International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 06 | June 2021 www.irjet.net p-ISSN: 2395-0072

Predictive Analysis and Detection of Heart Disease Using Machine Learning Approach

Bhavani S[1], K Raghuveer[2]

[1]*PG Student, Dept. of Information Science and Engineering, NIE, Mysore, India*
[2]*Professor, Dept. of Information Science and Engineering, NIE, Mysore, India*

Abstract – *cardiovascular diseases has become one of the biggest killers in the modern society. There in fields of medical data collection, predicting heart events is a major issue. Algorithms (ML) has really been demonstrated to be useful in supporting with the decision-making and analysis of huge amounts of data generated by the medical field. Throughout this research, we created an effective methodology for analysis and the results as well as Cardiac data in order to training a Machine Learning Algorithm to effectively treat the cardiac as well as predict future performance if any exist. Classification algorithms while using machine learning make use of huge quantities of training examples from 'cardiovascular disease patients' databases to discover comparisons. Machine learning algorithms employ huge quantities of training examples from 'cardiovascular disease patients' datasets to discover similarities among several factors in order to learn how to filter input messages from a healthcare worker in order to monitor the client's heart problem. In this research, we present a unique way for analyzing various characteristics using machine learning approach, which will improve the quality of heart disease predictions. Various types of characteristics and many well-known classifiers are used to build the predictive models. Along through prediction for cardiovascular diseases, we achieve a higher level of quality with a high degree of accuracy. This Disease Prediction Using Machine Learning is completely done with the help of Machine Learning and Python Programming language with Interface for it and also using the dataset that is available previously by the hospitals using that we will predict the heart disease.*

***Key Words*: machine learning, heart disease, prediction analysis.**

1.INTRODUCTION

Computer Vision Healthcare Detection seems to be a software that diagnoses diseases given the data given by the user. This even forecasts a hospital staff as well as person's condition given the facts as well as signs inputted, which delivers high accuracy based on knowledge. If somehow the individual that is not in any danger and even the consumer only asks what kind of sickness he or she has had. It really is a device that offers the individual advice and techniques on how to keep their nutrition remain in good state, as well as a technique to check exactly whether they have a problem utilizing such predictions. Since the medical field plays a key role in treating clients' illnesses, this really is a wonderful method therefore for medical field to educate individuals, and that it is beneficial to customers because user does not even go to the medical facility or really any surgery center, therefore by manually opening the health problems and perhaps other relevant information, the visitor could really know and understand about the illness from where user is undergoing as well as how to treat something. Also, several companies did such DPUML here together, although our motive is to achieve it unique as well as helpful to something like the clients of this systems. Such Health Detection Applying Algorithms was completed entirely with both the aid of Computer Vision and the Python with Tkinter Interface, as well as the datasets that was originally made accessible by such hospitals. Clinicians are using a series of laboratory methods and tools for not just identifying and evaluating minor illnesses, but also several severe disorders. This same proper and precise assessment is often credited with an early diagnosis. Specialists could occasionally make mistaken judgments during diagnosis someone medical symptoms; as both a solution, diseases supervised learning that employ algorithms methods aid in achieving reliable findings under such situations. As per investigation, 40% of people accept summary illness, that mostly contributes to harmful disease eventually, and the developer cardiovascular diseases evaluation utilizing algorithms was created to solve terms of attitude infection through previous phases. Since we all know, with in competitive atmosphere of economic and social development, human race has become so connected which user is not worried about security. One major cause of misunderstanding is a lack of willingness should seek medical help and a lack of money. Individuals became so preoccupied with their daily lives how they have enough ways to plan a consultation as well as seek medical advice, resulting in a dangerous condition.

1.1. Motivation

The main purpose of this application is to look at the feature selection methods, data preparation, including sample preparation used within the training models. The

application emerges as an important in the accuracy during training, verification, and practical validating. As little more than a result, this application had been carried out with the aim of learning little about the models and executing the techniques.

