and diagnosis of various diseases. These techniques discover the hidden pattern and relationship in medical data and therefore have been very important in designing clinical support. Now a day's data mining techniques are widely used in diagnosis of heart disease because of increasing death rate worldwide. The reason of this may be the complex and expensive tests conducted in labs to predict the heart disease. Systems based on these risk factors not only benefit healthcare professionals, but warn them of the potential presence of heart disease even before a patient is admitted to the hospital or undergoes an expensive medical examination. This in order to reduce the risk of this disease a better approach would to identify risk factor the result in heart disease. This study is an effort in this direction. This approach to predict the heart disease in early stage is developed in present study by analyzing risk factors. This technique developed weighted gain decision tree predicts the risk of heart disease with an accuracy of 90%.

KEYWORDS : Data mining, heart disease risk factors, classification, decision tree, prediction.

INTRODUCTION

Knowledge discovery now-a-days plays a vital role in decision making. Data mining is the process of discovering actionable information from large set of data, and can be used to extract patterns that can be converted to useful knowledge. Data mining includes various algorithms such as classification, association rule mining, and clustering for analyzing the data. Classification, an important concept of data mining, can be an effective solution to the knowledge acquisition or knowledge extraction problem. The objective of classification is to construct a model of the class label that can be used to classify new data that has unknown class labels based on training data. The training set of records is used to construct the classifier. Each record has a class label assigned to it. A setoff attribute values defines each record. The goal is to induce description for each class in terms of the attributes. This is then used by the classifier to classify future records whose classes are unknown.

There are many types of models built for classification, such as neural network, statistical models, genetic models, and decision tree models. In the domain of data mining, decision tree models are found to be the most useful as they can be computed relatively inexpensively and they have reasonable degree of accuracy. Decision tree analysis is frequently used for medical data mining.

A plethora of data about patient is constantly being added to the large number of medical records in the healthcare industry. It can be said that this industry is information rich, yet knowledge poor. However, data mining with its various analytical tools and techniques plays a major role in exploring the hidden patterns in the data sets of medical domains which can be utilized for clinical diagnosis or prevention.

LITERATURE REVIEW

Many diseases have become deadly today, including heart disease, breast cancer, and diabetes. Heart disease (HD) is considered one of the most complex and life-threatening human diseases in the world. The researchers have made efforts for early diagnosis or prevention of heart disease. Mozaffarian et al. (2008) identified and discussed the life style risk factors for CVD and suggested some preventive measures. Palaniappan & Awang (2008) developed a prototype Intelligent Heart Disease Prediction System using Decision Trees, Naïve Bayes and Neural Network and concluded that each technique has its unique strength in achieving the objectives of defined mining goals. Fan et al. (2011) developed a hybrid model using by integrating a case-based clustering method and fuzzy decision tree for medical data classification. Two datasets from UCI Machine Learning Repository, i.e., liver disorders dataset and Breast Cancer Wisconsin (Diagnosis), were employed for benchmark test. Soni et al. (2011) provided a survey of current techniques of knowledge discovery in databases using data mining techniques that are in use in today's medical research particularly in Heart Disease Prediction. Many experiments have been conducted to compare the performance of predictive data mining techniques on the same dataset, with the outcome showing Decision tree outperform Bayesian classification with similar accuracy.

Shouman et al. (2012) identified gaps in the research on heart disease diagnosis and treatment and proposed a model using data mining techniques for heart disease treatment. Nahar et al. (2013) investigated the sick and healthy factors which contribute to heart disease for males and females. Association rule mining was used to identify these factors. Joshi & Nair (2015) focused on the prediction of heart disease using three classification techniques namely Decision Trees, Naïve Bayes and K Nearest Neighbor. A fuzzy logic and decision tree (classification and regression tree [CART])-driven coronary heart disease prediction model was developed by (Kim et al. (2015)) for Koreans. Datasets derived from the Korean National Health and Nutrition Examination Survey VI was utilized to generate the proposed model.

Chadha & Mayank (2016) employed and analyzed different data mining techniques for the prediction of heart disease in a patient through extraction of interesting patterns from the dataset using vital parameters. Devi et al. (2016) developed a heart disease prediction model which implements data mining technique to help the medical practitioners in detecting the heart disease status based on the patient's clinical data. Gomathi & Shanmugapriyaa (2016) presented an analysis of the Heart disease for male patients using data mining techniques. 210 records from the preprocessed data set have all 8 available fields from the database. Babu et al.

