Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/54

Hybrid Learning Methods for Accurate Prediction of Cardiovascular Disease

Vrince Vimal[1], Kumud Pant[2]

[1]Department of Computer Science & Engineering,Graphic Era Hill University, Dehradun, Uttarakhand India, 248002
[2]Department of Biotechnology,Graphic Era Deemed to be University, Dehradun, Uttarakhand India, 248002

ABSTRACT

Death from heart disease ranks high among the world's leading killers. Healthcare data analysis has a significant difficulty in the domain of cardiovascular disease forecast. The vast amounts of data generated by the healthcare business present significant opportunities for machine learning to aid in decision making and prediction. Additionally, recent innovations in various IoT domains have made use of deep - learning approaches (IoT). Predicting cardiovascular illness with machine learning techniques is only partially explored in existing research. In this study, we offer a new approach to heart disease diagnosis that makes use of methods of machine learning to identify critically important traits. Numerous feature combinations are introduced, together with several well-established classification methods, to serve as the basis for the prediction method. By combining Hybrid Randomized Forest as well as a Linear Model in cardiac disease predictions, we achieve performance improvements with an accuracy level of 86.47 percent.

Keywords: Cardiovascular disease; Prediction model; Machine Learning; Heart disease prediction; Feature selection; Classification algorithms

INTRODUCTION

Because of the complexity of the health conditions that go into heart disease diagnosis, like diabetes, hypertension, abnormal cholesterol levels, and also an irregular heartbeat, it is often difficult to make an accurate diagnosis. Human cardiovascular disease severity has been studied using a number of data mining and neural network methods. Numerous algorithms, such as K Nearest Neighbour (KNN), Decision Trees (DT), Genetic Algorithms (GA), & Naive Bayes are used to categorise illness severity [1], [2]. Cardiovascular disease is a difficult condition that necessitates careful management. Avoiding doing this is associated with increased risk of heart disease and even death. Metabolism syndromes are identified using a medical research lens & data mining method. Predictions of cardiovascular disease and other data investigations benefit greatly from data mining using classification. Heart disease related events can also be accurately predicted using decision trees [3]. Data mining techniques are employed in a number of different ways for prediction of heart disease, allowing for knowledge to be abstracted in a number of different ways.

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/54

As part of this effort, we have read a large amount of material in order to develop the prediction model that incorporates not just individual methods but also the interconnection of three or more methods

Hybrid approaches [4] refer to methods that combine existing ones with new ones. Pulse rate time data are used to establish neural networks. Left bundle branch block, right bundle branch block (RBBB), atrial fibrillation (AFIB), atrial flutter (AFL), sinus rhythm (NSR), premature ventricular contraction (PVC), sinus Bradycardia (SBR) and second-degree block (BII) are all examples of conditions that can be predicted using this method. Classification is done on the dataset using a radial basis function network (RBFN), with 70% of a input being employed for development & 30% for classifying [5]. We also provide an overview of a Computer Assisted Decision Support System (CADSS) and how it might be used in the scientific and medical communities.

Research has demonstrated that using data mining techniques in healthcare canspeed up the process of prediction for heart disease while increasing its precision. We advocate using the GA to identify cardiac issues. In this approach, the GA is used to infer successful association rules in random selection, crossover, as well as the mutation that ultimately leads towards the functional properties function's novelty. Specifically, we are using the Cleveland datasets, which is compiled from a UCI machine learning repository, to validate our experiments. In the following sections, we will demonstrate how our findings stand out in comparison to other well-known supervised learning methods. Particle Swarm Optimization (PSO), one of most potent evolutionary method, is presented, and some rules are produced for cardiovascular illness. Encoder rules have now been applied arbitrarily, leading to an increase in accuracy. Pulse rate, sex, age, and numerous other factors are used as predictors of cardiovascular disease. With the use of ML model and Neural Network models, we may achieve higher accurate and trustworthy outcomes, as demonstrated in. Within the scope of this paper, we provide a method known as the Hybrid Random Forest with Linear Model (HRFLM). The primary aim of this study is to enhance the effectiveness of cardiovascular disease prediction. There have been a lot of investigations, and as a result, there are now boundaries placed on selecting features for algorithms to work inside. On the other hand, the HRFLM approach does not limit itself to a particular subset of features but instead makes use of everything available. In this work, we employ a hybrid approach and conduct experiments to determine the characteristics of a machine learning model. Experimental results demonstrate that our suggested hybrid technique outperforms state-of-the-art methods in predicting cardiovascular disease. The remaining sections of the paper are as follows: In Part II, we will talk about the works done on the heart and the existing procedures and techniques used to complete them. In Section III, we also present a summary of our findings. HRFLM is discussed in Section IV. Initially, data must be cleaned and organised before moving on to the next steps of feature extraction, suitable classification model, then finally, evaluation of performance. The deployed algorithms and experimental framework are described in Section V. Data as well as experimental design evaluations are included in Section VI. The methods used and the final results are also presented.

