Determination of Significant Features for Building an Efficient Heart Disease Prediction System

Ekta Maini, Bondu Venkateswarlu, Arbind Gupta

Abstract*: *Heart diseases are responsible for the greatest number of deaths all over the world. These diseases are usually not detected in early stages as the cost of medical diagnostics is not affordable by a majority of the people. Research has shown that machine learning methods have a great capability to extract valuable information from the medical data. This information is used to build the prediction models which provide cost effective technological aid for a medical practitioner to detect the heart disease in early stages. However, the presence of some irrelevant and redundant features in medical data deteriorates the competence of the prediction system. This research was aimed to improve the accuracy of the existing methods by removing such features. In this study, brute force-based algorithm of feature selection was used to determine relevant significant features. After experimenting rigorously with 7528 possible combinations of features and 5 machine learning algorithms, 8 important features were identified. A prediction model was developed using these significant features. Accuracy of this model is experimentally calculated to be 86.4%which is higher than the results of existing studies. The prediction model proposed in this study shall help in predicting heart disease efficiently.

Index Terms*: *Feature selection, heart disease prediction Machine Learning

I. INTRODUCTION

Heart diseases are responsible for the greatest number of deaths worldwide. The situation is alarming in low- and middle-income countries. Nearly 3 million people died because of heart diseases India during past 2 years [1]. The healthcare industry in developing countries like India faces a lot of challenges. The cost of medical diagnostics and treatment is not affordable for a major portion of the population. There is a shortage of physicians, especially in the rural areas. There is a lack of awareness among the people regarding healthy lifestyle habits. Because of all these factors, people do not visit the doctor/hospital for the regular medical checkups and the disease is not diagnosed in time. The disease is usually detected when the symptoms like weakness of physical body, swollen feet, shortness of breath, cold sweats and fatigue develop. If the disease is detected in the advanced stage, the cost of treatment gets higher and survival rate decreases. Thus, there is a great necessity to develop cost effective and easily accessible tools which can diagnose the heart diseases in time. Past few years, the concepts of machine learning and deep learning have used to develop prediction models for various diseases [2].Such models are a great step towards providing affordable and quality healthcare to the masses. The performance of such prediction system not only depends on the choice of machine learning algorithm used but also on the quality of medical data. Raw medical data may have a few missing values, some outliers , some irrelevant or redundant features. Feature selection is an important technique in data preparation where the irrelevant and redundant features are excluded to build more efficient systems .Literature mentions the usage of feature selection rules in medical disease prediction [3]. Study done by Kavitha et al [4] highlighted the fact that redundant, irrelevant and inconsistent data adversely affects the performance of the prediction system. Dey et al,2016 [5] used PCA algorithm of feature selection which resulted in achieving 80% accuracy for prediction system . Studies done by Nahar [6], Mohammad Shafenoor Amina et al [7] and Ali Muhammad Usman et al [8] highlight the importance of exploring more techniques of feature selection in order to build high performance systems. Jabbar et al. [9] used feature selection methods for identifying correlation in the medical data. The resulting model attained an accuracy of 81% in forecasting cardiac ailments. Study done by Shah et al [10] was based on the usage of probabilistic principal component analysis for feature selection .The resulting system attained an accuracy of 82.18% .In the research done by Vivekanandan and Iyengar, an accuracy of 83% was obtained by applying modified differential evolution for feature selection [11].

In this paper, an exhaustive attempt has been made to employ a novel and robust algorithm for feature selection which would result in building an accurate and efficient prediction system. Heart disease dataset provided by UCI machine learning repository was utilized in this work [12]. This dataset has been used for many research studies as this dataset needs minimal preprocessing .Five techniques namely logistic regression, k-NN, decision trees ,Naïve Bayes, and Support Vector Machine were used for designing the prediction models. Brute force method was used for feature selection. selection.7528 possible combinations of input attributes were

Revised Manuscript Received on July 12, 2019.