(2017) extracted fourteen attributes such as age, sex, blood pressure, etc. from the medical profiles to predict the likelihood of patient getting heart disease. These attributes are fed in to K-means algorithms, MAFIA algorithm and Decision tree classification in heart disease prediction to provide reliable performance in diagnosing heart disease. Salman et al. (2017) evaluated and scored the big data of patients with chronic heart disease and of those who require urgent attention. The assessment is based on multi criteria decision making in a telemedical environment to improve the triage and prioritization processes.

Ahmad et al. (2019) identified the data mining techniques and algorithms that are commonly implemented for various disease risk prediction model as well as finding the accuracy of the algorithms. Haq et al. (2018) developed a machine-learning-based diagnosis system for heart disease prediction by using heart disease dataset. Seven machine learning algorithms, three feature selection algorithms, the cross-validation method, and seven classifiers performance evaluation metrics were used for analysis. For building a classification model for heart disease patients Kumar et al. (2018) used four different classification algorithms: NaiveBayes, MultilayerPerceptron, RandomForest and Decision Table to classify that whether a patient is tested positive or negative for heart diseases, based on some diagnostic measurements integrated into the dataset. Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System (MKL with ANFIS) based deep learning method was proposed by (Manogaran et al. (2018)) for heart disease diagnosis. Ghorbani & Ghousi (2019) reviewed 168 articles associated with the implementation of data mining for diagnosing heart diseases, and determined the most efficient data mining methods used for medical diagnosing purposes. Authors also identified research gaps in the application of data mining in health care. Large set of medical instances were taken as input by (Nagamani et al. (2019)) to extract the required information from the record of heart patients using Mapreduce technique. The functionality of the proposed Mapreduce Algorithm across similar and distributed systems was analyzed using the Cleveland data set. Kumar et al. (2020) used the data mining techniques such as Naïve Bayes, Support Vector Machine and Decision Tree to predicted heart problem for patients of diabetes dependent. Preethi & Selvakumar (2020) explained big data and machine learning techniques for predicting the heart problem. The estimate of heart problem is a very difficult task in medical field. Sowmiya & Sumitra (2021) provided a latest method with novel feature selection and classification technique to estimate the heart disease. The death rate during heart disease will be decrease when use this approach. To select the best features for hybrid K-nearest neighbor classifier are used ant colony optimization algorithm. When Compared this developed algorithm to predefined algorithms such as SVM, Naïve Bayes, KNN and C4.5 than developed algorithm give best Accuracy. Sahoo, Das, Mishra & Suman (2021) used many predefined algorithms such as J48, Naïve Bayes, REPTREE, CART, and Bayes Net in this research. Rathi & Gupta (2021) developed mobile-based disease detection system using hybrid data mining method. In this research the voting ensemble technique is used for combining Sequential Minimal Optimization and Naïve Bayes. In this methodology user enter the symptoms and proposed model predict the diseases online.

Cardiovascular disease is one of the most serious human diseases in the world and has a very bad effect on people's lives. Accurate and timely diagnosis of heart disease is important in preventing heart failure and treatment. Heart disease prevention system can help health professionals in predicting heart disease status based on the medical data of the patient. This research presents a data mining system based on weighted decision tree for predicting the heart disease.

MATERIALS AND METHODS

Classification is the technique of data mining. Classification technique is used to classify the data on the basis of different classes. Decision trees are one of the classification methods that classify labeled training data in the form of trees or rules. After checking the accuracy, the unlabeled data is classified using the tree or rules learned in the training phase. Decision trees support both quantitative and qualitative data. Large volume of meteorological data can be analyzed in a reasonable amount of time.

The structure of a decision tree is similar to a tree with a root node, a left subtree, and a right subtree. Leaf nodes in the tree represent class labels. An arc from one node to another represents a condition for an attribute.

A tree can be built like this:

Attribute selection as root node is based on attribute splitting.
Determination of a node to be presented as an end node or a node to continue the node split.

Assign a terminal node to a class. The attribute splits depend on the impurity measures such as Information gain, gain ratio, Gini index etc.After getting the tree or rule in the training phase, the test data is taken randomly from the training data to test the accuracy of the classifier. After the tree is built, it is pruned to check for overfitting and noise. Finally, the tree is an optimized tree. The advantages of the tree approach are ease of understanding and interpretation, handling of categorical and numeric attributes, and tolerance for outliers and missing values. Decision tree classifiers are used extensively for diagnosis of diseases such as breast cancer, ovarian cancer and heart sound diagnosis and so on (Mendis et al. (2011)).