Results from the HRFLM technique and comparisons to other models are discussed in Section VII. Part VIII concludes with a summary of the work done so far and some suggestions for further improvement. neurological disorders is where neural networks really shine. We utilise a model that

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/54

incorporates 13 factors for predicting cardiovascular disease. The outcomes demonstrate a higher degree of performance in comparison to the existing methods in literature such as Carotid artery stenting (CAS) is another popular treatment option in modern medicine. Major adverse cardiovascular events (MACE) are more likely to occur in older people with heart disease when the CAS is present. They take on a crucial role in the evaluating process.

We use an ANN, which has shown promising results in the prediction of cardiovascular illness [6]. In this study, we propose neural network methods, that integrate not just posterior probabilities but also estimated values from such a variety of prior approaches. The maximum accuracy of such a model being 89.01%, that is impressive in comparison to other studies. As we demonstrated in using a Neural Network on the Various heart data improves the performance of heart disease. The machine learning field has also advanced recently, with applications in the Internet of Things (IoT) . Accurate identification of networked IoT devices has indeed been demonstrated using ML algorithms applied to network traffic information. Meidan et al. gathered and categorised data on network activity from nine different Internet of Things gadgets, computers, and mobile phones. They developed a multi-stage meta classifiers using learning algorithm. In the first phase, the classifier is able to tell the difference between traffic from IoT as well as non-IoT sources. As a second step, a unique category for each type of IoT device is established. It has been shown that deep learning may be used to successfully extract useful information from large amounts of unstructured sensor data generated by Internet of Things devices operating in dynamic contexts [7, 8]. Multilayered deep learning is well-suited to the edge computing environment [9].

RELATED WORK

The disciplines that are immediately relevant to this publication are rich with prior research. Artificial neural networks (ANNs) have been implemented to make the most precise predictions possible with in medical industry [6]. An ANN technique called back propagation multilayer perception (MLP) is being utilised to make diagnoses of heart disease. We find that the obtained results are superior to those of previously existing models within the same field. NN, DT, Support Vector Machine and Naive Bayes are used to analyse data out from UCI cardiac lab to find trends. The results are compared to these algorithms in terms of speed and precision. The suggested hybrid method competes with other known methods, providing values of 86.8 percent for the F-measure. This paper introduces Convolutional Neural Networks (CNN) for classification without segments. As part of its training phase, this technique takes into account ECG signals containing heart cycles beginning at a variety of locations. CNN can create features from a variety of angles during patient testing stage Previous attempts to make good utilization of the medical industry in terms massive data dumps have largely failed. Prediction of cardiovascular disease can be simplified and made more accurate with the help of the new methods provided here, which also help to save costs.High precision is established in establishing the efficacy of the ML and DL algorithms explored within that study on prediction & categorization of cardiac disease.

In HRFLM, we search using UCI Cleveland database using a computational method with the 3 association rules in mine, which are known as apriori, predictive, & Tertius. These rules help us uncover the elements that contribute to heart disease. Based on the evidence that has been collected, one can draw the conclusion that women have a lower risk of developing heart disease in