Ekta Maini, Department of Computer Science &Engineering Dayananda Sagar University ,Bengaluru, India Email:ekta.marwaha@gmail.com

Bondu Venkateswarlu, , Department of Computer Science & Engineering Dayananda Sagar University ,Bengaluru, India .Email: bonduvenkat-cse@dsu.edu.in

Arbind Gupta, Department of Computer Science &Engineering ,Dayananda Sagar College of Engineering ,Bengaluru, India. Email: arbind.gupta@gmail.com

Retrieval Number: B3393078219/19©BEIESP
DOI: 10.35940/ijrte.B3393.078219

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

experimented to determine the significant features. After conducting these experiments, 8 features were identified to be significant. Top 3 machine learning techniques were also recognized by experimental results. Based on these results, a prediction model was developed whose performance was also checked on UCI Statlog Heart disease dataset[13].The prediction model developed in this study achieved an accuracy of 86.4% which is quite appreciable when compared to the existing studies. This research contributes significantly by proposing a novel way of feature selection which achieves high accuracy and reduced risk of overfitting the data. This model is quite efficient and can be used by medical practitioners effectively for prevention of heart disease.

A thorough explanation of the Cleveland dataset has been done in Section II. Section III discusses the methodology used to develop the model. It includes steps for preprocessing the data, selecting the important features, training and testing the prediction model followed by evaluating the performance. The results of the study have been discussed thoroughly in section IV. Conclusion and future scope of this study have been discussed in the last section of the paper.

II. DESCRIPTION OF THE MEDICAL DATASET

UCI machine learning repository has four databases for heart disease prediction. These include heart disease datasets of VA Long Beach, Hungary, Cleveland, and Switzerland [12]. Out of these four datasets, Cleveland heart disease dataset was chosen for this study as this dataset has the most complete medical records with very little missing data. This dataset provides information of 303 medical records. There are 76 parameters in the dataset, of which, the details of 14 attributes are provided by the repository. The comprehensive description of this dataset is given in Table I. There are 13 input clinical attributes that act as input attributes. There is a 'Num' attribute which is the predicted risk score of heart disease. This attribute acts as an output attribute. The predicted risk score takes the value between 0 to 4. A score of 0 signifies little danger of heart illness while a score in range 1-4 indicates more likelihood of getting a heart related medical problem. Thus, it can be said that a risk score 0 indicates healthy person while score 1-4 shows vulnerability of a patient to develop some heart ailment. Distribution of the risk score over 303 records of Cleveland dataset is shown in Fig.1.

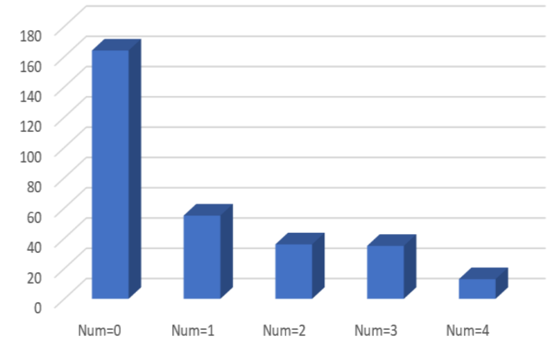

Fig.1Distribution of 'Num' in the dataset

Table I UCI heart Disease Dataset [12]

S. No	Feature name	Feature code	Explanation	Datatype
1	Age	Age	Age in years	Numeric

2	Gender	Gender	Female =1
			Male=0
				Nominal

3	Serum Cholesterol	Chol	In mg/dl	Numeric
4	Fasting blood sugar	Fbs	0: fbs<120mg/dl 1: fbs<120mg/dl	Nominal
5	Blood pressure(resting)	Trestbps	In mm Hg	Numeric
6	Type of pain in chest	cp	1 =typical angina 2=atypical angina 3 = non-anginal 4= asymptomatic	Nominal
7	Electrocardiographic results	Restecg	0 = normal 1 = having ST-T wave defect 2= hypertrophy	Nominal
8	Heart rate achieved(max)	Thalach	Heart rate	Numeric
9	Exercise induced angina	Exang	1=Yes 0=No	Nominal
10	Slope of peak exercise ST segment	Slope	1 = up sloping 2 = flat 3 =down sloping	Nominal
11	Old peak for ST depression	Old peak	ST depression induced by exercise relative to rest	Numeric
12	Number of major vessels colored by fluoroscopy	Ca	Number of major vessels (0–3) colored by fluoroscopy	Nominal
13	Thallium scan	Thal	Heart Status 3 =normal 6 =fixed defect 7=reversible defect	Nominal
14	Risk score of heart disease	Num	Presence/absence Of disease 0=Low risk 1–4= High risk of heart disease	Nominal

