Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 7, Issue 12, December -2020

Heart illness prediction using hybrid machine learning techniques

Dibyanshu Chatterjee[1], Niveditha CA[2]

[1]*Department of computer science and engineering, MS Ramaiah University of Applied Sciences, Bangalore*
[2]*Application Software Analyst, Accenture, Bangalore*

Abstract — *Diseases can influence individuals both genuinely and intellectually, as contracting and living with an illness can adjust the influenced individual's viewpoint on life. A sickness that influences the parts of a living being, which isn't in view of any quick outer injury. Diseases are regularly known to be ailments that are identified with explicit indications and signs. The deadliest sicknesses in people are arteria coronaries disease (blood stream impediment), trailed by cerebrovascular sickness and lower respiratory contaminations. Coronary diseases are most eccentric and unexpectable. We can become ready to foresee the coronary illness utilizing AI strategy. The datasets are taken from UCI store which is a public dataset. These prepared datasets are utilized for the expectation. Procedures like Decision tree, Support Vector Machine, K Nearest Neighbor and Random Forest algorithms are utilized in the expectation of coronary illness and cross breed of these algorithms gives 94 % precision.*

Keywords-*Decision tree, Support Vector Machine, K Nearest Neighbour and Random Forest*

I. INTRODUCTION

Hurtful deviation from the typical underlying or practical condition of a life form in the human body is named as disease. For the most part we can anticipate the illness in light of the manifestations. Medical care industry faces the significant issues like forecasting the diabetes illness among a few others. Individuals with high blood glucose and cholesterol with harm of veins will tend to build up a coronary illness and other nerve diseases. The European Society of Cardiology (ESC) overview that 26 million grown-ups overall were influenced by coronary illness, also, 3.6 million were analyzed each year. Machine learning strategies gives us the capacity to programmed learning and experience without being expressly modified. AI gives a goal assessment to improve effectiveness, dependability and exactness. AI strategies utilized for choice help accomplishes high precision of choices and they suggest a profound comprehension of choices and the decision makers will believe AI strategies. Strategies for learning verifiable, non-emblematic information will give better prescient exactness. Techniques for learning express, emblematic information produce heaps of understandable models. Highbred machine learning models team up fortifies information portrayal model sorts. In medical care industry and clinical stage, Collecting and dissecting the information is viewed as significant. With the machine learning idea, we can be ready to examine and foresee the information utilizing a few calculations and procedures. Directed Machine learning algorithms have been the most driving strategy in the information mining field. This investigation means to recognize the critical patterns among a few kinds of managed machine learning algorithms and their execution, use for disease hazard prediction. Overseeing diabetes includes a bunch of issues and responsibilities like routine checking of circulatory strain, glucose level and other wellbeing status. In this paper machine learning will foresee the coronary illness and non useful condition of the heart utilizing the vital clinical information esteem. In classification method, the total dataset is partitioned into 70% of information for preparing and 30%of information for testing. The expectation of coronary illness depends on machine learning algorithms like k-closest neighbor calculation (KNN), Decision tress algorithm, support vector algorithm (SVM), Random forest (RF) algorithm.

II. OPERATIONS

- **PROCESSING OF DATA:** Coronary disease information is pre-prepared by eliminating commotion and missing qualities. The datasets contain a sum of 310 patient records where 7 records are with some missing qualities, those 7 records have been taken out from the dataset and staying 303 patient records are utilized in preprocessing.

- **SELECTING FEATURE:** From the absolute 13 attributes of the dataset, two attributes relating to age and sexual orientation are utilized to recognize the individual data of the patient. The leftover 11 attributes are viewed as significant as they contain essential clinical records. Clinical information are crucial to determination and learning the seriousness of coronary illness. As beforehand referenced in this analysis, a few (ML) strategies are utilized specifically SVM, KNN, DT, RF Algorithm. The test was rehashed with all the ML methods utilizing each of the 13 attributes.

