International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2019): 7.583

A Hybrid Machine Learning Technique for Prediction of Heart Diseases

Reshu Choubey[1], Dr. Nishchol Mishra[2], Dr. Jitendra Agrawal[3], Yogendra Pratap Maravi[4]

[1]M. Tech Scholar, School of Information Technology, RGPV, Bhopal, India

[2]Associate Professor, School of Information Technology, RGPV, Bhopal, India

[3]Associate Professor, School of Information Technology, RGPV, Bhopal, India

[4]Assistant Professor, School of Information Technology, RGPV, Bhopal, India

Abstract: ***The heart is significant organ of human body part Heart disease describes a scope of conditions that influence your heart. Machine learning (ML) is the scientific study of algorithms and statistical models that PC systems use to play out a specific task without using unequivocal instructions, depending on patterns and deduction instead. There are various algorithms, which predict the heart disease. Accuracy is a key parameter to judge the algorithm. This paper proposed the hybrid algorithm based on naïve bayes with random forest. Accuracy, classification error, F-measure, Recall and Precision parameters are calculated. 92% accuracy achieved by proposed hybrid algorithm.***

Keywords: Machine learning, F-measure, Recall and Precision

1. Introduction

Heart disease is perhaps the biggest cause of grimness and mortality among the number of inhabitants on the planet. Prediction of cardiovascular disease is viewed as one of the most significant subjects in the section of clinical information analysis. The measure of information in the healthcare industry is tremendous. Information mining turns the huge assortment of crude healthcare information into data that can assist with making educated decisions and predictions.

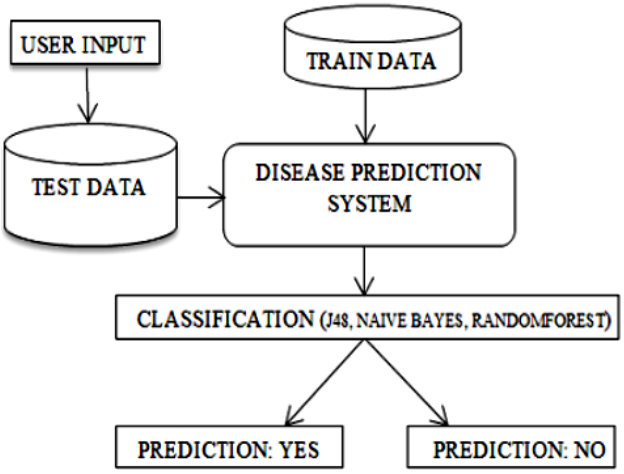

Figure 1: Basic architecture of prediction model

It consists of preparing dataset and user contribution as the test dataset. Classification techniques in data mining are capable of processing a large amount of data. It can be used to predict categorical class labels and classifies data based on training set and class labels and it can be used for classifying newly available data. The term could cover any context in which some decision or forecast is made on the basis of presently available information. Classification procedures recognized method for repeatedly making such decisions in new situations. Here assume that problem is a concern with the construction of a procedure that will be applied to a continuing sequence of cases in which each new case must be assigned to one of a set of pre defined classes on the basis of observed features of data.

Several learning algorithms target discovering better representations of the inputs gave during training.[11] Classic examples incorporate head components analysis and cluster analysis. Highlight learning algorithms, also called representation learning algorithms, often endeavor to preserve the data in their information yet additionally transform it such that makes it useful, often as a pre-processing step before performing classification or predictions. This procedure allows reconstruction of the inputs originating from the unknown information producing distribution, while not being necessarily devoted to configurations that are implausible under that distribution. This replaces manual element designing, and allows a machine to both get familiar with the features and use them to play out a specific task.

Feature learning can be either supervised or unsupervised. In supervised component learning, features are found out using named input information. Examples incorporate fake neural networks, multilayer perceptrons, and supervised lexicon learning. In unsupervised element learning, features are found out with unlabeled information. Examples incorporate word reference learning, autonomous segment analysis, autoencoders, network factorization and various forms of clustering.

Complex learning algorithms endeavor to do as such under the constraint that the scholarly representation is low-dimensional. Sparse coding algorithms endeavor to do as such under the constraint that the scholarly representation is sparse, implying that the scientific model has a huge number. Multilinear subspace learning algorithms intend to gain low-dimensional representations legitimately from

tensor representations for multidimensional information, without reshaping them into higher-dimensional vectors. Deep learning algorithms discover multiple levels of representation, or a chain of command of features, with more significant level, progressively abstract features characterized in terms of (or producing) lower-level features. It has been contended that an intelligent machine is one that learns a representation that disentangles the fundamental factors of variety that clarify the observed information.