The identification of cardiac disease is a severe difficulty. Generally, devices that would identify heart disease, because they're either too expensive or unproductive on calculating the cholesterol levels in human. This same risk of death that chronic implications of ventricular conditions can all be diminished even when they are found early. Unfortunately, this is better to watch such individuals on a large scale. Because that need significant competence, time, and technique to successfully treat situations plus discuss with something like a physician for 24hrs, it really isn't possible. We sometimes use different classifiers so examine the data finding unstructured data considering they have had so much data in modern day society. Using medicine data, the hidden patterns might be mistreated for care and emergency.

1.2. Objectives

An aim of this study seeks to see if a complete current parameter, also including personality, weight, excessive sweating, overnight blood pressure, and so on, predict that if they are common in patients of cardiac heart attacks. Another collection with both the doctor as well as parameters is retrieved from those in the Extracted features. One should estimate whereas if person could undergo a heart attack using the same knowledge. Regardless as well as not, they include a biostability. To provide it, researchers categories individual business owners on 14 medical indicators if user really at risk for heart disease. Different techniques are now used to train these medical personality traits: artificial neural network (Ann, KNN, and Random Forest Classifier). Classifier is really the most accurate of these strategies, with either an accuracy of 88.52 basis points. Ultimately, we characterize persons of developing a disease If you've a biostability or not, this process is extremely low cost.

2. RELATED WORK

As such M. Sanz et al. [1], breathing disability is one of the most dangerous injuries and loss safety. The medicine of heart problems has been mentioned in a study that has acquired a wide acceptance in the healthcare professionals throughout the world. Cardiac chronic conditions are one of the major public health problems around the world. Only in North America, 17.7 people die every week as a result chronic cardiovascular disease. Which according to Health Organization (WHO), disability accounted for roughly 31% of all annual deaths in 2016.

[2] Who the variety of heart and lung cases, and that's the focus of this study, shows that 82 percent of these people are from middle class and lower countries, 17 million people are under 35 years of age and likely to lead to non-infectious diseases, 6.7 million people are affected by stroke, and 7.4 million people include heart disease (WHO, 2016). S. N. Blair but mostly workmates [3], The attributes that need to be studied in order to study the cardiovascular diseases mischance are those related to attitudes. Furthermore, patients will have been applied to a series of experiments, which include blood pressure, glucose, vital signs, neck stiffness, medications, maximal heart rate, and moderate sugar levels, so there is some good news. It may be possible that industry expertise is possible if the requirement is realized and assumed early, but treatment for all of these cardiac patients is based on research studies, the patient's physical examination, and the patient's responses to questions and complaints [4,5] Abdel et al. Individually understanding the chances of having cardiac disease is dependent on as a risk factor. A number of different of knowledge discovery techniques also recently being proposed to tackle technical challenges.

Machine learning, however according Aljanabi et al [6], will serve us recognize patterns and their useful knowledge. While computer vision has a lot of functions in drug, it is commonly used to assess cardiovascular problems. Multiple organizations are interested through using deep learning techniques to detect conditions since it minimizes diagnostic time and the cost. Expresses selectivity and productivity Several diseases can be spotted using machine learning techniques, but the fundamental purpose of the article would to identify heart disease since it's the biggest cause of mortality today and a successful heart attacks diagnosis is greatly beneficial in saving lives.

[7] Obasi et al. Throughout prediction and diagnosis, machine learning (ML) plays a significant role. It incorporates an effective educational data to classify whether such a patient has a specified diseases type [5–8].

To classification cardiovascular dysfunction in to the acceptable or unhealthy, N. S. C. Reddy et al. [9] adopted RF, SVM, NB, NN, and K - NN with many features selection such as correlation matrix, recursive feature elimination (RFE), and learning vector quantization (LVQ) model. The debug findings reveal that RF obtained better achievement.

To predict cardiac disorders, R. Atallah et al. [10] used stochastic gradient descent (SGD), KNN, RF, logistic regression (LR), and voting ensemble learning. The best accuracy of 90 percent was realized first by -e voting ensemble learning model.