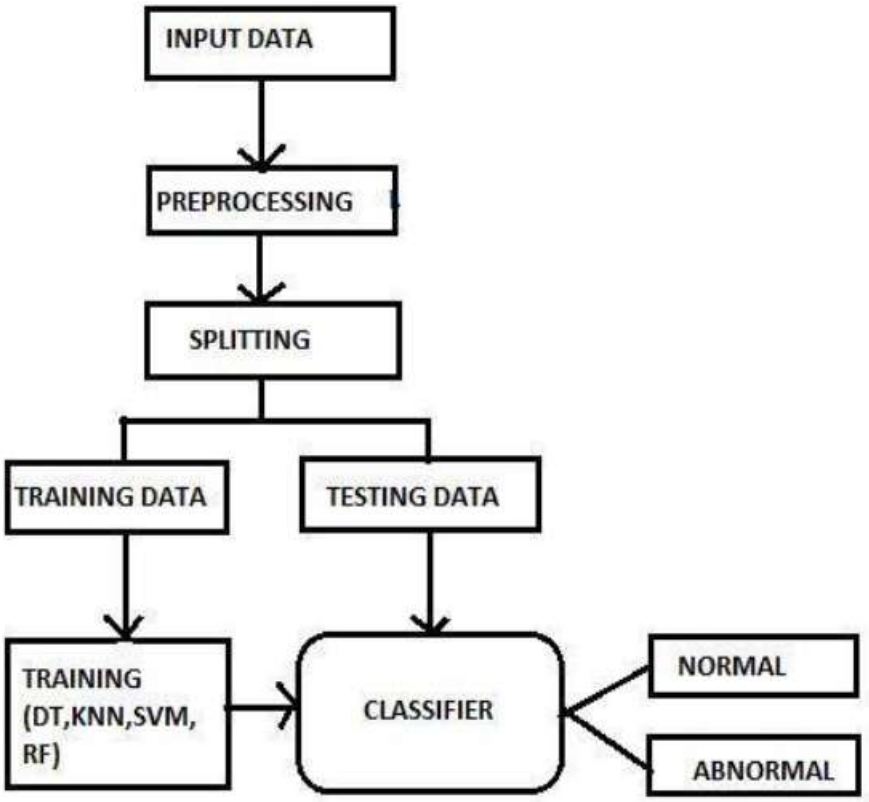

Fig 1. Workflow utilizing UCI dataset

- **CLASSIFICATION:** The grouping of the dataset is done based on the factors and models. At that point the different classifiers are applied to each dataset to gauge its presentation. The methods are, Decision tree, Support Vector Machine, K Nearest Neighbor, Random forest algorithms.

1. **DECISION TREE:** For preparing tests of information the trees are developed in view of data sources. Relapse and order issues are tackled utilizing choice tree. This strategy is performed on the premise of Top down separation and conquer approach. Tree pruning assists with eliminating immaterial examples of information.

2. **SUPPORT VECTOR MACHINE:** SVM is supposed to be an administered machine learning algorithm which might be utilized for grouping or relapse issues. It utilizes a way called the kernel trick to modify your information, at that point upheld these changes it finds an ideal limit between the potential yields.

3. **RANDOM FOREST:** A Random forest algorithm is one among the premier successful troupe arrangement approach. This algorithm has been utilized in forecast and likelihood. The Random Forest technique comprises of different choice trees. Every choice tree gives a data that shows the choice about the class of the article. RF technique mix sacking and irregular determination of highlights. There are three distinct parameters in RF, they are, number of the trees (n tree), Minimum node size and also, number of factors utilized in parting every hub.

4. **K NEAREST NEIGHBOUR:** K-Nearest neighbor (KNN) might be a basic, lethargic and nonparametric classifier. KNN is favored when all the highlights are direct. It is under regulated learning space and also, finds exceptional application in pattern detection, data mining and interruption recognition. KNN is also called as case-based thinking and has been used in numerous applications like pattern identification, statistical assessment. Order is acquired by distinguishing the nearest neighbor to work out the class of an obscure example. KNN is favored over other characterization calculations due to its high convergence speed and ease.

III. RESULTS

The prediction of illness is created utilizing 13 highlights what's more, 4 classifiers to improves the precision of the models. The most elevated exactness is accomplished by K-closest neighbor classification technique when contrasted with existing strategies.