In this section, an algorithm DT-classifier is developed for decision-tree classifier that performs classification in two phases: Tree Building and Tree Pruning. In tree building, a decision tree model is built by recursively partitioning the training data set according to the local best-fit criterion until all or most records belonging to each partition have the same class label. To improve decision tree generalization, tree pruning is used to prune leaves and branches that serve to classify one or a very small number of data vectors. While building and pruning the tree, the algorithm takes into account the weights for the attributes also.

The steps of the DT-classifier algorithm are:

Step 1. Identify the attributes ($A_i, i = 1, \dots, I$) with their values (v_i^T, T = 1, ..., t) related to the healthcare problem.

Step 2. Calculate the Importance Weight (W_i) of the attributes using Eigen vector approach.

Step 3. Calculate the Weighted Gain (WG_i) of all the attributes using the equations given below

$$WG_i = W_i * G_i \qquad \text{Eq (1)}$$

Where

$$G_i = I_i^T - E_i \qquad \text{Eq (2)}$$

$$E_i = \sum_{T=1}^{t} \frac{p_i^T + n_i^T}{\sum_{c=1}^{t} p_i^c + n_i^c} \times I_i^T \qquad \text{Eq (3)}$$

And

$$I_i^T = -\frac{p_i^T}{p_i^T + n_i^T} log_2\left(\frac{p_i^T}{p_i^T + n_i^T}\right) - \frac{n_i^T}{p_i^T + n_i^T} log_2\left(\frac{n_i^T}{p_i^T + n_i^T}\right) \qquad \text{Eq (4)}$$

Here
p represents the positive outcomes
n represents the negative outcomes
G_i is the Gain
E_i is the Entropy
I_i^T is the Information Gain

Step 4. Compute the Split Attribute Value (V_i) of all the attributes.

$$V_i = -\sum_{T=1}^{t} \frac{p_i^T}{p_i^T + n_i^T} log_2\left(\frac{p_i^T}{p_i^T + n_i^T}\right) - \sum_{T=1}^{t} \frac{n_i^T}{p_i^T + n_i^T} log_2\left(\frac{n_i^T}{p_i^T + n_i^T}\right) \quad \text{Eq (5)}$$

Step 5. Calculate the Weighted Gain Ratio (R_i) of every attribute.

$$R_i = \frac{WG_i}{V_i} \quad \text{Eq (6)}$$

Step 6. Find attribute with maximum weighted gain.

Step 7. Make a root node of the attribute of the decision tree which contains maximum weighted gain ratio.

Step 8. Then take the value of this root node as sub root node.

Step 9. Repeat step 4 to step 6 for remaining attributes.

Step 10. Again, find attribute which contains the maximum weight gain for remaining attributes with respect to the entire sub root node.

Step 11. This attribute will be the sub root of the sub root node.

Step 12. Repeat the step 7 to Step 11 until all remaining attributes are covered.

Step 13. Perform pruning of the tree by calculating cost complexity as:

$$Cost\ complexity = Error\ rate\ of\ the\ sub\ tree + \alpha * (Leaf\ nodes\ in\ the\ sub\ tree) \quad \text{Eq (7)}$$

In order to compute the error rate, the following formula is used:

- $Error\ rate = 1 - Accuracy$ Eq (8)
- $= 1 - ((TP + TN)/(P + N))$ Eq (9)

Where TP, TN, P and N are the True Positive, True Negative, Positive sample and Negative sample.

Second for computing . following steps are performed:

- Compute the weight of every node of sub trees. Those sub trees are pruned
- Compute the height of the sub tree using the following equation

Height of sub tree=((n-1)/2) Eq (10)
Where n is the number of nodes in the sub tree.

- Compute as

α = Height of sub tree. Eq (11)

The sub trees that have low accuracy and low-cost complexity are pruned and replaced by leaf node

CASE STUDY

Although heart diseases have been identified as the leading cause of death in the world in the past decade, they have been introduced as the most preventable and controllable diseases (Poirier, (2008)). Complete and accurate treatment of the disease depends on the timely diagnosis of this disease. Accurate and systematic tools for identifying high-risk patients and extracting data for timely diagnosis of heart disease appear urgent.

