Ameliorate Performance of Predicting Diseases Related to Heart by Utilizing ML Techniques

[1]**Manali Modi**,[2]**Neeshu Kumari**

[1,2]Department of Computer Science & Engineering, Parul Institute of Engineering & Technology, Parul University, Vadodara, India.

Abstract: Background: Disease Prediction framework dependent on prescient displaying predicts the infection of the client based on the indications that client gives as a contribution to the framework. The framework examines the indications gave by the client as info and gives the likelihood of the ailment as a yield. Ailment Prediction is finished by actualizing 5 systems, for example, Naïve Bayes, KNN, Decision Tree, Linear Regression and Random Forest Algorithms.

Aim: To study different ML algorithms to predict diseases related to heart through using considering various issues such as accuracy and precision.

Results: Random Forest algorithm outperforms from all the algorithms with accuracy around 88%.

Index Terms: **Machine learning, Algorithms, Diseases related to Heart.**

I. INTRODUCTION

In everyday life numerous variables that influence heart of human being. Numerous issues that happening at a rapidly and various ailments are quickly being distinguished for heart. We can say that the strength of heart of any person depends upon encounters in an individual's existence which can be totally subject to expert and individual practices of an individual. As per the WWO, consistently in excess of twelve million short-lived are happening because of different kinds of maladies that can be likewise familiar as term cardio-vascular infection. Heart ailment incorporates numerous ailments that are various and explicitly influence the heart and the corridors of an individual. Expansion for chance of coronary illness amongst youthful might because of the terrible dietary patterns, absence of rest, eager nature, wretchedness and various different factors, for example, corpulence, horrible eating routine, family ancestry, hypertension, high BP, inactive conduct, ancestry. Manifestations for coronary illness significantly rely on which of the distress felt by an person. A few manifestations are not normally recognized by the everyday citizens. In any case, basic side effects incorporate chest agony, shortness of breath, and heart palpitations. happens at the point when a piece of the heart doesn't get enough oxygen. Angina might be activated by upsetting occasions or physical effort and typically keeps going under 10 minutes. Cardiovascular failures can likewise happen because of various kinds of coronary illness. The indications of a respiratory failure resemble angina aside from that they can happen during rest and will in general be progressively serious. The side effects of a coronary episode can in some cases look like heartburn. Indigestion and a stomach throb can happen, just as an overwhelming feeling in the chest.

Cardiovascular breakdown can be likewise a result for coronary illness, shortness of breath can happen when the heart turns out to be too feeble to even think about circulating blood. Some heart conditions happen without any side effects by any means, particularly in more seasoned grown-ups and people with sugar.

enable the patients to play it safe and take administrative measures. As of late, the social insurance industry has been producing immense measures of information about patients and their ailment analysis reports are by and large particularly taken for the expectation of heart assaults around the world. At the point when the information about coronary illness is tremendous, the AI methods can be executed for the investigation.

II. TYPE STYLE AND FONTS

The fundamental inspiration for carrying out proposed research is that for exhibiting a coronary illness expectation form to forecast the event for coronary illness. Moreover, the exploration effort might be pointed for distinguishing finest characterization calculation to recognize probability for coronary illness. Proposed research can be advocated to configure a relative report of investigation utilizing three grouping calculations specifically NB, KNN, Logistic Regression so on are utilized for various degrees. In spite of the fact that these are generally utilized AI calculations, the coronary illness forecast is a crucial errand including most noteworthy conceivable precision. Subsequently, the three calculations are assessed at various levels and kinds of assessment procedures. This will give scientists and clinical specialists to set up a better understanding and assist them with recognizing an answer for distinguish the best technique for anticipating the heart diseases[5][6].

A key test going up against human services association (clinics, clinical focuses) is the office of value administrations at sensible costs. The heart is the basic bit that can be consider omnipotent.

These can be few components that assembles hazard of diseases related to heart [9].