2. Background

S. Mohan, et al. presents heart disease is one of the most significant causes of mortality on the planet today. Prediction of cardiovascular disease is a basic test in the zone of clinical information analysis. Machine learning (ML) has been shown to be successful in assisting in making decisions and predictions from the huge amount of information delivered by the healthcare industry. [1]

W. Chang, et al., presents the changes throughout individuals' life beat and improvement in material levels that occurred as of late increased the quantity of individuals suffering from hypertension on the planet. Along these lines, as a cardiac confusion of hypertension, the pervasiveness of hypertensive heart disease has increased every year, it has seriously imperiled the safety of human life, and the compelling prediction of hypertensive heart disease has become a worldwide issue. [2]

B. Wang et al., presents Renal dysfunction, which is associated with awful clinical outcomes, is one of the most well-known complications of heart failure (HF). Opportune prediction of renal dysfunction can enable clinical staffs to mediate right on time to evade catastrophic consequences. Right now, is proposed a multi-task deep and wide neural network (MT-DWNN) for foreseeing lethal complications during hospitalization.[3]

T. S. Brisimi, et al., presents urban living in present day huge cities has significant adverse effects on health, increasing the risk of several incessant diseases. it is focus on the two driving clusters of ceaseless diseases, heart disease and diabetes, and create information driven methods to anticipate hospitalizations because of these conditions. it is base these predictions on the patients' clinical history, later and increasingly distant, as described in their Electronic Health Records (EHRs).[4]

A. Mdhaffar, et al., presents this work presents a novel health analysis approach for heart failure prediction. It is based on the use of complex event processing (CEP) innovation, joined with statistical approaches. A CEP motor processes approaching health information by executing threshold-based analysis rules. Instead of having to physically set up thresholds, our novel statistical calculation naturally computes and updates thresholds as indicated by recorded historical information.[5]

J. Zhang et al., presents as of late, the use of intelligent technologies in clinical decision making in the telehealth condition has started to assume a crucial job in improving the nature of patients' lives and diminishing the costs and workload associated with their day by day healthcare. Right now, powerful clinical suggestion system that uses a fast Fourier transformation-coupled machine learning ensemble model is proposed for short-term disease risk prediction to furnish ceaseless heart disease patients with proper recommendations about the need to take a clinical test or not on the coming day based on examining their clinical information. The info sequence of sliding windows based on the patient's time series information are decomposed by using the fast Fourier transformation so as to separate the recurrence data. [6]

N. Alshurafa, et al., presents Remote health monitoring (RHM) systems are getting all the more widely received by clinicians and hospitals to remotely screen and speak with patients while streamlining clinician time, decreasing hospital costs, and improving nature of care.[7]

G. S. Karanasiou et al., presents Heart failure (HF) is a ceaseless disease described by low quality of life, intermittent hospitalization and high mortality. Adherence of patient to treatment suggested by the experts has been demonstrated a significant hindrance of the previously mentioned serious consequences. In any case, the non-adherence rates are significantly high; a reality that highlights the significance of anticipating the adherence of the patient and empowering experts to adjust in like manner understanding monitoring and the executives.[8]

D. Tay, et al., presents myocardial infarction (MI) is one of the main causes of death in many created countries. Thus, early discovery of MI events is basic for powerful preventative therapies, possibly diminishing avoidable mortality. One methodology for early disease prediction is the use of risk prediction models created using machine learning techniques. One significant segment of these models is to furnish clinicians with the adaptability to customize (e.g., the prediction range) and use the risk prediction model that they considered most gainful for their patients. [9]

S. Nikolaiev et al., presents the worldview change from deferred interventional to Prescient, Preventive and Personalized Medication is a main worldwide test in the 21st century. Ubiquitous union of portable applications, new intelligent sensors and machine learning methods make possible making of new age personalized programmed healthcare monitoring and pathologies location systems. These systems will assist with making platforms for progressively viable treatments custom fitted to the person, that is considered as the "medication of things to come".[10]

J. S. Sonawane et al., presents in clinical field the disease diagnosis is often made based on the knowledge and experience of the clinical specialist. Because of this there are chances of errors, undesirable biases and furthermore takes longer time in precise diagnosis of disease. In case of heart disease, its diagnosis is most troublesome task. It depends on the cautious analysis of various clinical and neurotic information of the patient by clinical experts, which is confused process. Because of progression in machine learning, PC and data innovation, the researchers and

clinical practitioners in huge degree are interested in the improvement of computerized system for the prediction of heart disease.[11]

D. R. Patil et al., presents in clinical field the diagnosis of heart disease is most troublesome task. It depends on the cautious analysis of various clinical and obsessive information of the patient by clinical experts, which is confounded process. Because of headway in machine learning and data innovation, the researchers and clinical practitioners in huge degree are interested in the improvement of mechanized system for the prediction of heart disease that is exceptionally exact, successful and supportive in early diagnosis. Right now is present a prediction system for heart disease using multilayer perceptron neural network. [12]

3. Proposed Methodology

The main contributions of this paper will be summarized as follows.