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/54

comparison to men. Accurate diagnosis is the first and most important step in treating cardiac problems. The conventional methods, on the other hand, are not sufficient for producing reliable forecasts and diagnoses. The input for HRFLM consists of 13 clinical characteristics and is generated by an artificial neural network (ANN) trained by back propagation. The data that were acquired are then compared and analysed against by the conventional procedures. When the risk levels reach such a high level, accurate diagnosis of the disease requires the consideration of a variety of different characteristics. Because of the nature and complexity of cardiac disease, a treatment strategy that is both effective and comprehensive is required. In the sphere of medicine, using methods of data mining can be helpful in corrective situations. The methods of data mining are even further utilised, taking into consideration DT, NN, SVM, and KNN. The results from the support vector machine (SVM) appear to be effective in boosting accuracy in the prediction of illness, and is one of the methods that are applied. A nonlinear technique that includes a component for measuring heart activity has been developed in order to identify arrhythmias such as bradycardia, tachycardia, atrial, as well as atrial-ventricular flutters, as well as a great number of additional types. The precision of the end outcomes that are derived from ECG data allows for an accurate estimation of the performance efficacy of this method. [10] Both the correct identification of disease and the prediction of possible anomalies with in patient are accomplished through the utilisation using ANN training. In recent years, a variety of data mining methodologies with prediction methods, including KNN, LR, SVM, NN, and Vote, have gained a lot of traction in the quest to recognise and forecast cardiovascular disease. In this study, we suggest the unique method Vote, which should be used in conjunction with a hybrid strategy that makes use of both LR and NB. The trials of the suggested method are carried out using the UCI database, which results in an accuracy of 87.4% in the prediction of heart disease [11], Using three distinct data sets from Cleveland, Switzerland, and Hungarian in UCI correspondingly, a PCA method is proposed to evaluation. This technique will extract the vectors that have a large covariance in addition to the vectors projection that will be utilised to reduce its feature dimension. A radial basis function that enables kernel-based support vector machines (SVM) is given the feature selection along with the minimising dimension instruction. The outcomes of the approaches are as follows: 82.18% of the UCI data sets for Cleveland, 85.82% for Switzerland, and 91.30% for Hungarian [12]. The key original contribution of this study is the introduction of a hybrid method that combines linear regression (LR), multivariate adaptive regression splines (MARS), and artificial neural networks (ANN). This method is introduced using rough set methods. The recommended strategy was successful in reducing the number of essential characteristics. The subsequent step for ANN involves inputting the principle components. The data on cardiovascular disease are utilised to illustrate that the construction of a hybrid strategy is effective [13], [14]. It is hypothesised that cardiac illness can be predicted using multilayer neural network perception. This technique takes 13 clinical attribute features as the input, and after being trained via back propagation, it produces remarkably accurate results for determining whether or not an person has cardiovascular disease [39]. To make a much more accurate prediction of heart disease, we also present the Apriori algorithm combined with SVM and evaluate how it stacks up against 9 other classification strategies. The findings of both the classification approach have demonstrated a greater level of accuracy and performance in the forecasting of heart disease when comparison to the outcomes of the other methodologies that are currently in use [15]. In the process of diagnosing cardiac disease, the selection of relevant features is an important factor. The use of an ANN using back propagation

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/54

has been suggested as a method for improving disease prediction. The application of ANN yields findings that are exceptionally exact and exceedingly precise [16]. For the purpose of detecting heart disease, an algorithm that incorporates fuzzy NN and is given the name Recurrent Fuzzy Neural Network (RFNN) is presented. Within the UCI given dataset, a maximum of 297 different patient records were taken into consideration; of these, 252 are put into use for learning, while the other records are put into use for tests. In light of the evaluation, it has been determined that the outcomes live up to expectations [17]. The prediction of heart disease using SVM and ANN has been proposed. In this procedure, the premise of the correctness and timing of the testing is determined by using two different approaches. As can be seen in [18], the suggested framework divides the data records into two categories that can then be analysed further using SVM and ANN respectively. The Back Propagation Neural Network (BPNN) using classification approach is introduced. This is when the hypertension sequence was formed, and then, following that, the precise sequence is produced. The effectiveness of BPNN approaches has been evaluated through the training phase in addition to the testing stage using a variety of sample sizes. The results of these evaluations have been compared. The increased use of this method has led to greater accuracy [19], which can be directly correlated with the increase in the quantity of records.

PROPOSED METHOD

In this research, we classified heart disease data from the Cleveland UCI repository using a R studio rattle. A clear picture is painted for the user of the dataset, the workspace, and the process of developing predictive analytics. The ML process begins with data processing, continues with selecting features using DT entropy, then assesses the efficacy of the model used for classification and finally yields refined results. ForA many permutations of characteristics, feature extraction and modelling must be repeated. Values and their ranges are displayed in Table. Each model's generation and performance, along with the ML techniques being used create the products based on the 11 features, have been recorded. Data pre-processing is summarised in Section A, feature extraction with entropy is discussed in Section B, classification with ML techniques is explained in Section C, and efficiency is displayed in Section D.