Some of them are recorded beneath:

• The family ancestry of coronary illness

• Smoking

• Cholesterol

• High circulatory strain

Obesity Coronary illness Prediction utilizing Machine Learning Algorithms..As a result of a wide openness of superlative proportion of data and a need to change over this open enormous proportion of data to accommodating information requires the use of data mining procedures. Data Mining and KDD (learning divulgence in the database) have ended up being conspicuous as of late [10][11]. The ubiquity of data mining and KDD (data disclosure in database) shouldn't be a wonder since the proportion of the data builds that are open are amazingly broad to be examined genuinely and even the methods for modified data examination in

perspective on built up bits of knowledge and machine adjusting much of the time undermine issues while getting ready huge, powerful data builds including complex things [12].
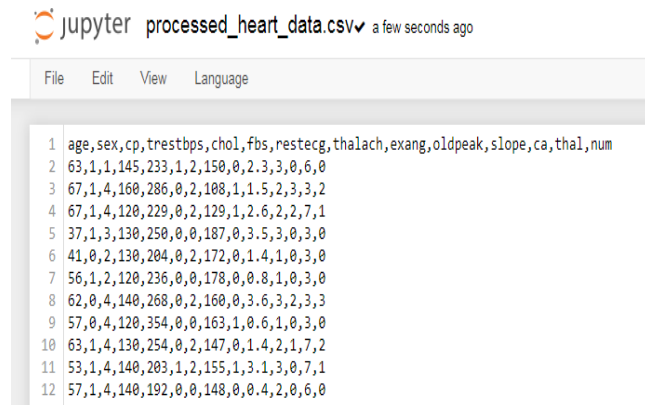
III. LITERATURE REVIEW

Ahmet Ilhan et.al discovered a portion of strategies of examining diseases related to heart by utilizing Recurrent Fuzzy Neural Network, inherited calculation that can be helped to clinical specialists that provids food a few parameters that may affect the dynamic procedure. A aggregate of 297 examples of patient information are considered, in the midst of which forty five were appointed to perform examine and two fifty two were utilized to prepare test report. This examination achieves a precision of around 98%. This work of coronary illness on test database, examination can be conveyed effectively. Subsequent components can be determined: precision, Root-Mean-Square Error, likelihood of false classification error, particularity, affectability, accuracy and F score [6]. A. Lamgunde et. al., prescribes the hereditary calculation utilizing reverse proliferation method way for dealing with foresee coronary illness adequately. The examination uses various input highlights for looking at the framework for coronary illness forecast. In general 13 clinical qualities are used through the framework such as sex, blood pressure, and so forth… to anticipate odds from victim toning coronary illness [8]. T..R et.al., proposed study identified with different grouping strategies that can anticipate chance elements identified with each individual considering the components, for example, sex, age, BP, cholesterol, beat rate. By methods for different information mining arrangement systems like NB, KNN, and so on., victim hazard stage might arranged. As great deal for traits can be in use, more exactness can be accomplished[8]. Indhu et.al., prescribes the NB calculation to foreseeing danger of coronary illness in victims. To prepare dataset, all the information is considered. This order stage includes Cleansing of data whereas accompanying assignments are executed: noise removing, standardization, information decreases and so on. The expectation stage includes order and expectation of illness types. Thus preparing set incorporates infection type and the testing set is manufactured utilizing the inquiries. The yield created is sent to the specialist/authority [9]. S. Dangare.et.al., presents the primary three layers: input, covered up and the yield layer. The info is taken care of to the information layer also, the yield layer extends the outcome procured. From there on, correlation of both real and the normal yield is performed. Utilizing the back engendering, blunder can be decided and weight in the midst of the yield and earlier covered up layers can be balanced. After fruition of back spread, forward procedure begins and is conveyed forward till the blunder is decreased [10]. K. Pramanik.et.al., suggests a Hybrid Algorithm that being a mix of ID3 and KNN calculation and are embraced for anticipating coronary illness. The information is pre-handled utilizing the KNN calculation consequently it's additionally alluded to as pre-handled calculation.

IV. RESEARCH PROPOSED METHODOLOGY

In this research paper, online Cleveland disease prediction data set for heart has been used from UCI Repository of ML. On these dataset in proposed research five different algorithms Decision Tree, Logical Regression(LR), KNN, and Random forest(RF) applied to find the maximum accuracy and compare with each other for better results. Figure I show proposed methodology.