To forecast cardiovascular disorders, R. Kannan et al. [11] deployed a training algorithm (RNN), a genetic algorithm, and K-mean. The maximum performance was produced with RNN, although the lowest accuracy has been achieved using K-mean.

To estimates cardio disorders, K. Raza et al. [12] created three machine learning algorithms: LR, RF, SVM, and stochastic gradient boosting (SGB). Based on the current model, LR has the better performance of 86.5 degree Celsius.

When diagnose cardiovascular illnesses, A. N. Oo et al. [13] can use an ensemble learning model, multilayer perceptron, LR, and NB. The finding suggests that ensemble learning exceeds competing methods. the results of cardiac disorder prediction.

When diagnose cardiovascular disorders, S. Nalluri et al. [14] have used ensemble learning model, multilayer perceptron, LR, and NB. The data indicate that ensemble learning model outperformed conventional system in terms of cardiac problem prediction.

To anticipate cardiovascular disease, R. Bhat et al. [15] utilized random feature selecting (CFS) with sequential minimal optimization (SMO). The CFS-SMO algorithm, with either an accuracy of 86.96 percent, had been shown to be the maximum precise.

L. Sapra et al. [16] adopted approaches to improving cardiovascular diseases forecasting (XG Boost and LR). - LR, with an accuracy of 85.68 percent, topped XG Boost, who had an accuracy of 84.46 percent, according with results.

Computer vision is a methodology for applying innovative methodologies that contribute themselves to predicting also in context of information systems; in commercial application, this would be known as computer vision. With learning from past connections and patterns in the data, those statistical methods help researchers, business analysts, technicians, and marketers to "create dependable, predictable conclusions and solutions" and identify "hidden knowledge."

Projects for data analysis Normally, machine learning tasks are broken down into various segments:

Supervised classification: The "master" presents the system with problem description and target output, with both the objective of trying to learn a guideline that communication infrastructure to output values. In certain situations, this input data may be provided in portions or be limited to specific responses.

Semi-supervised: individuals learn when the system was given just an imperfect instruction signals, such as a training phase that lacks certain (typically many) of the desired outputs.

Learning environment: This system only can collect implement in order for a fixed couple of iterations (because to a limitation), and this must simultaneously optimize those entities for which it receives categories. Those could be made available for identification when utilized dynamically.

Unsupervised learning: is when the training data is not provided tags and is left to identify pattern in the data in itself. Unsupervised learning can become a goal by itself (finding hidden patterns and relationships) or a matter of life and death (finding hidden patterns in data) (feature learning).

reinforcement learning: Information (in the form of recognition and penalties) are still only supplied as reaction to the system's activities in a dynamic world, like as riding a motorcycle or playing the game, in reinforcement learning.

3. EXISTING SYSTEM

Prediction using traditional methods and models involves various risk factors and it consists of various measures of algorithms such as datasets, programs and much more to add on. High-risk and Low-risk patient classification is done on the basis of the tests that are done in group. But these models are only valuable in clinical situations and not in big industry sector. So, to include the disease predictions in various health related industries, we have used the concepts of machine learning and supervised learning methods to build the predictions system.

After doing the research and comparison of all the algorithms and theorems of machine learning we have come to conclusion that all those algorithms such as Decision Tree, KNN, Naïve Bayes, Regression and Random Forest Algorithm all are important in building a disease prediction system which predicts the disease of the patients from which user is suffering from and to do this we have used some performance measures like ROC, KAPPA Statistics, RMSE, MEA and various other tools. After using various techniques such as neural networks to make predictions of the diseases and after doing that we come to conclusion that it can predicts up to 90% accuracy rate after doing the experimentation and verifying the results. The information of patient statistics, results, disease history in recorded in EHR, which enables to identify the potential data centric solution, which reduces the cost of medical case studies. Existing system can predict the disease but not the sub type of the disease and it fails to predict the condition of the people, the

predictions of disease have been indefinite and non-specific.