This section demonstrates the implication of the proposed DT-classifier algorithm in the area of Heart disease. Data of patients is collected from sources like website, doctors, patients and survey done by American Heart Association. Data is also collected from the patients in the form of questionnaires. Mostly heart disease patients had many similar risk factors (Isern et al., (2010)). 58% of the data was used for system development and 42% was used each for testing.Table 1 shows the identified risk factors and their corresponding values. These risk factors and their values were used as input to the system.

TABLE-1: Risk factor and their corresponding value

S.No.	Risk factor	Value
1.	Sex	Male, female
2.	Age(year)	21-25,26-30,31-35,36-40,41-45,46-50,51-55,56-60,61-65,66-70
3.	Blood Cholesterol	High > 240 mg/dL, Low< 200 mg/dL, Normal= 200-239 mg/dL dL means milligram per deciliter
4.	Blood Pressure	High (140/90 MMhg or higher) Low (90/60 MMhg or lower) Normal (between 90/60 MMHg and 139/89 MMHg) Blood pressure measured in millimeters of mercury (MMHg) 140 MMHg (Systolic) 90 MMHg(diastolic)
5.	Heritable	Yes, no
6.	Smoking	Yes, no
7.	Drunker	Yes, no
8.	Exercise Work	High, Low, Normal
9.	Diabetes	Yes, No
10.	Meal	Good, Poor, Normal
11.	Over Weight	Yes, No
12.	Tension	Yes, No
Result	Heart Disease	Yes, no

Pairwise comparison rating matrices were developed using a 9-point scale to determine the Importance Weight for all the attributes. Next the weighted gain, weighted gain ratio for the attributes are computed using equation Eq (1) and Eq (6).Considering the maximum value at each stage, the decision tree is constructed which is pruned as described in Step 13.

RESULTSTable 2 displays the computed importance weights for all the identified attributes. Table 3 illustrates the computed weighted gain ratios. It can be seen that blood cholesterol has the highest value; thus, it is considered as the root node.

TABLE-2: Risk Factors and their Weights

Risk factor	Weights
Sex	0.009
Age	0.012
Blood Cholesterol	0.364
Blood Pressure	0.153
Heritable	0.032
Smoking	0.116
Drunker	0.034
Exercise Work	0.018
Diabetes	0.120
Meal	0.025
Over Weight	0.053
Tension	0.064

TABLE- 3: Weighted Gain Ratio

Risk factor	Weighted gain ratio
Sex	0.000275
Age	0.0011230
Blood Cholesterol	0.0367361 (Root Node)
Blood Pressure	0.0123291
Heritable	0.0011490
Smoking	0.0092300
Drunker	0.0005710
Exercise Work	0.0017140
Diabetes	0.000505
Meal	0.000386
Over Weight	0.006996
Tension	0.000128

Table 4 presents the computed Weighted Gain Ratios with respect to the attributes to get the child node of Blood

Cholesterol. Proceeding in the same manner, the decision tree is constructed as shown in Figure 1.

Table-4: Weighted Gain Ratio with respect to Blood Cholesterol

Risk factors	Weighted gain ratio (BC = High)	Weighted gain ratio (BC = Normal)	Weighted gain ratio (BC = Low)
Sex	0	0.000987	0.000319
Age	0.002985	0.002154	0.000318
Blood Pressure	0.033617 (sub-node)	0.008605	0.001376
Heritable	0.002394	0.001139	0.006400
Smoking	0.005929	0.012269 (sub-	0
Drunker	0.003314	0.001265	0.001449
Exercise	0.000415		0.000922
Diabetes	0.000652	0.009565	0.002296
Meal	0	0.000211	0
Over Weight	0.000288	0.010886	0.318867 (sub-
Tension	0.000347	0.006239	0.020062

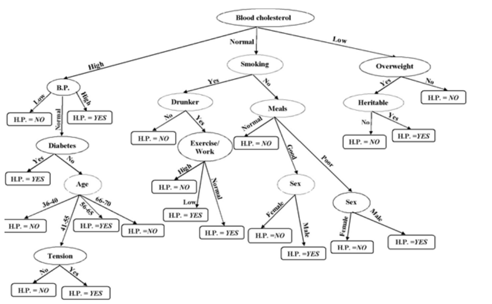
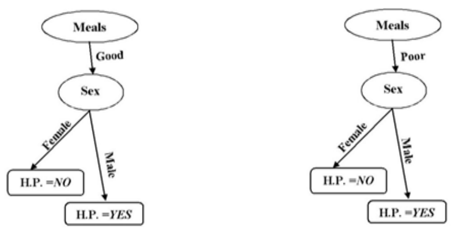


Figure 1: Decision Tree before Pruning

From the above figure, it is clear that this is not an optimized tree, thus tree pruning is applied as mentioned in Step 13.