III. METHODOLOGY

This study was carried out with an aim to classify the records on the basis of risk score of heart diseases. A risk score of 0 shall reflect a low risk of heart disease while a risk score of 1 indicates a high risk of heart disease. All the computations were performed in Python programming language. The complete study was carried out in 4 stages namely (i) preprocessing of dataset to remove the missing values (ii) feature selection to identify the significant features (iii) training the model using classification algorithm (iv) estimating the performance of the system.

The flow diagram is represented in Fig. 2. Feature selection and model creation were carried out iteratively for all combinations of features. For each iteration, the performance of each model was recorded. A detailed description of all the steps is done in sub-sections.

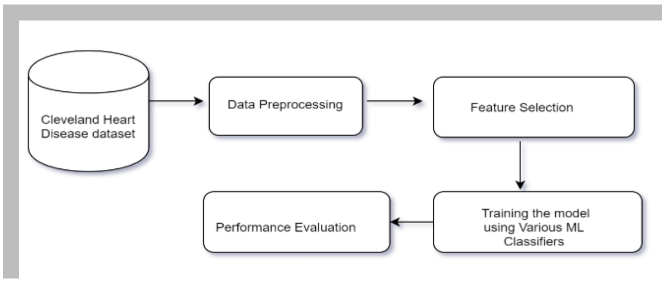
Cleveland Heart Disease dataset
Data Preprocessing
Feature Selection
Training the model using Various ML Classifiers
Performance Evaluation

Fig.2 Scheme of the proposed project

A. Preprocessing of dataset: This is the first and foremost step in a data science project. Preprocessing plays a very crucial role in building a robust prediction system [14]. During the exploratory data analysis, it was observed that the Cleveland dataset has 6 records which were incomplete and had missing values. It was decided to remove such records. The remaining dataset had 297 records. It was observed that 'num' attribute had values in the range 0 to 4.The values between 1 to 4 indicate high risk of heart disease while 0 shows a low risk of the disease. It was decided to change 'Num' into binary class. The records where the risk score is between 1-4 were kept in one class while the records indicating low risk were kept in a separate class. The final dataset had only 2 possible values of 'num' 0 or 1. '1' reflects an existence of disease while '0' shows nonappearance of heart disease. After performing these operations, it was observed that the dataset had 297 records, of which 160 records had high danger of heart disease while remaining 137 had low danger of heart disease. This distribution is shown in Fig.3. Here, 'Num'=0 is regarded as minimum danger of heart illness or nonappearance of heart disease while 'Num'=1 reflects a high danger of heart illness.

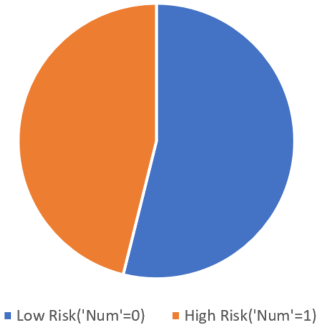

Fig.3 Distribution of risk score of heart disease in preprocessed dataset

B. Feature selection: It is observed from various studies that there may be a few attributes in the dataset which don't have much importance in the prediction of output. Either such attributes are unrelated, or these are redundant. These attributes reduce the accuracy of the system and make the execution slow. Feature selection is carried out to remove such features. In this study all the possible combinations of attributes were considered along with 5 machine learning classifiers namely Support Vector Machine ,decision tree ,logistic regression, k-NN, and Naïve Bayes to develop the prediction model. Brute force method was used with a lower bound of 4 features. Initially, each possible combination of 4 features was tested with 5 machine learning techniques. Later, all conceivable combination of 5 attributes were tested with above mentioned machine learning techniques. In the say way, this procedure was repeated for all the features.

It is known that the total number of combinations for 'p' features is 2^p. Excluding the empty set, the possible number of combinations is 2^p-1. In this study, minimum 4 features need to be selected in a subset. It means that all the combinations with 1,2 and 3 attributes are neglected.