Drawbacks:

1. The effects of heart events are hard to forecast.
2. Information systems would not contribute towards creating profits.
3. Statistical methods for medical information are too heavy.

4. PROPOSED SYSTEM

This suggested method of illness predictions utilizing ml algorithms is that we've had utilized a variety of methods, algorithms, as well as other tools to implement a situation which its predictions a health diagnosis based on their signs, which we compare to the software's datasets that has been before published. Through correlating those databases to the particular diagnosis, we can approximate the client's disorder %. The dataset and symptoms go to the prediction model of the system where the data is pre-processed for the future references and then the feature selection is done by the user, where they will enter the various symptoms. Then the classification of those data is done with the help of various algorithms and techniques such as Decision Tree, KNN, Random Forest and etc. Then the data goes in the recommendation model, there it shows the risk analysis that is involved in the system and it also provides the probability estimation of the system such that it shows the various probability like how the system behaves when there are n number of predictions are done and it also does the recommendations for the patients from their final result and also from their symptoms like it can show what to use and what not to use from the given datasets and the final results. Here we have combined the overall structure and unstructured form of data for the overall risk analysis that is required for doing the prediction of the disease. Using the structured analysis, we can identify the chronic types of disease in a particular region and particular community. In unstructured analysis we select the features automatically with the help of algorithms and techniques. This system takes symptoms from the user and predicts the disease accordingly based on the symptoms that it takes and also from the previous datasets, it also helps in continuous evaluation of viral diseases, heart rate, blood pressure, sugar level and much more which is in the system and along with other external symptoms it predicts the appropriate and accurate disease.

Benefits:

1. Decrease the amount of work it required before receive these results.
2. There are still no proof about cheating.
3. Lower rates.

5. SYSTEM REQUIREMENTS

5.1 Hardware Requirements

System : Pentium4, Intel Core i3, i5, i7 and 2 GHz Minimum
RAM : 512 Mb or above
Hard Disk: 10GB or above
Input Device: Keyboard and Mouse
Output Device: Monitor or PC

5.2 Software Requirements

Operating System: Windows7,10 or Higher Versions
Platform : Jupyter Notebook
Front End : HTML, CSS
Back End: SQLite3 and Python
Programming Lang: Python