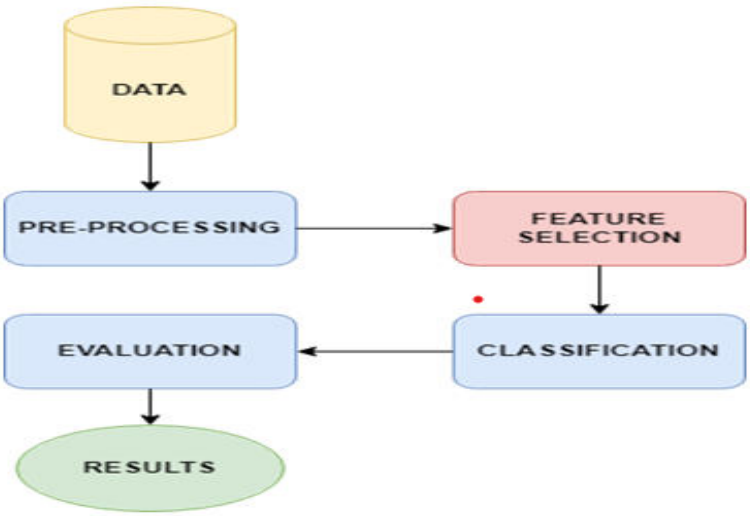

FIGURE 1 : Experimental workflow

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/54

A. PRE-PROCESSING

After compiling a wide range of sources, information on cardiovascular disease is put through a preliminary processing stage. Out of a total of 303 patient records, 6 have partial missing values. Six of the records were found to be unnecessary, so we removed them from dataset before beginning pre-processing on the remaining 297 health records. Features in the provided dataset are presented to the multiclass variable and binary classification. Testing for heart disease is done with the help of the multi-class parameter. A score of 1 signifies that the individual does have cardiovascular disease, whereas a score of 0 shows that perhaps the individual has no cardiovascular disease. For data analysis to begin, medical information must first be translated into concrete diagnostic values. From a total of 297 medical files, pre-processing determined that 137 used to have a value of 1 confirming the presence of heart disease, while the remaining 160 had a value of 0 showing the occurrence of heart illness.

B. FEATURE SELECTION AND REDUCTION

Just two of the dataset's eleven attributes—those relating to the patient's age and gender—are used to determine the individual's details. The nine remaining characteristics are also significant because they contain essential clinical data. Vital to diagnosis and understanding the extent of heart disease, clinical records are a must. Several machine learning (ML) methods, including NB, GLM, LR, DL, DT, RF, GBT, and SVM, are employed in this test. All ML methods were re-tested, this time with all nine attributes. HRFLM's illustrative prediction method is depicted in Figure 2.

C. CLASSIFICATION MODELLING

Datasets are grouped together using the characteristics of Decision Trees (DTs), which include criteria as well as variables. Afterwards, classifiers are used on each clustered dataset to gauge its efficacy. From these outcomes, the models with the lowest error rates are selected as the best performers. Selecting the DT cluster with the highest error rate and then extracting the features used by the relating classifier to optimise performance further. With this dataset, we can assess the classifier's efficiency in terms of minimising classification errors.

1) Language Model

For a statistics D input of length x_a and feature pairs x_a, x_b Following are the values that complete the solution to the linear equation f$(x) = mx + c$:

$$m = \frac{(\sum_a x_a y_a) - n\overline{x_a y_a}}{(\sum_a x_a^2) - n\overline{x_a}^2} \quad (1)$$

2) Decision Trees

The trees are built with inputs that have high entropy from information-dense training examples. An easy recursive partitioning technique at the top allows for rapid construction of these trees.

$$\text{Entropy} = -\sum_{k=1}^{n} p_{ab} \log_2 p_{ab} \quad (2)$$

3) Random Forest

To achieve optimal performance, this classifier constructs multiple decision trees and combines

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/54

them. Primarily, it uses a technique called bootstrap aggregation or bagging to learn trees. Bagging is repeated from b = 1 to B times for a set of data represented by $X = [x_1, x_2, x_3,..., x_n]$ with corresponding responses $Y = [x_1, x_2,..., x_n]$. The predictions D=1t o d are averaged from each tree on x' to create the unseen samples x'.