- To collect heart decease data set from from the UCI machine learning repository (i.e. Cleveland).
- To make hybrid approach based on random forest and Naive Bayes.
- To implement proposed method on spyder python 3.7 software.
- To prediction of various parameters.
- To generate results graph and compare from previous work.

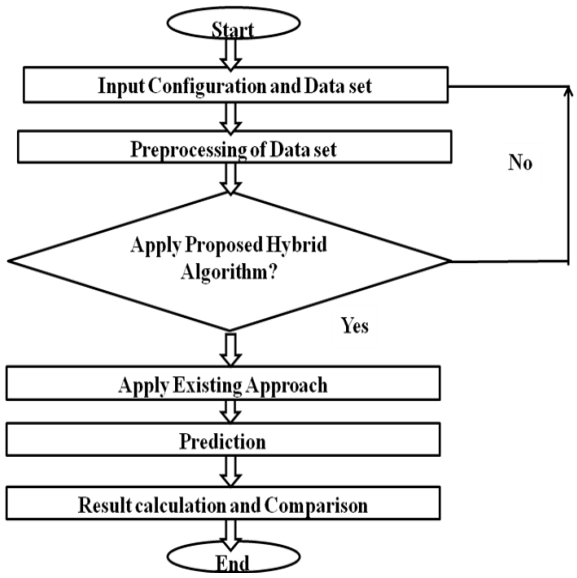

Figure 2: Flow Chart

Figure 2 is showing the proposed flow chart. The flow of work starts with to define input configuration and taken heart disease data set from the UCI machine learning repository. Now before apply machine learning techniques, firstly apply the steps for data preprocessing. The sample of data is taken in this step, it is also known as training data.

Now apply proposed hybrid approach based on the random forest and Naive Bayes. At last all training data is process and give predication of diseases. Now, Results graph generation and calculation of necessary parameters is done.

Step 1 − First, start with the selection of random samples from a given dataset.
Step 2 − Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.
Step 3 − In this step, voting will be performed for every predicted result.
Step 4 − At last, select the most voted prediction result as the final prediction result.

4. Simulation Result

The implementation of the proposed algorithm is done over python spyder 3.7. The sklearn, numpy, pandas, matplotlib, pyplot, seaborn, os library helps us to use the functions available in spyder environment for various methods like decision tree, random forest, naive bayes etc.