6. ARCHITECTURE DAIGRAM

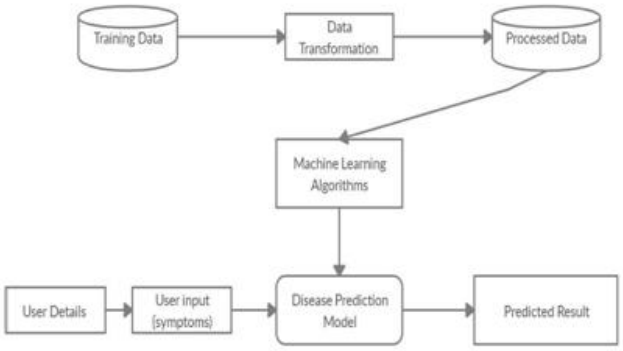

Fig -1: System Architecture

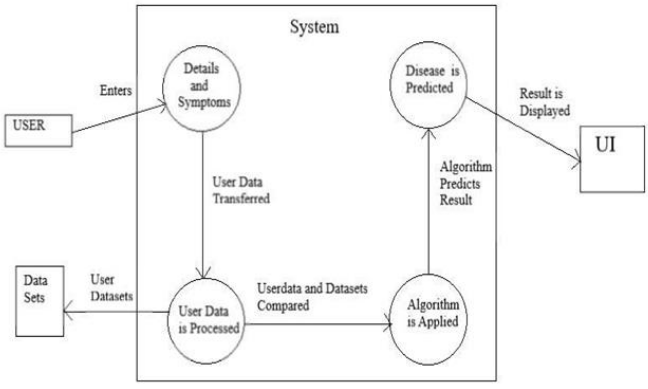

Fig -2: Data Flow Diagram

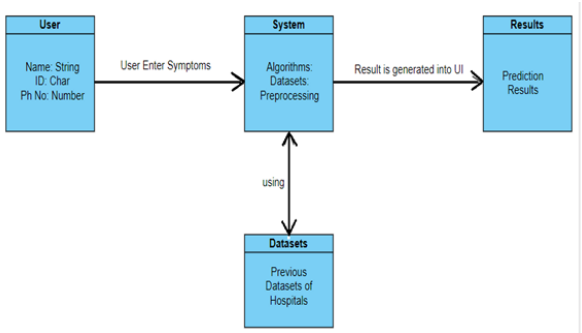

Fig -3: Class Diagram

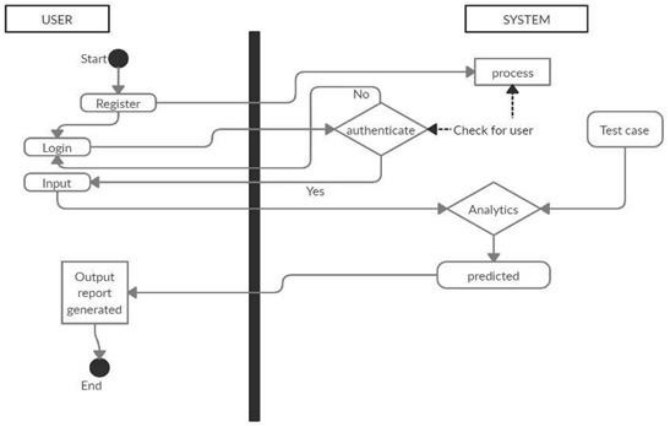

Fig -4: Activity Diagram

7. METHODOLOGY:

We'll start by exploring that collection with Panda's package, and we'll evaluate information. Training and experimentation on datasets. The heart Disease Prediction model will be trained on the dataset of diseases to do the prediction accurately and produce Confusion matrix. In this project different algorithms were used – Decision tree (DT), K-nearest neighbor (KNN), Random Forest (RF), Logistic regression (LR), Support vector machine (SVM), voting classifiers hard and soft (DT, LR, SVM). We will represent that data samples utilizing pie, bar, or bar plots again using EDA technique. We'll choose certain characteristics from either the database besides research during filtering. Separating the dataset into two for testing and training: 75% and 25%. Utilizing machine learning methods to find as well as compare the performance, thereafter determining Accuracy, Recollect, as well as F1 Point total results. This information gets maintained in order to detect each user inputs. That visitor would determine their consequence by providing mistreated via an Interface built with the Python System. Initially, there were maybe two options (Maybe or Just no), but now we aim to offer multiple methods: Low risk, increased risk and High risk, and hence suggest feeding practices. Deployment and analysis on real life scenario the trained and tested prediction model will be deployed in a real-life scenario made by the human experts & will be leveraged for further improvement in the methodology.

8. RESULTS

Table-1: comparison table

Model	F1 Accuracy Score
Decision tree	0.91
Random forest	0.96
KNN	0.73
SVM	0.69
Logistic regression	0.70
Voting classifiers hard and soft (LR, SVM, DT)	0.75 & 0.87 respectively