$$j = \frac{1}{D}\sum_{d=1}^{D} fd(x')$$

4) Neural Networks

Inputs x_j, hidden layers, and output y_j are all parts of a neuron. The output is generated using an activation function (such as a sigmoid) and a bias constant (c).

$$f\left(c + \sum_{j=1}^{n} x_j u_j\right) \quad (4)$$

D. PERFORMANCE MEASURES

Standard performance measures like accuracy, precision, and error in classification have been factored into the calculation of this model's performance efficacy. In this setting, accuracy is defined as the fraction of training data that produces predictions that are true. Accuracy is measured by how many times a prediction ends up being right for the "yes" category of cases. What we mean by "classification error" is the amount of missing or incorrect information in a given set of examples. Three performance metrics are used to identify the important features of cardiovascular disease, with the goal of better examining the behavior of the different feature combinations. In the ML method, the goal is to find a model that performs better than any others already in use. We present HRFLM, which reduces classification error and improves accuracy in predicting heart disease. Each classifier's performance is analysed separately, and the aggregated data is carefully stored so that it can be analysed later.

EXPERIMENTAL SETUP

To classify heart diseases found in the Cleveland UCI machine learning repository, we employed a R studio rattle. The experiment assessment process is shown in sequential steps in Figure 1. Data from the UCI dataset is loaded and prepared for pre-processing in the first stage. From the pre-processed set of data of cardiovascular disease, a subset of 11 attributes is chosen. The classification is based on the results from the three previously established models for predicting cardiovascular disease (DT, RM, LM). Model performance is measured with a confusion matrix. True positive (TP), true negative (TN), false positive (FP), and false negative (FN) are the four possible results from the confusion matrix to determine precision, sensitivity, and specificity, we employ the following metrics.

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/54

TABLE 1 : RESULT OF PROPOSED MODEL WITH OTHER MODELS

Models	Accuracy	Precision	Classification error	Specificity	F-measure	Sensitivity
Naive Bayes	73.8	88.5	22.2	58	82.5	77.8
Generalized Linear Model	83.1	86.8	12.9	18	89.6	92.9
Logistic Regression	80.9	87.6	15.1	23	88.2	89.1
Deep Learning	85.4	88.7	10.6	31.3	90.6	93
Decision Tree	83	84	13	0	89.8	96.8
Random Forest	84.1	85.1	11.9	8	90.4	96.8
VOTE	85.41	89.2	10.59	-	82.4	-
Proposed Model	86.4	87.1	9.6	80.6	88	90.8

EVALUATION RESULTS

Models for making predictions are built with 11 features, and their accuracy is determined using statistical methods. Table 1 provides an overview of the most effective approaches to classification. Data on sensitivity, specificity, F-measure, and accuracy are all summarised below. HRFLM classification method achieves highest accuracy among existing methods.

Performance of Classifiers with all the features

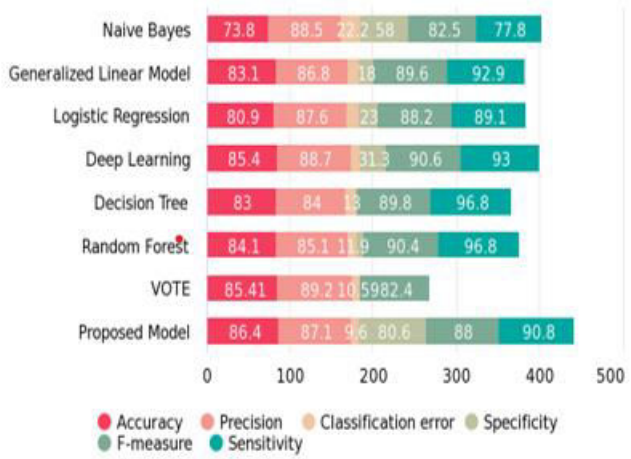

FIGURE 2: Performance of Classifiers with all the features

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/54

CONCLUSION

Through into the analysis and identification of raw health records of cardiac information, we can save lives and uncover anomalies in cardiovascular issues at an early stage. Using machine learning techniques, we have been able to uncover previously unreported data and arrive at some surprising findings on coronary heart disease. Heart disease prognostication is a vital yet challenging area of medical practise. However, with proper diagnosis and treatment, the likelihood of mortality can be dramatically lowered. Ideally, the research would expand to incorporate real-world datasets in addition to theoretical approaches and simulations. The suggested HRFLM approach integrates the most useful aspects of the Linear Method and the Random Forest (RF) (LM). HRFLM was found to be highly effective in predicting cardiovascular disease. In the future, this research can benefit from using a wider variety of machine learning technique combinations to enhance prediction approaches. Novel feature selection approaches can be created to acquire a more holistic knowledge of the relevant characteristics, hence improving the effectiveness of heart disease predictions.