Thus, the total possible combinations of features can be calculated as

$$2^p - \left(\frac{p!}{1!\,(p-1)!}\right) - \left(\frac{p!}{2!\,(p-2)!}\right) - \left(\frac{p!}{3!\,(p-3)!}\right) - 1$$

In Cleveland dataset, there are 13 input attributes. Substituting p=13 in the above formula resulted in a total of 7528 possible combinations. All these combinations were selected and tested during this work.

C. Training the model: Studies have shown the presence of powerful machine learning techniques. In this study 5 such algorithms were used to train the model. After the feature selection ,models were created by using these techniques. It is necessary to validate the performance of the built model. We used k-fold cross validation method in this study. The entire dataset was segregated into 10 parts. Out of these parts, 1 subset was used to train the model while 9 subsets are used to test the model. The process was carried out iteratively 10 times. The results were recorded by examining the mean of each 10 iterations. Stratified sampling was used to divide the subsets so that the all subsets had an equal class ratio as the original dataset.

D. Evaluation of the model: The performance of a system can be measured through various criteria like accuracy, F-score and precision of the system [15]. Accuracy is calculated by computing the mathematical ratio of the number of correct estimates to the total number of predictions. The model should have a high value of accuracy. Ratio of correct predictions to the positive class is called precision. Weighted mean of recall and precision is calculated as F-score. The overall behavior of the system is well understood using these performance indices.

IV. RESULTS AND DISCUSSIONS

The results obtained in this research work have been showcased in this section. The results of this study show that out of 13 attributes, 8 were found to be quite substantial for envisaging the risk score of heart disease. These are: gender, type of chest pain, fasting blood sugar ,exercise induced angina, old peak from ST depression, slope of peak exercise ST segment, number of major vessels colored by fluoroscopy and findings of Thallium scan test. In the dataset, these are represented by attributes 'gender', 'cp', 'fbs', 'ca', 'exang' ,

DOI: 10.35940/ijrte.B3393.078219

’oldpeak’ , ’slope’ and ’thal’ ,respectively. The performance of Support Vector Machine and Naïve Bayes was found to be the best in terms of precision, F-score and accuracy.

Five machine learning algorithms were applied on all the 7528 combinations of attributes. The performance was evaluated for each case. Accuracy, precision and F-measure were recorded. These results have been tabulated in Tables II-IV, respectively. It is clear from the results tabulated in Table II that SVM achieved the highest accuracy (86.7%) with 7 attributes (‘gender’, ‘cp’, ‘chol’, ‘fbs’, ‘slope’, ‘ca’, ‘age’).

Table II. Significant features for the best accuracy

ML algorithm	Accuracy	Significant Features
Decision Tree	82.1%	‘cp’, ‘fbs’, ‘gender’, ‘ca’, ‘thal’, ‘oldpeak’, ‘slope’, ‘exang’
k-NN	82.0%	‘cp’, ‘gender’, ‘ca’, ‘fbs’, ‘thal’, ‘old peak’
Logistic Regression	85.31%	‘cp’, ‘gender’, ‘age’, ‘restecg’, ‘thal’, ‘ca’, ‘slope’,’ exang’
Naïve Bayes	85.62%	‘Gender’, ‘cp’, ‘thalach’, ‘ca’, ‘old peak’, ‘exang’, ‘fbs’, ‘thal’
SVM	86.7%	‘Gender’, ‘cp’, ‘chol’, ‘fbs’, ‘slope’, ‘ca’, ‘age’

As indicated by Table III, k-NN and decision tree algorithms obtained the highest precision using 4 attributes. These attributes are ‘gender’, ‘exang’, ‘restecg’, ‘cp’.