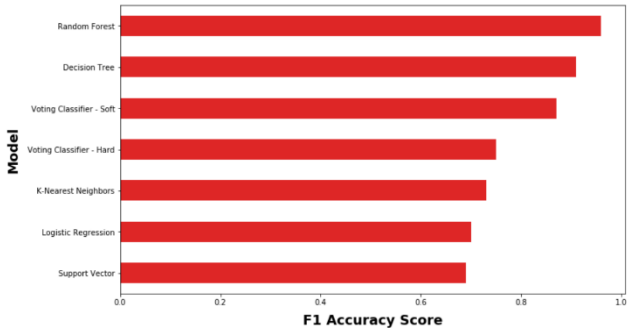

Chart-1: F1 Accuracy score

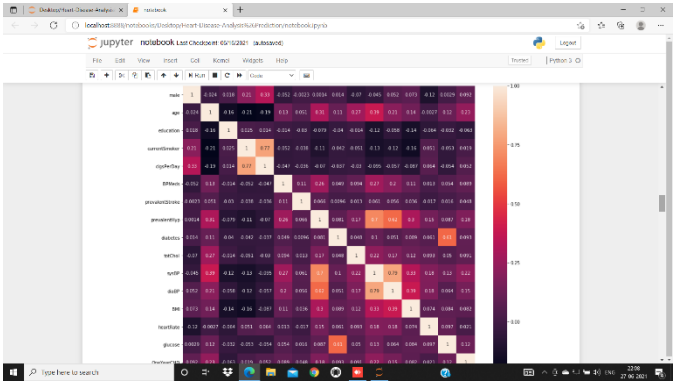

Chart-2: plot correlation to see if columns can be dropped

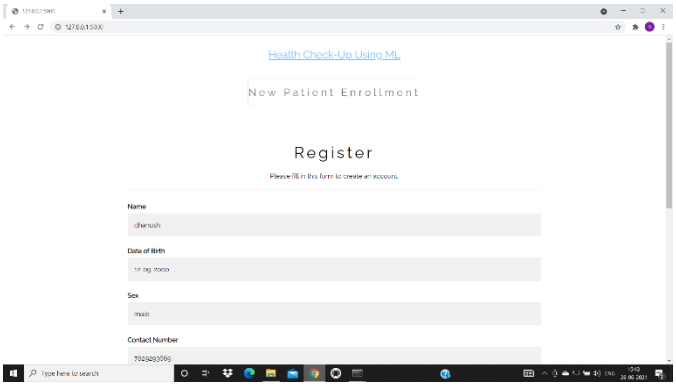

Fig -5: home and registration page

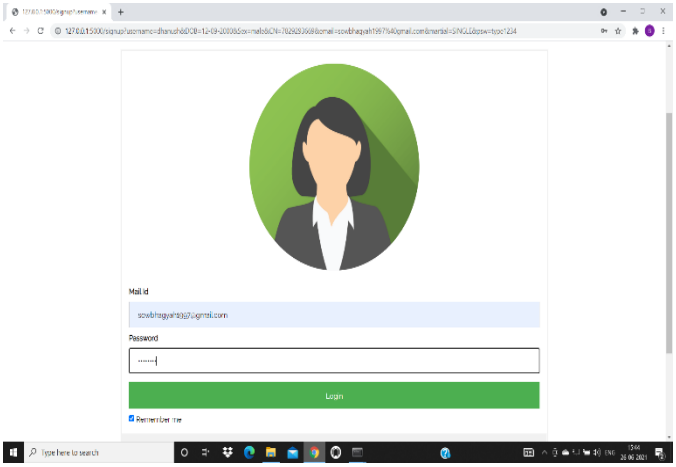

Fig -6: Login page

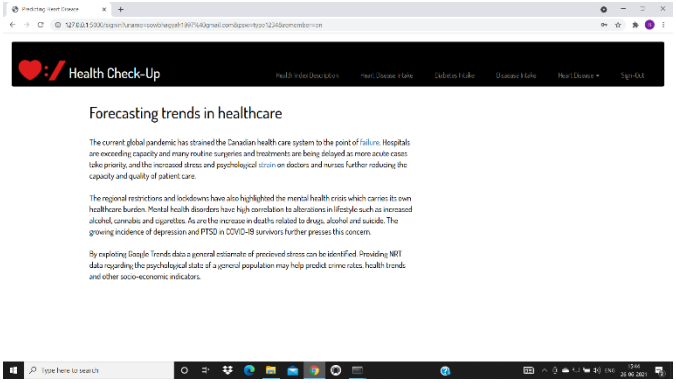

Fig -7: dashboard

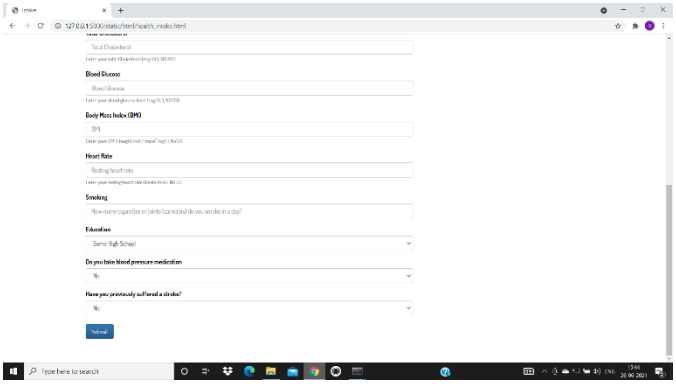

Fig -8: input information

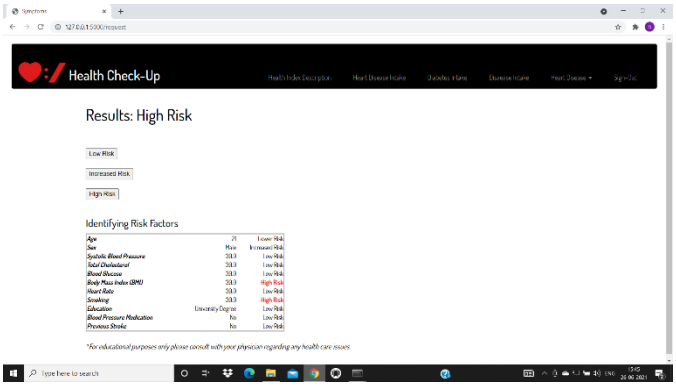

Fig- 9: Prediction result and medication

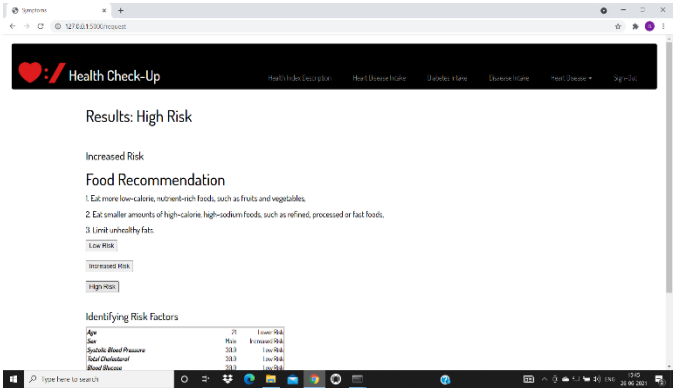

Fig- 10: food recommendation

9. CONCLUSION AND FUTUREWORK

We implemented RF, SVM, LR, DT, KNN and voting classifiers hard and soft. Models enables more accurate predictions and as just a result, to boost and increase the precision. All method's results are evaluated. This Entire collection was examined, and indeed the findings being evaluated. If health industry adopts this project, then the work of the doctors can be reduced and they can easily predict the disease of the patient. The Disease prediction is to provide prediction for the various and generally occurring heart diseases that when unchecked and sometimes ignored can turns into fatal disease and cause lot of problem to the patient and as well as their family members. This same primary goal of this study is to improve the accuracy of cardiovascular diseases predictions. Modification of user information is possible. Operating system that is even more responsive, Additional information as well as the most recent disorders may be found in the options during backups storage, that could be handled as little more than a smartphone app.

REFERENCES

[1] M. Sanz, A. Marco del Castillo, S. Jepsen et al., "Periodontitis and cardiovascular diseases: consensus report," Journal of Clinical Periodontology, vol. 47, no. 3, pp. 268–288, 2020.

[2] World Health Organization. http://www.who.int/cardiovascular diseases/en. 2019.