The cost complexity is computed using Eq (7) as follows

Meal = Good
Sub-tree size = 4
Sub-tree Height = 2
Total weight = 1.131
Wt/size = 0.28275
Wt/ht ( ) = 0.5655
Accuracy=0.4285
Error rate=0.5715
Cost complexity=1.7025

Meal = Poor
Sub-tree size = 6
Sub-tree Height = 3
Total weight = 1.567
Wt/size = 0.2611
Wt/ht ( ) = 0.5223
Accuracy=0.625
Error rate=0.375
Cost complexity=1.942

In figure 2 left hand side sub tree accuracy and cost complexity are less than right hand side sub tree accuracy and complexity so left-hand side sub tree will be pruned. Thus, the decision tree after pruning is illustrated in figure 3.

True positive, True negative, False positive and False negative are summarized in the confusion matrix, which is a useful tool for analyzing how well the classifier can recognize tuples of different classes. TP and TN specify that the classifier is getting things right, while FP and FN imply the classifier is getting mislabeling.

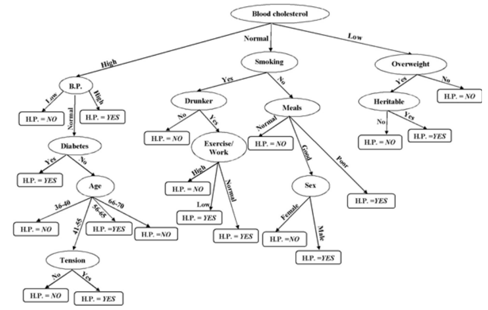


Figure 3: Decision Tree after Pruning

Table- 5: Confusion matrix, shown with totals for positive and negative tuples.

Actual Class	Predicted class			
		Yes	No	Total
	Yes	TP	FN	P
	No	FP	TN	N
	Total	P'	N'	P+N

The proposed algorithm computes the value of terms (True positive, True negative, False positive and False negative), which are summarized in the form of confusion matrix in Table 6 and Table 7.

Table 6: Confusion matrix (Training data), shown with totals for positive and negative tuples.

Actual Class	Predicted class			
		Yes	No	Total
	Yes	13	1	P
	No	0	26	N
	Total	P'	N'	P+N

Table-7: Confusion matrix (Testing data), shown with totals for positive and negative tuples.

Actual Class	Predicted class			
		Yes	No	Total
	Yes	11	3	P
	No	0	16	N
	Total	P'	N'	P+N

The tables below show the Class accuracy on the basis of training and testing data.

Table- 8: Class Accuracy (Training data).

Algorithms	Class	Training data	
		TP Rate	FP Rate
J48	Yes	1.000	0.192
	No	0.808	0.000
Decision stump	Yes	0.714	0.154
	No	0.846	0.286
LMT	Yes	0.929	0.077
	No	0.923	0.071
Hoeffding tree	Yes	0.857	0.115
	No	0.885	0.143
REPTree	Yes	0	0
	No	1	1
Id3	Yes	1.0	0.0
	No	1.0	0.0
Developed Algorithm	Yes	0.928	0.0
	No	1.0	0.037

Table- 9: Class Accuracy (Testing data).

Algorithms	Class	Testing data	
		TP Rate	FP Rate
J48	Yes	0.929	0.250
	No	0.750	0.071

Decision stump	Yes	0.929	0.188
	No	0.813	0.071
LMT	Yes	0.929	0.125
	No	0.875	0.071
Hoeffding tree	Yes	0.857	0.125
	No	0.875	0.143
REPTree	Yes	0	0
	No	1	1
Id3	Yes	0.769	0.133
	No	0.867	0.231
Developed Algorithm	Yes	0.785	0.0
	No	1.0	0.157

The developed algorithm is compared with other predefined algorithms on the basis of measures for evaluating the classifier performance. These measures (Accuracy, Error Rate, Sensitivity, F, Specificity, Precision and Recall) are computed using Weka 3.9.3.