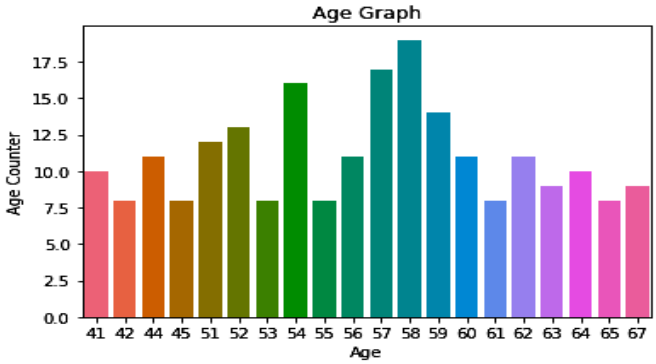

Figure 3: Graph of data feature- Age

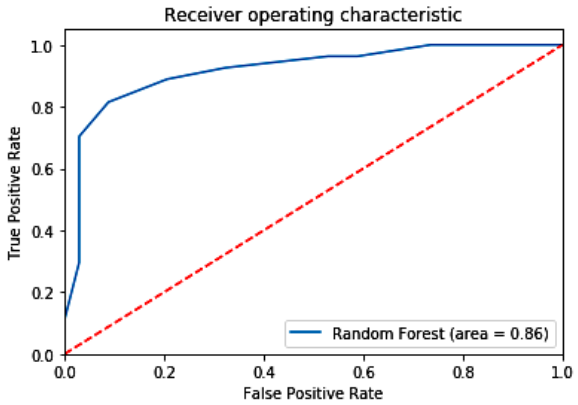

Figure 4: ROC result graph of Random forest

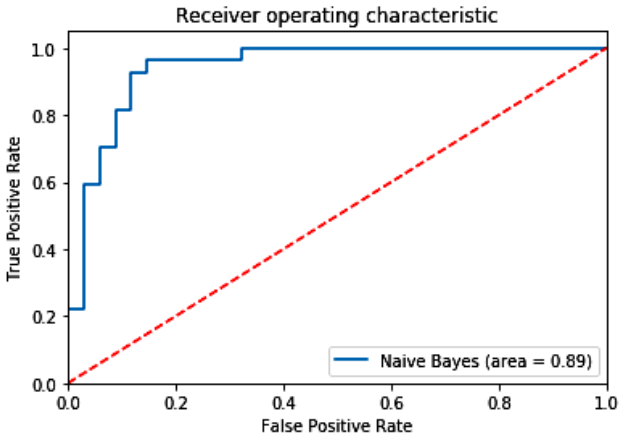

Figure 5: ROC result graph of Naïve Bayes

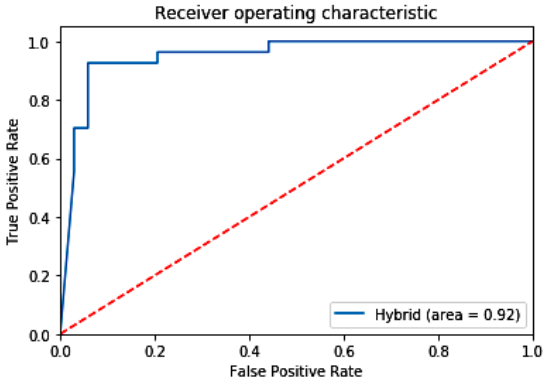

Figure 6: ROC result graph of proposed method (Hybrid)

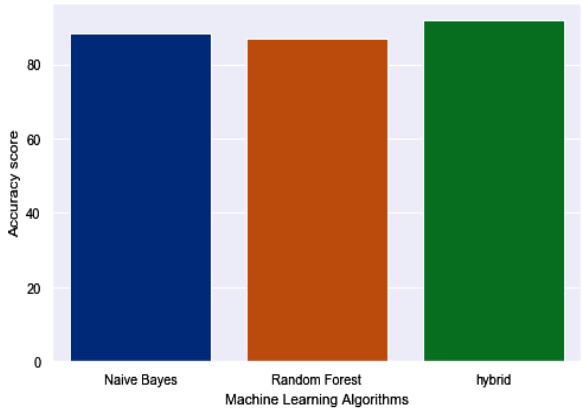

Figure 7: Result comparison of applied methods

From all results graph figures and receiver operating characterises curve, it is clear the proposed hybrid algorithm gives significant improved results than individual random forest and naïve bayes methods.

Table 1: Simulation Result of Random Forest method

Sr. No.	Parameters	Random Forest Method
1	Accuracy	86
2	Classification error	14
3	Precision	86
4	Recall	91
5	F-measure	89

Table 2: Simulation Result of Naive Bayes method

Sr. No.	Parameters	Naive Bayes
1	Accuracy	89
2	Classification error	11
3	Precision	91
4	Recall	88
5	F-measure	90

Table 3: Simulation Result of Proposed Hybrid method

Sr. No.	Parameters	Proposed Hybrid Method
1	Accuracy	92
2	Classification error	8
3	Precision	94
4	Recall	91
5	F-measure	93

Table 1, 2 and 3 are showing simulation results when applying random forest, naïve bayes and proposed hybrid algorithm.

Table 4: Comparison of Proposed work with previous work

S. No.	Parameters	Previous work	Proposed Work
1	Method	Hybrid random forest with a linear model	Hybrid Naive Bayes with random forest
2	Accuracy	88.4	92
3	Classification error	11.6	8
4	Precision	90.1	94
5	F-measure	90	93

Figure 4 shows that the comparison of proposed and previous work results. It is clear that proposed hybrid gives 92% accuracy while in previous there is 88.4% accuracy. The classification error is 8% in proposed while 11.6% in previous approach. The precision value is 94% and F-measure is 93% in proposed while previously it 90.1% and 90% respectively. Therefore it is clear from this comparison table, proposed hybrid approach gives better result than previous approaches.

5. Conclusion

Present day medication generates a lot of data stored in the clinical database. For instance, clinical information may contain MRIs, signals like ECG, clinical data like glucose, circulatory strain, cholesterol levels, and so forth. Therefore this paper focused to develop a Hybrid random forest and Naive Bayes (HRFNB) for predicting the threat of heart disease to a patient with the medical records got from the patients. HRFNB had the different attributes that it did not use already determined number of hidden units, but the hidden units got summed with one another till the error was decreased using proposed hybrid algorithm. The experimental results have proved that the proposed approach has achieved improvement in accuracy.