Table III Significant features for the best precision

ML algorithm	Precision	Significant Features
Logistic Regression	85.4%	‘cp’, ‘fbs’, ‘gender’, ‘ca’, ‘thal’, ‘gender’, ‘trestbps’, ‘old peak’
SVM	85.63%	‘cp’, ‘gender’, ‘ca’, ‘chol’, ‘thalach’, ‘slope’
Naïve Bayes	87.31%	‘cp’, ‘gender’, ‘age’, ‘thal’, ‘ca’, ‘slope’, ‘old peak’, ‘thalach’
Decision Tree	93.62%	‘gender’, ‘exang’, ‘restecg’, ‘cp’
k-NN	94.7%	‘gender’, ‘exang’, ‘restecg’, ‘cp’

The F-score achieved by the machine learning algorithms is reflected in Table IV. It is evident that the highest f-score was obtained on 9 attributes by SVM. These are ‘cp’, ‘thalac’, ‘restecg’, ‘ca’, ‘slope’, ‘thal’, ‘old peak’, ‘trestbps’ and ‘exang’.Performance of k-NN and decision tree was found to be best on precision but not on accuracy and F-score.

Table IV Significant features for the best F-score

ML algorithm	F-score	Significant Features
Decision Tree	83.2%	‘cp’, ‘fbs’, ‘gender’, ‘restecg’, ‘ca’, ‘thal’, ‘trestbps’, ‘oldpeak’
k-NN	83.1%	‘cp’, ‘gender’, ‘ca’, ‘oldpeak’, ‘thal’, ‘chol’, ‘fbs’
Logistic Regression	84.6%	‘cp’, ‘gender’, ‘age’, ‘thal’, ‘ca’, ‘slope’, ‘exang’, ‘restecg’, ‘chol’, ‘fbs’
Naïve Bayes	86.62%	‘gender’, ‘exang’, ‘slope’, ‘ca’, ‘cp’, ‘age’
SVM	87.7%	‘cp’, ‘restecg’, ‘exang’, ‘slope’, ‘ca’, ‘thal’, ‘old peak’, ‘trestbps’, ‘thalach’

It is observed that the features like gender, fasting blood sugar, cholesterol and type of chest pain are significant in predicting the risk of heart diseases. Attributes like serum cholesterol and fasting blood sugar are highly dependent on a person’s lifestyle and one should adopt a healthy lifestyle to keep these parameters in control. This shall reduce the susceptibility of a person to develop heart disease.

Average accuracy, precision and F-measure of all machine learning techniques on all combinations of features is tabulated in Table V. It is evident from the table that Naïve Bayes and SVM achieved maximum accuracy and precision while Logistic Regression and SVM yielded the best results for F-score. This information is pictorially represented in Fig 4.

Table V Performance of classifiers on all combination of features in UCI Cleveland dataset

ML algorithm	Avg. accuracy	Avg. Precision	Avg. F-score
Logistic Regression	81.7%	76.2%	81.30 %
Decision tree	73.05%	66.43%	65.8 %
k-NN	75%	66.4%	65.3 %
Naïve Bayes	83.8%	78.7%	80.1 %
SVM	85.2%	78.1%	80.2 %

Fig 4. Performance of ML algorithms on all 13 features of Cleveland dataset

The models with maximum performance metrics were investigated to identify the frequency of occurrence of an attribute. Among the likely combinations of 7528 cases, sets of features which performed well for a particular machine learning technique were selected. Careful monitoring was done to identify the combinations which yielded the best results. All the features which were a part of these combinations were noted.Fig.5 clearly depicts the frequency of occurrence of such attributes. It is clear that the attributes ‘gender’ and type of chest pain ‘cp’ appeared the greatest number of times in achieving the best performance. Appearance over 7 times in the best performance metrics combinations was the qualifying criterion for an attribute to be considered as significant. The attributes which featured less than 7 times were ignored. It is evident that the selection of these attributes results in better performance of the prediction system.

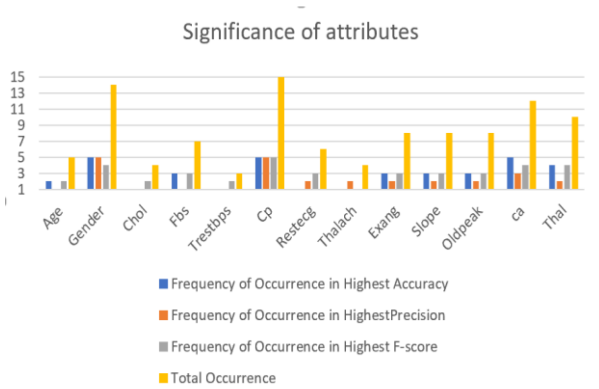

Fig.5 Significance of attributes as measured by the frequency of occurrence in accuracy, precision and F-score

On this basis, we determined that 8 important attributes that result in the best performance of the prediction system are: ‘gender’ ,’cp’ ,’ca’ ,’thal’ ,’old peak’ ,’slope’, ‘exang’ and ‘fbs’.