[3] S. N. Blair, "Commentary on Wang Y et al. "An Overview of Non-exercise Estimated Cardiorespiratory Fitness: estimation Equations, Cross-Validation and Application"" Journal of Science in Sport and Exercise, vol. 1, no. 1, pp. 94-95, 2019.

[4]. A.-H. Abdel-Aty, H. Kadry, M. Zidan et al., "A quantum classification algorithm for classification incomplete patterns based on entanglement measure," Journal of Intelligent and Fuzzy Systems, vol. 38, no. 9, pp. 1–8, 2020.

[5] A. Sagheer, M. Zidan, and M. M. Abdelsamea, "A novel autonomous perceptron model for pattern classification applications," Entropy, vol. 21, no. 8, p. 763, 2019.

[6] M. Aljanabi, H. Qutqut, and M. Hijjawi, "Machine learning classification techniques for heart disease prediction: a review," International Journal of Engineering and Technology, vol. 7, pp. 5373–5379, 2018.

[7] T. Obasi and M. O. Shafiq, "Towards comparing and using machine learning techniques for detecting and predicting heart attack and diseases," in Proceedings of the 2019 IEEE International Conference on Big Data (Big Data), pp. 2393–2402, Los Angeles, CA, USA, April 2019.

[8] A. A. Ali, H. S. Hassan, and E. M. Anwar, "heart diseases diagnosis based on a novel convolution neural network and gate recurrent unit technique," in Proceedings of the 2020 12th International Conference on Electrical Engineering (ICEENG), vol. 145–150, Cairo, Egypt, July 2020.

[9] N. S. C. Reddy, S. S. Nee, L. Z. Min, and C. X. Ying, "Classification and feature selection approaches by machine learning techniques: heart disease prediction," International Journal of Innovative Computing, vol. 9, 2019.

[10] R. Atallah and A. Al-Mousa, "heart disease detection using machine learning majority voting ensemble method," in Proceedings of the 2019 2nd International Conference on New Trends in Computing Sciences (ICTCS), pp. 1–6, Amman, Jordan, October 2019.

[11] N. S. R. Pillai, K. K. Bee, and J. Kiruthika, "Prediction of heart disease using rnn algorithm," International Research Journal of Engineering and Technology, vol. 5, 2019.

[12] R. Kannan and V. Vasanthi, "Machine learning algorithms with roc curve for predicting and diagnosing the heart disease," InSoft Computing and Medical Bioinformatics, pp. 63– 72, 2019.

[13] K. Raza, "Improving the prediction accuracy of heart disease with ensemble learning and majority voting rule," InUHealthcare Monitoring Systems, pp. 179–196, 2019.

[14] A. N. Oo and K. T. Win: Feature Selection Based Sequential Minimal Optimization (Smo) Classifier for Heart Disease classification.

[15] S. Nalluri, R. V. Saraswathi, S. Ramasubbareddy, K. Govinda, and E. Swetha, "Chronic heart disease prediction using datamining techniques," In Data Engineering and Communication Technology, pp. 903–912, 2020.

[16] R. Bhat, S. Chawande, and S. Chadda, "Prediction of test for heart disease diagnosis using artificial neural network," Indian Journal of Applied Research, vol. 9, 2019.

[17] L. Sapra, J. K. Sandhu, and N. Goyal, "Intelligent method for detection of coronary artery disease with ensemble approach," Advances in Communication and Computational Technology, vol. 1033–1042, 2021.

[18] C. Ricciardi, A. S. Valente, K. Edmund et al., "Linear discriminant analysis and principal component analysis to predict coronary artery disease," Health Informatics Journal, vol. 26, no. 3, pp. 2181–2192, 2020.

[19] A. K. Garate-Escamilla, A. H. E. Hassani, and E. Andres, "Classification models for heart disease prediction using feature selection and pca," Informatics Medicine Unlocked, vol. 19, Article ID 100330, 2020.