Fig I: Proposed Algorithm

Fig II: Cleveland Dataset for Heart Disease

For this research, python language has been used to implement proposed algorithm. In Anaconda Jupiter IDE, Cleveland dataset has been imported using numpy and pandas library of python.

```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns

        %matplotlib inline

        import os
        print(os.listdir())

        import warnings
        warnings.filterwarnings('ignore')

        ['.config', 'heart.csv', 'sample_data']
```

Fig III: Importing dataset in Jupiter

For the successful examination, below is the range of different parameters to predict heart diseases.

```
age:         age
sex:         1: male, 0: female
cp:          chest pain type, 1: typical angina, 2: atypical angina, 3: non-anginal pain.
trestbps:    resting blood pressure
chol:        serum cholestoral in mg/dl
fbs:         fasting blood sugar > 120 mg/dl
restecg:     resting electrocardiographic results (values 0,1,2)
thalach:     maximum heart rate achieved
exang:       exercise induced angina
oldpeak:     oldpeak = ST depression induced by exercise relative to rest
slope:       the slope of the peak exercise ST segment
ca:          number of major vessels (0-3) colored by flourosopy
thal:        thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
```

Fig IV: Range of parameters

After applying five different algorithms results are as per below figures.

Logistic Regression

```
from sklearn.linear_model import LogisticRegression
logreg = LogisticRegression()

logreg.fit(X_train, Y_train)

y_pred_lr = logreg.predict(X_test)
print(y_pred_lr)
```

accuracy score : 85.25 %

Fig V: Logistic Regression applied on dataset

Naive Bayes

```
#Gaussian Naive Bayes
from sklearn.naive_bayes import GaussianNB
model = train_model(X_train, Y_train, X_test, Y_test, GaussianNB)
```

```
Train accuracy: 83.47%
Test accuracy: 85.25%
```

Fig VI: Naïve Bayes applied on dataset

K-Nearest Neighbor

```
from sklearn.neighbors import KNeighborsClassifier
model = train_model(X_train, Y_train, X_test, Y_test, KNeighborsClassifier)
```

```
Train accuracy: 78.10%
Test accuracy: 63.93%
```

Fig VII: KNN applied on dataset

Random Forest

```
#Random forest with 100 trees
from sklearn.ensemble import RandomForestClassifier
rf = RandomForestClassifier(n_estimators=100, random_state=0)
rf.fit(X_train, Y_train)
print("Accuracy on training set: {:.3f}".format(rf.score(X_train, Y_train)))
print("Accuracy on test set: {:.3f}".format(rf.score(X_test, Y_test)))
```

```
Accuracy on training set: 1.000
Accuracy on test set: 0.885
```

Now, let us prune the depth of trees and check the accuracy.

```
rf1 = RandomForestClassifier(max_depth=3, n_estimators=100, random_state=0)
rf1.fit(X_train, Y_train)
print("Accuracy on training set: {:.3f}".format(rf1.score(X_train, Y_train)))
print("Accuracy on test set: {:.3f}".format(rf1.score(X_test, Y_test)))
```

```
Accuracy on training set: 0.876
Accuracy on test set: 0.869
```

Fig VIII: Random Forest applied on dataset

Decision Tree

```
from sklearn.tree import DecisionTreeClassifier
tree1 = DecisionTreeClassifier(random_state=0)
tree1.fit(X_train, Y_train)
print("Accuracy on training set: {:.3f}".format(tree1.score(X_train, Y_train)))
print("Accuracy on test set: {:.3f}".format(tree1.score(X_test, Y_test)))
```

```
Accuracy on training set: 1.000
Accuracy on test set: 0.787
```