REFERENCES

1. Durairaj, M. Revathi, V., 2015. Prediction Of Heart Disease Using Back Propagation MLP Algorithm., 4(08), pp.235−239.
2. Gavhane, A., 2018. Prediction of Heart Disease Using Machine Learning. 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), (Iceca), pp.1275−1278.
3. Al-milli, N., 2013. Backpropogation Neural Network for Prediction of Heart Disease., 56(1), pp.131−135.
4. Jpdlo, V. et al., 2018. Heart diseases prediction with Data Mining and Neural Network Techniques. , 6(7 2), pp.1−6.
5. A. Devi,S. Rajamhoana, C, K. Umamaheswari, R. Kiruba, K. Karunya and R. Deepika, Analysis of Neural Networks Based Heart Disease Prediction System, 2018 11th International Conference on Human System Interaction (HSI), Gdansk, 2018, pp. 233-239.
6. Baccour, L., 2018. Amende d fuse d TOPSIS-VIKOR for classification (ATOVIC) applied to some UCI data sets R. Expert Systems With Applications, 99, pp
7. J. Wu, S. Luo, S. Wang and H. Wang, "NLES: A Novel Lifetime Extension Scheme for Safety-Critical Cyber-Physical Systems Using SDN and NFV," IEEE Internet of Things Journal, no. 6, no. 2, pp. 2463-2475, 2018.
8. G. Li, J. Wu, J. Li, K. Wang, T. Ye, Service Popularity-based Smart Resources Partitioning for Fog Computing-enabled Industrial Internet of Things, IEEE Transactions on Industrial Informatics, vol. 14, no. 10, pp. 4702-4711, Oct. 2018.
9. Li, He, Kaoru Ota, and Mianxiong Dong. "Learning IoT in edge: deep learning for the internet of things with edge computing." IEEE Network 32, no. 1 (2018): 96-101.
10. Raju, C. et al., 2018. Mining Techniques. 2018 Conference on Emerging Devices and Smart Systems (ICEDSS), (March), pp.253−255.
11. Sabahi, F., 2018. Bimodal fuzzy analytic hierarchy process (BFAHP) for coronary heart disease risk assessment. Journal of Biomedical Informatics, 83(April), pp.204−216. Available at: https://doi.org/10.1016/j.jbi.2018.03.016.
12. Shafenoor Amin, M., Kia Chiam, Y. Dewi Varathan, K., 2018. Identification of significant features and data mining techniques in predicting heart disease. Telematics and Informatics.

Webology, Volume 18, Number 5, 2021
ISSN: 1735-188X
DOI: 10.29121/WEB/V18I5/54

Available at: https://linkinghub.elsevier.com/retrieve/pii/S0736585318308876.

13. Shah, S.M.S. et al., 2017. Feature extraction through parallel Probabilistic Principal Component Analysis for heart disease diagnosis. Physica A: Statistical Mechanics and its Applications, 482, pp.796−807. Available at: http://dx.doi.org/10.1016/j.physa.2017.04.113.
14. S. Radhimeenakshi, "Classification and prediction of heart disease risk using data mining techniques of Support Vector Machine and Artificial Neural Network," 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 3107-3111
15. Sonawane, J.S. Student, P.G., 2014. Prediction of Heart Disease Using Multilayer Perceptron Neural Network. , (978).
16. Sowmiya, C., 2017. Analytical Study of Heart Disease Diagnosis Using Classification Techniques.
17. Tran, V.P. Al-jumaily, A.A., 2017. Non-Contact Doppler Radar Based Prediction of Nocturnal Body Orientations Using Deep Neural Network for Chronic Heart Failure Patients. , pp.3−7.
18. Vivekanandan, T. Sriman Narayana Iyengar, N.C., 2017. Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease. Computers in Biology and Medicine, 90(April), pp.125−136.
19. Williams, O. et al., 2017. An integrated decision support system based on ANN and Fuzzy AHP for heart failure risk prediction. Expert Systems With Applications, 68, pp.163−172.