Table- 10: Comparison with other algorithms (Training data)

	Proposed algo	J48	LMT	REPtree	ID3	Decision stump	Hoeffding tree
Accuracy	97.5	87.5	92.5	82.5	95.0	80	87.5
Error rate	2.5	12.5	7.5	17.5	5.0	20	12.5
Sensitivity	0.92	1	0.928	0.71	0.92	0.714	0.8571
F	0.95	1.57	0.88	0.734	0.92	0.714	0.8275
Specificity	1.0	0.8	0.92	0.88	0.96	0.846	0.8846
Precision	1.0	0.73	0.86	0.76	0.92	0.714	0.8
Recall	0.92	1	0.92	0.71	0.92	0.714	0.857

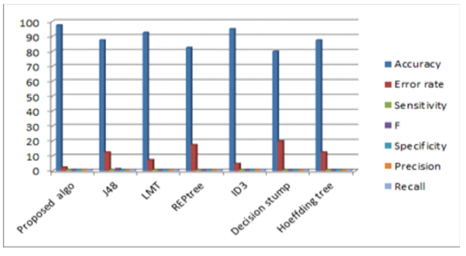


Figure 4: Graph showing measures for evaluating the classifier performance (Training data).

Table- 11: Comparison with other algorithms (Testing data)

	Proposed algo	J48	LMT	REPtree	ID3	Decision stump	Hoeffding tree
Accuracy	90.0	83.3	90.0	70.0	76.6	86.66	82.14
Error rate	10.0	16.7	10.0	30.0	23.3	13.33	17.86
Sensitivity	0.78	0.92	0.92	0.57	0.71	0.928	0.857
F	0.876	0.8323	0.88	0.6362	0.7383	0.6392	0.5999
Specificity	1.0	0.75	0.87	0.81	0.81	0.812	0.875
Precision	1.0	0.76	0.86	0.72	0.72	0.866	0.857
Recall	0.78	0.92	0.92	0.57	0.57	0.5	0.4615

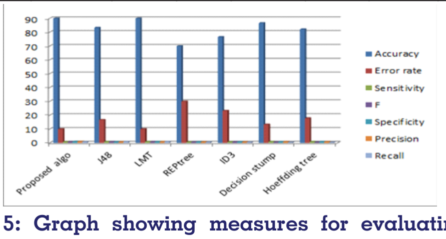


Figure 5: Graph showing measures for evaluating the classifier performance (Testing data).

On the basis of the computations it is clear that the performance of the developed algorithm is better than other predefined algorithms.

CONCLUSION

Various data mining techniques are applied in medical areas to improve the patient's health and overall quality of medical services but from patient point of view a system must be able to predict the heart disease at an early stage. In this research, a new model for generating weighted gain decision tree is developed to predict the heart disease in early stages using the risk factors. The model of weighted gain decision tree uses identified important risk factors for the prediction of heart disease and does not require costly medical tests. Data of 70 patients was collected and the results obtained showed training accuracy of 97.5% and testing accuracy of 90% as shown in Table 10 and Table 11. Such an intelligent system helps in predicting the disease using risk factors and hence saves cost and time to undergo medical tests and checkups. The system is capable of ensuring that the patient can monitor his health on his own and plan preventive measures and treatment at the early stages of the diseases. In order to further improve the working of the system hybrid data mining techniques will be used to design more accurate decision support system for diagnosis of diseases.