Fig IX: Decision Tree applied on dataset

V. RESULTS

This paper is introduced utilizing various machine learning algorithms. From Random forest, logistic regression, KNN, Naive Baye's Algorithm, and Decision Tree are utilized to construct framework. Proposed algorithm portrays, Random Forest demonstrates the better outcomes and helps the specialists in this area and even the individual identified with the clinical field to get ready for a superior and early determination for the patient.
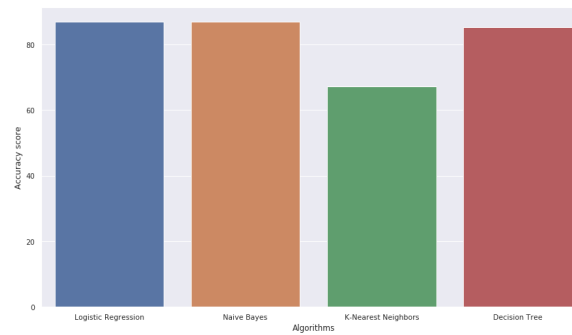
.

Fig X: Accuracy Comparison

REFERENCES

1. M. K. A. and F. Siraj, "Utilization of an artificial neural network in the prediction of heart disease," Int. J. Bio-Science Bio-Technology, vol. 5, no. 4, pp. 159–165, 2013.
2. P. Selvakumar and S. P. Rajagopalan, "Survey on neural network models for heart disease prediction," J. Theor. Appl. Inf. Technol., vol. 67, no. 2, pp. 485–497, 2014.
3. A. Methaila, P. Kansal, H. Arya, and P. Kumar, "Early Heart Disease Prediction Using Data Mining Techniques," vol. 6956, no. October, pp. 53–59, 2014.
4. T. Karayilan and Ö. Kiliç, "Prediction of Heart disease using neural network," in 2nd International Conference on Computer Science and Engineering, UBMK 2017, 2017, pp. 719–723.
5. M. Mardiyono, R. Suryanita, and A. Adnan, "Intelligent Monitoring System on Prediction of Building Damage Index using Neural-Network," TELKOMNIKA vol. 10, no. 1, p. 155, 2015.
6. N. Guru, A. Dahiya, and N. Rajpal, "Decision support system for heart disease diagnosis using Neural Network," 2007.
7. J. S. Singh Navdeep, "No Title," Intell. Hear. Dis. Predict. Syst. using CANFIS Genet. Algorithm, Biological Medical Science., no. International Journal of Advance Research, Ideas and Innovations in Technology 987, 2008.
8. R. G. S. Rajkumar Asha, "No Title," Diagnosis Hear. Dis. Using Datamining Algorithm, vol. Issue 10 V, no. Global Journal of Computer Science and Technology, p. Page 38, 2010.
9. S. Radhimeenakshi and G. M. Nasira, "Prediction of Heart Disease using Neural Network with Back Propagation," Int. J. Data Min. Tech. Appl., vol. 4, no. 1, pp. 19– 22, 2016.
10. "Prediction of Heart Disease using Artificial Neural Network," VFAST Trans. Softw. Eng., pp. 102–112, 2019.
11. Z. M. A. Mehta Rupa G. ,Rana Dipti P., "No Title," A Nov. Fuzzy Based Classif. Data Min. using Fuzzy Discret., no. World Congress on Computer Science and Information, 2009.
12. A. A. Shinde, S. N. Kale, R. M. Samant, A. S. Naik, and S. A. Ghorpade, "Heart Disease Prediction System using Multilayered Feed Forward Neural Network and Back Propagation Neural Network," Int. J. Comput. Appl., vol. 166, no. 7, pp. 975– 8887, 2017.
13. N. Al-Milli, "Backpropogation neural network for prediction of heart disease," J. Theor. Appl. Inf. Technol., vol. 56, no. 1, pp. 131–135, 2013.
14. G. S. Kumari Milan, "No Title," Comp. Study Data Min. Classif. Methods Cardiovasc. Dis. Predict., vol. Vol. 2, no. IJCST Vol. 2, Iss ue2, June 2011, 2011.
15. B. L. and K. H. R. Kiyong Noh, HeonGyu Lee, Ho-Sun Shon, "Associative Classification Approach for Diagnosing Cardiovascular Disease," Intell Computation Signal Process. Pattern Recognition, vol. 345, pp. 721–727, 2006.