Out of the 13 attributes, remaining 5 insignificant attributes were removed. A prediction model was developed on the 8 significant features using Naive Bayes and SVM techniques. The evaluation of this study was carried out on UCI Statlog dataset. This dataset is similar to Cleveland dataset as it has the same set of attributes. This set has 270 records with no missing values. k-fold cross validation algorithm (k=10) was used to evaluate the models. 2 models were built .All features were used in the first case while only 8 significant features were used in the second case. Accuracy of both models was calculated. Comparison of performance of both the models is discussed in Table VI.

Table VI Performance evaluation of two prediction models

ML algorithms	Accuracy with all 13 Features	Accuracy with 8 significant features
SVM	85.2%	**86.40%**
Naïve Bayes	83.8%	**84.30%**
Logistic Regression	81.70%	**82.90%**

The importance of feature selection algorithm in building high performance prediction model is evident from Table VI. Accuracy of model developed using Naïve Bayes and Logistic Regression increased from 83.8% to 84.3% and 81.7% to 82.9% respectively. Performance of SVM was found to be better than the other machine learning algorithms. Accuracy of SVM increased from 85.2% to 86.4% when the insignificant features were removed. This is graphically shown in Fig 6.

These experimental results clearly reflect that the algorithm developed for feature selection in this study is quite efficient in increasing the accuracy of the system.

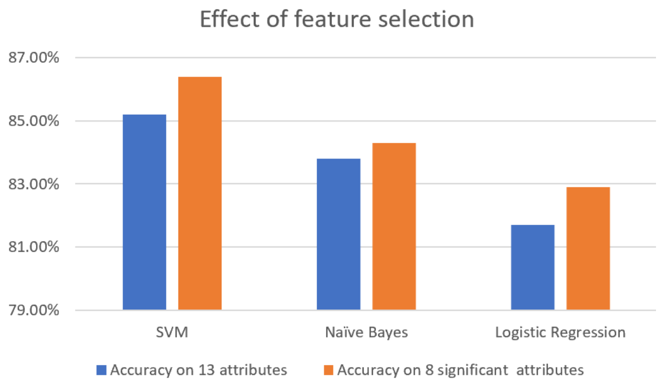

Fig.6 Analysis of performance on UCI Statlog dataset

V. CONCLUSION

Healthcare sector generates a lot of data. Machine learning algorithms can effectively fetch interesting information from this raw data. Such information can be used to build prediction models which can assist in diagnosing the heart diseases at an early stage. The performance of these models can be increased by removing insignificant features from the medical data. We conducted experiments on Cleveland heart disease dataset to identify significant features and the best machine learning algorithms . 7528 trials of all possible combination of features were conducted in this study to develop a robust and efficient heart disease prediction system.8 attributes namely ‘gender’ ,’cp’ ,’ca’ ,’thal’ ,’old peak’ ,’slope’, ‘exang’ and ‘fbs’ in the dataset were found to be significant. Performance of SVM machine learning algorithm was found to be the best. A Prediction model was built on 8 significant features using SVM machine learning algorithm. The performance of this system was evaluated on UCI Statlog dataset. It was observed that the accuracy of this model was 86.4% which is more than the existing studies. We plan to extend this research by conducting similar experiments on real dataset taken from an Indian hospital. It is also proposed to apply other feature selection techniques to identify important features which may increase the accuracy of the prediction model further.