REFERENCES

[1] Ahmad, W. M. T. W., Ghani, N., & Drus, S. M. (2019). Handling imbalanced class problem of measles infection risk prediction model. Int. J. Eng. Adv. Technol., 9(1), 3431-3435.
[2] Babu, S., Vivek, E. M., Famina, K. P., Fida, K., Aswathi, P., Shanid, M., & Hena, M. (2017, April). Heart disease diagnosis using data mining technique. In 2017 international conference of electronics, communication and aerospace technology (ICECA) (Vol. 1, pp. 750-753). IEEE.
[3] Chadha, R., & Mayank, S. (2016). Prediction of heart disease using data mining techniques. CSI transactions on ICT, 4(2), 193-198.
[4] Devi, S. K., Krishnapriya, S., & Kalita, D. (2016). Prediction of heart disease using data mining techniques. Indian journal of Science and technology, 9(39), 1-5.
[5] Fan, C. Y., Chang, P. C., Lin, J. J., & Hsieh, J. C. (2011). A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. Applied Soft Computing, 11(1), 632-644.
[6] Ghorbani, R., & Ghousi, R. (2019). Predictive data mining approaches in medical diagnosis: A review of some diseases prediction. International Journal of Data and Network Science, 3(2), 47-70.
[7] Gomathi, K., & Shanmugapriyaa, D. (2016). Heart disease prediction using data mining classification. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 4(II), 60-63.
[8] Haq, A. U., Li, J. P., Memon, M. H., Nazir, S., & Sun, R. (2018). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Information Systems, 2018.
[9] Isern, D., Sánchez, D., & Moreno, A. (2010). Agents applied in health care: A review. International journal of medical informatics, 79(3), 145-166.
[10] Joshi, S., & Nair, M. K. (2015). Prediction of heart disease using classification based data mining techniques. In Computational Intelligence in Data Mining-Volume 2 (pp. 503-511). Springer, New Delhi.
[11] Kim, J., Lee, J., & Lee, Y. (2015). Data-mining-based coronary heart disease risk prediction model using fuzzy logic and decision tree. Healthcare informatics research, 21(3), 167-174.
[12] Kumar, A., Kumar, P., Srivastava, A., Ambeth Kumar, V. D., Vengatesan, K., & Singhal, A. (2020, April). Comparative analysis of data mining techniques to predict heart disease for diabetic patients. In International Conference on Advances in Computing and Data Sciences (pp. 507-518). Springer, Singapore.
[13] Kumar, M., Shambhu, S., & Sharma, A. (2018). Classification of heart diseases patients using data mining techniques. International journal of research in electronics and computer engineering, 6(3), 1495-1499.
[14] Manogaran, G., Varatharajan, R., & Priyan, M. K. (2018). Hybrid recommendation system for heart disease diagnosis based on multiple kernel learning with adaptive neuro-fuzzy inference system. Multimedia tools and applications, 77(4), 4379-4399.
[15] Mendis, S., Puska, P., Norrving, B. E., & World Health Organization. (2011). Global atlas on cardiovascular disease prevention and control. World Health Organization.
[16] Mozaffarian, D., Wilson, P. W., & Kannel, W. B. (2008). Beyond established and novel risk factors: lifestyle risk factors for cardiovascular disease. Circulation, 117(23), 3031-3038.
[17] Nagamani, T., Logeswari, S., & Gomathy, B. (2019). Heart disease prediction using data mining with mapreduce algorithm. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN, 2278-3075.
[18] Nahar, J., Imam, T., Tickle, K. S., & Chen, Y. P. P. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. Expert Systems with Applications, 40(4), 1086-1093.
[19] Palaniappan, S., & Awang, R. (2008, March). Intelligent heart disease prediction system using data mining techniques. In 2008 IEEE/ACS international conference on computer systems and applications (pp. 108-115). IEEE.
[20] Poirier, P. (2008). Healthy lifestyle: even if you are doing everything right, extra weight carries an excess risk of acute coronary events. Circulation, 117(24), 3057-3059.
[21] Preethi, M., & Selvakumar, J. (2020). A Survey of Predicting Heart Disease. JOIV: International Journal on Informatics Visualization, 4(2), 109-111.
[22] Rathi, M., & Gupta, A. (2021). Mobile-based prediction framework for disease

detection using hybrid data mining approach. In Proceedings of International Conference on Artificial Intelligence and Applications (pp. 521-530). Springer, Singapore.

[23] Sahoo, S., Das, M., Mishra, S., & Suman, S. (2021). A hybrid DTNB model for heart disorders prediction. In Advances in electronics, communication and computing (pp. 155-163). Springer, Singapore.

[24] Salman, O. H., Zaidan, A. A., Zaidan, B. B., Naserkalid, F., & Hashim, M. J. I. J. (2017). Novel methodology for triage and prioritizing using "big data" patients with chronic heart diseases through telemedicine environmental. International Journal of Information Technology & Decision Making, 16(05), 1211-1245.

[25] Shouman, M., Turner, T., & Stocker, R. (2012, March). Using data mining techniques in heart disease diagnosis and treatment. In 2012 Japan-Egypt Conference on Electronics, Communications and Computers (pp. 173-177). IEE

[26] Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications, 17(8), 43-48.

[27] Sowmiya, C., & Sumitra, P. (2021). A hybrid approach for mortality prediction for heart patients using ACO-HKNN. Journal of Ambient Intelligence and Humanized Computing, 12(5), 5405-5412.