References

[1] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in *IEEE Access*, vol. 7, pp. 81542-81554, 2019.

[2] W. Chang, Y. Liu, X. Wu, Y. Xiao, S. Zhou and W. Cao, "A New Hybrid XGBSVM Model: Application for Hypertensive Heart Disease," in *IEEE Access*, vol. 7, pp. 175248-175258, 2019.

[3] B. Wang *et al.*, "A Multi-Task Neural Network Architecture for Renal Dysfunction Prediction in Heart Failure Patients With Electronic Health Records," in *IEEE Access*, vol. 7, pp. 178392-178400, 2019.

[4] T. S. Brisimi, T. Xu, T. Wang, W. Dai, W. G. Adams and I. C. Paschalidis, "Predicting Chronic Disease Hospitalizations from Electronic Health Records: An Interpretable Classification Approach," in *Proceedings of the IEEE*, vol. 106, no. 4, pp. 690-707, April 2018.

[5] Mdhaffar, I. Bouassida Rodriguez, K. Charfi, L. Abid and B. Freisleben, "CEP4HFP: Complex Event Processing for Heart Failure Prediction," in *IEEE Transactions on NanoBioscience*, vol. 16, no. 8, pp. 708-717, Dec. 2017.

[6] J. Zhang *et al.*, "Coupling a Fast Fourier Transformation With a Machine Learning Ensemble Model to Support Recommendations for Heart Disease

Patients in a Telehealth Environment," in *IEEE Access*, vol. 5, pp. 10674-10685, 2017.

[7] N. Alshurafa, C. Sideris, M. Pourhomayoun, H. Kalantarian, M. Sarrafzadeh and J. Eastwood, "Remote Health Monitoring Outcome Success Prediction Using Baseline and First Month Intervention Data," in *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 2, pp. 507-514, March 2017.

[8] G. S. Karanasiou *et al.*, "Predicting adherence of patients with HF through machine learning techniques," in *Healthcare Technology Letters*, vol. 3, no. 3, pp. 165-170, 9 2016.

[9] D. Tay, C. L. Poh, E. Van Reeth and R. I. Kitney, "The Effect of Sample Age and Prediction Resolution on Myocardial Infarction Risk Prediction," in *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 1178-1185, May 2015.

[10] S. Nikolaiev and Y. Timoshenko, "Reinvention of the cardiovascular diseases prevention and prediction due to ubiquitous convergence of mobile apps and machine learning," *2015 Information Technologies in Innovation Business Conference (ITIB)*, Kharkiv, 2015, pp. 23-26.

[11] J. S. Sonawane and D. R. Patil, "Prediction of heart disease using learning vector quantization algorithm," *2014 Conference on IT in Business, Industry and Government (CSIBIG)*, Indore, 2014, pp. 1-5.

[12] D. R. Patil and J. S. Sonawane, "Prediction of heart disease using multilayer perceptron neural network," *International Conference on Information Communication and Embedded Systems (ICICES2014)*, Chennai, 2014, pp. 1-6.

[13] R. Wijaya, A. S. Prihatmanto and Kuspriyanto, "Preliminary design of estimation heart disease by using machine learning ANN within one year," *2013 Joint International Conference on Rural Information & Communication Technology and Electric-Vehicle Technology (rICT & ICeV-T)*, Bandung, 2013, pp. 1-4.

[14] T. R. Tavares, A. L. I. Oliveira, G. G. Cabral, S. S. Mattos and R. Grigorio, "Preprocessing unbalanced data using weighted support vector machines for prediction of heart disease in children," *The 2013 International Joint Conference on Neural Networks (IJCNN)*, Dallas, TX, 2013, pp. 1-8.

[15] N. G. B. Amma, "Cardiovascular disease prediction system using genetic algorithm and neural network," *2012 International Conference on Computing, Communication and Applications*, Dindigul, Tamilnadu, 2012, pp. 1-5.

[16] K. R. Taylor *et al.*, "AudioGene: Computer-based prediction of genetic factors involved in non-syndromic hearing impairment," *2011 9th IEEE/ACS International Conference on Computer Systems and Applications (AICCSA)*, Sharm El-Sheikh, 2011, pp. 75-79.

[17] Thakkar, M. I. Hasan and M. A. Desai, "Health Care Decision Support System for Swine Flu Prediction Using Naïve Bayes Classifier," *2010 International Conference on Advances in Recent Technologies in Communication and Computing*, Kottayam, 2010, pp. 101-105.