REFERENCES

1. India State-Level Disease Burden Initiative Collaborators,” Nations within a nation: variations in epidemiological transition across the states of India”, 1990–2016 in the Global Burden of Disease Study. Lancet 2017; 390:2437–60.
2. Abdelaziz, A., Elhoseny, M., Salama, A. S., & Riad, A. M,” A machine learning model for improving healthcare services on cloud computing environment ”, Measurement, 2018 , pp 117–128 .doi:10.1016/j.measurement.2018.01.022
3. Paul, A.K., Shill, P.C., Rabin, M.R.I., Akhand, M.A.H.,” Genetic algorithm based fuzzy decision support system for the diagnosis of heart disease. (ICIEV)” In: 5th International Conference on Informatics, Electronics and Vision. IEEE, pp. 145–150,2016

4. Kavitha, R., Kannan, E., “An efficient framework for heart disease classification using feature extraction and feature selection technique in data mining” International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS),2016 pp. 1–5
5. Dey, A., Singh, J., Singh, N., “Analysis of supervised machine learning algorithms for heart disease prediction with reduced number of attributes using principal component analysis”, Analysis 140 (2) 2016., 27–31.
6. Nahar, J., Imam, T., Tickle, K.S., Chen, Y.P.P” Computational intelligence for heart disease diagnosis: a medical knowledge driven approach” Expert Syst. Appl. 2016. 40 (1), 96–104.
7. Mohammad Shafenoor Amina, Yin Kia Chiama, Kasturi Dewi Varathan,"Identification of significant features and data mining techniques in predicting heart disease”, Telematics and Informatics 36 (2019) 82–93
8. Ali Muhammad Usman, Umi Kalsom Yusof, Syibrah Naim, "Cuckoo inspired algorithms for feature selection in heart disease prediction" International Journal of Advances in Intelligent Informatics Vol. 4, No. 2, July 2018, pp. 95-106
9. M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, “Prediction of Heart Disease Using Random Forest and Feature Subset Selection,” 2016, pp. 187–196, doi: https://doi.org/10.1007/978-3-319-28031-8_16.
10. S. M. S. Shah, S. Batool, I. Khan, M. U. Ashraf, S. H. Abbas, and S. A. Hussain, “Feature extraction through parallel Probabilistic Principal Component Analysis for heart disease diagnosis,” Phys. A Stat. Mech. its Appl., vol. 482, pp. 796–807, Sep. 2017, doi: https://doi.org/10.1016/j.physa.2017.04.113.
11. T. Vivekanandan and N. C. S. N. Iyengar, “Optimal feature selection using a modified differential evolution algorithm and its effectiveness for prediction of heart disease,” Computers in Biology,vol.90,pp.125–136,Nov.2017,doi:https://doi.org/10.1016/j.compbiomed.2017.09.011.
12. https://archive.ics.uci.edu/ml/datasets/heart+disease
13. https://archive.ics.uci.edu/ml/support/Statlog+(Heart)
14. Ekta Maini and Bondu Venkateswarlu," Implementing an End to End Solution for Data Science Project Cycle-A Complete Roadmap for Data Aspirants" International Journal of Modern Electronics and Communication Engineering (IJMECE) Volume No.-7, Issue No.-3, May, 2019
15. Maini E., Venkateswarlu B., Gupta A., Applying Machine Learning Algorithms to Develop a Universal Cardiovascular Disease Prediction System. In: J. Hemanth et al. (Eds.): ICICI 2018, LNDECT 26, pp. 627–632, 2019. doi:10.1007/978-3-030-03146-6_69

AUTHORS PROFILE

Ekta Maini received her B.Tech degree from Kurukshetra University in 2002.She studied ME from Panjab University in 2004.Currently she is a research scholar in Dayananda Sagar University, Bangalore. Her interest areas include Data Mining and Machine Learning. She has participated in many national and international conferences.

Bondu Venkateswarlu received his Ph.D. degree in 2016. He is currently working as an Associate Professor in the Department of Computer Science and Engineering of Dayananda Sagar University, Bengaluru. India. His current research interests include Data Mining, Soft Computing Techniques & Software Engineering, He has published many papers in refereed journals.

Arbind Gupta did BE,ME and PhD from IIT Kharagpur. He is currently working as Professor in Department of Computer Science and Engineering in Dayananda Sagar College of Engineering. His areas of interest include image processing ,computer vision and medical imaging. He is actively involved in research activities.