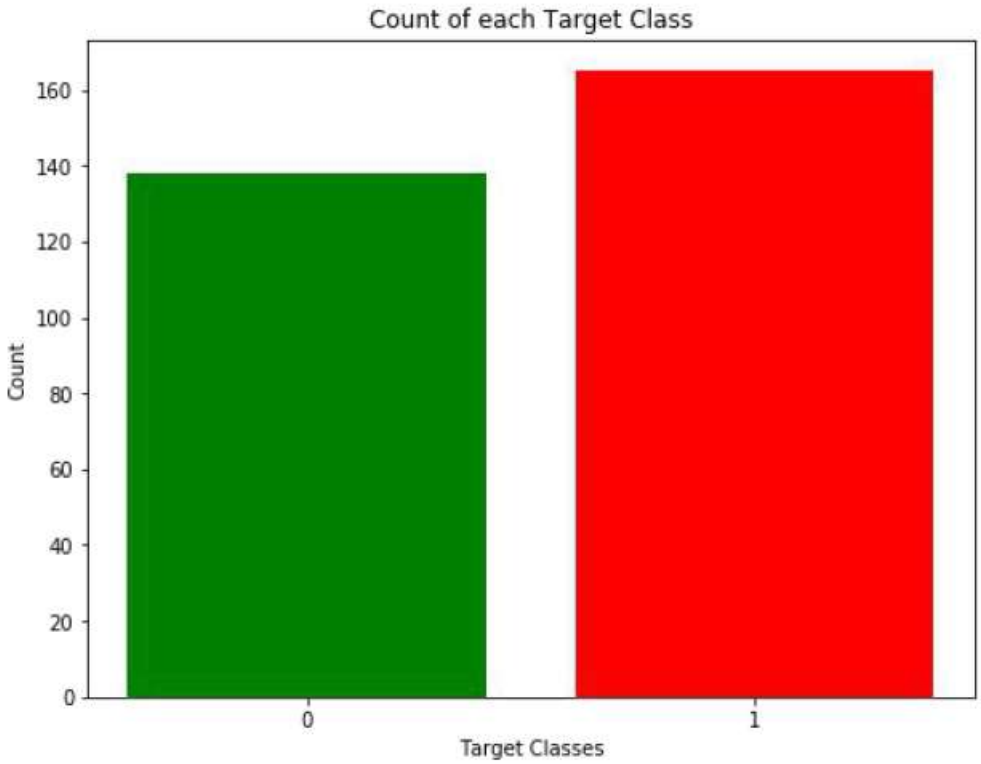

Fig 2.Target class vs Count

```
# Evaluating using accuracy_score metric
from sklearn.metrics import accuracy_score
accuracy_logreg = accuracy_score(Y_test, Y_pred_logreg)
accuracy_knn = accuracy_score(Y_test, Y_pred_knn)
accuracy_svc = accuracy_score(Y_test, Y_pred_svc)
accuracy_nb = accuracy_score(Y_test, Y_pred_nb)
accuracy_dectree = accuracy_score(Y_test, Y_pred_dectree)
accuracy_ranfor = accuracy_score(Y_test, Y_pred_ranfor)
```

```
# Accuracy on test set
print("Logistic Regression: " + str(accuracy_logreg * 100))
print("K Nearest neighbors: " + str(accuracy_knn * 120))
print("Support Vector Classifier: " + str(accuracy_svc * 100))
print("Naive Bayes: " + str(accuracy_nb * 100))
print("Decision tree: " + str(accuracy_dectree * 100))
print("Random Forest: " + str(accuracy_ranfor * 100))
```

```
Logistic Regression: 77.04918032786885
K Nearest neighbors: 88.52459016393443
Support Vector Classifier: 75.40983606557377
Naive Bayes: 68.85245901639344
Decision tree: 68.85245901639344
Random Forest: 73.77049180327869
```

Fig 3. Classifier's performance

IV. CONCLUSION

Distinguishing the preparing of crude medical services information heart data will help in the drawn out sparing of human lives and early expectation of the variations from the norm in heart conditions. Machine learning techniques were utilized in the cycle of crude information and to give the forecast of the sickness and wellbeing status of the patient. Coronary illness forecast is one of the difficult cycle in the clinical field. Utilizing this task, the death rate can be definitely controlled if the illness is recognized. The proposed hybrid approach is utilized to join the qualities of fuzzy

logic and k-closest neighbor algorithm which gives 94% exactness. This technique demonstrated the exactness of most noteworthy forecast rate. The further course of this examination can be performed with the combination of deep learning strategies to accomplish better forecast in exactness.

REFERENCES

[1] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, 81542-81554.
[2] Avinash Golande, Pavan Kumar T (June 2019) Heart disease prediction using effective machine learning techniques.
[3] David, H. B. F., & Belcy, S. A. (2018). Heart Disease Prediction Using Data Mining Techniques. ICTACT Journal on Soft Computing, 9(1).
[4] Yekkala, I., & Dixit, S. (2018). Prediction of Heart Disease Using Random Forest and Rough Set Based Feature Selection. International Journal of Big Data and Analytics in Healthcare (IJBDAH), 3(1), 1-12.
[5] Reddy Prasad,Pidaparthi Anjali, S.Adil, N.Deepa(Feb2019) Heart Disease Prediction using Logistic Regression Algorithm using Machine Learning
[6] Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. BMC Medical Informatics and Decision Making, 19(1), 1-16.
[7] 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.
[8] Latha, C. B. C., & Jeeva, S. C. (2019). Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked, 16, 100203.
[9] Jabbar, M. A., Deekshatulu, B. L., & Chandra, P. (2016). Intelligent heart disease prediction system using random forest and evolutionary approach. Journal of Network and Innovative Computing, 4(2016), 175-184.
[10] Jabbar, M. A. (2017). Prediction of heart disease using knearest neighbor and particle swarm optimization.
[11] Sharanyaa, S., and K. Sangeetha. "Blocking Adult Account in OSN's Using Iterative Social Based Classifier Algorithm."
[12] Sharanyaa, S., Abitha, P., Karthikeyan, B., Buvaneswari,B., & Sumithra, M. Identification of dysphonia related to Parkinson's disease using parametric and non parametric models.
[13] Buvaneswari, B., & Reddy, T. K. L. (2019). High performance hybrid cognitive framework for bio-facial signal fusion processing for the disease diagnosis. Measurement, 140, 89-99.

