International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 04 | April - 2022 Impact Factor: 7.185 ISSN: 2582-3930

# Different Diseases Prediction System

**Ayansh Namdeo[1] , D. Nihal Reddy[2] , Darshita Khanna[3] , Devyani Rathore[4] , Monika Barfa[5] , Nikhil Dubey[6] , Prof. Ronak Jain[7]**

[1,2,3,4,5,6,7]Department of Information Technology,
[7] Faculty of Information Technology,
[1,2,3,4,5,6]Acropolis Institute of Technology and Research, Indore, Madhya Pradesh
[7]Acropolis Institute of Technology and Research, Indore, Madhya Pradesh

***Abstract* —** With huge facts boom in biomedical and healthcare communities, correct evaluation of scientific data advantages early disease detection, affected individual care, and neighborhood services. However, the contrast accuracy is reduced when the satisfactory on medical data is incomplete. Moreover, one-of-a-kind areas show off unique qualities of positive regional diseases, which may weaken the prediction of health problem outbreaks. In this paper, we streamline laptop computer getting to know algorithms for effective prediction of general sickness outbreak in disease-frequent communities. The utility of laptop computer gaining information of in the challenge of medical analysis is developing gradually. This can be contributed particularly to the enchantment in the classification and recognition buildings used in disease prognosis which is successful to furnish data that aids medical professionals in early detection of fatal illnesses and therefore, make bigger the survival charge of victims significantly.

The effects of the learn about give a boost to the notion of the software of computing system getting to know in early detection of diseases. Compared to several ordinary calculating algorithms, the scheming accuracy of our proposed algorithm reaches ---- with a regular speed which is quicker than that of the unimodel disease risk prediction algorithm and produces report.

***Keywords***— Disease Prediction System, Machine Learning, Random Forest(RF), Support Vector Machine (SVM), Symptoms.

## I. .INTRODUCTION

It is a tool this is made via the usage of computing tool gaining knowledge of algorithms for guessing the possible illnesses based totally at the affected person's symptoms. The growth of science has been improving our lives thus far. It offers many equipment that could keep hundreds and hundreds of lives, and computer gaining knowledge of is one in all them. Machine gaining knowledge of is used to reinforce systems which can assist us predict such a lot of ailments based totally on signs. It is able to Endorse the medical doctors, likelihood of the feasible diseases. And evaluation can be accomplished primarily based on suggestion, because of this price might be reduced. We are residing inside the age of technological know-how and nowadays human beings can say that almost some aspect is possible with the help of generation. These days we've such a lot of equipment and techniques to get proper of entry to statistics from any area of this international and statistics at this age is so critical that without records we would now not continue to exist. We've tools which could provide us or suggest relevant facts at our fingertips and the internet is the sort of gear. Nowadays billions of search queries are accomplished every day and on occasion there given results are relevant and on occasion they are not. In these search queries, lots of searches are associated to clinical advice. Humans regularly favor to understand if they've any critical ailments primarily based mostly on their signs and symptoms and symptoms. However, there aren't any device accessible to deliver them proper records. This lookup attempts to supply them equipment so that viable illness prediction information can be provided to the end-user.

## II.LITERATURE SURVEY

MIN CHEN et al, proposed a disease prediction system in his paper where he used machine learning algorithms. In the prediction of disease, he used techniques like CNN-UDRP algorithm, CNN-MDRP algorithm, Naive Bayes, K-Nearest Neighbor, and Decision Tree. This proposed system had an accuracy of 94.8%.

( MIN CHEN)

Sayali Ambekar et al, [2] recommended Disease Risk Prediction and used a convolution neural network to perform the task. In this paper machine learning techniques like CNN-UDRP algorithm, Naive Bayes, and KNN algorithm are used. The system uses structured data to be trained and its accuracy reaches 82% and achieved by using Naïve Bayes.

(Sayali Ambekar)

Darcy A. Davis, Nitesh V. Chawla, Nicholas Blumm, Nicholas Christakis, Albert-Laszlo Barabasi have found that global treatment of chronic disease is neither time or cost efficient. So the authors conducted this research to predict future disease risk. For this CARE was used (which relies only on a patient's medical history using ICD- 9-CM codes in order to predict future diseases risks). CARE combines collaborative filtering methods with clustering to predict each patient's greatest disease risks based on their own medical history and that of similar patients. Authors have also described an Iterative version, ICARE, which incorporates ensemble concepts for improved performance. These novel systems require no specialized information and provide predictions for medical conditions of all kinds in a single run. The impressive future disease coverage of ICARE represents more accurate early warnings for thousands of diseases, some even years in advance. Applied to full potential, the CARE framework can be used explore a broader disease. Predictor 7 histories, suggest previously unconsidered concerns, and facilitating discussion about early testing and prevention.

(A.Davis, V. Chawla, Blumm, Christakis, & Barbasi, 2008)

Jyoti Soni, Ujma Ansari, Dipesh Sharma and Sunita Soni have done this research research paper into provide a survey of current techniques of knowledge discovery in databases using data mining techniques that are in use in today's medical research particularly in Heart Disease Prediction. Number of experiments has been conducted to compare the performance of predictive data mining technique on the same dataset and the outcome reveals that Decision Tree outperforms and sometime Bayesian classification is having similar accuracy as of decision tree but other predictive methods like KNN, Neural Networks, Classification based on clustering is not performing well.

(Jyoti Soni, Ansari, Sharma, & Soni, 2011)

Shadab Adam Pattekari and AsmaParveen have conducted research using Naïve Bayes Algorithm to predict the heart diseases where user provides the data which is compared with trained set of values. So from this research, patients were able to provide their basic information which is compared with the data and the heart disease is predicted.

(Adam & Parveen, 2012)

M.A. Nishar Banu, B Gomathy used medical data mining techniques like association rule mining, classification, clustering I to analyze the different kinds of heart-based problems. Decision tree is made to illustrate every possible outcome of a decision. Different rules are made to get the best outcome. In this research age, sex, smoking, overweight, alcohol intake, blood sugar, hear rate, blood pressure are the parameters used for making the decisions. Risk level for different parameters is stored with their id's ranging (1-8). ID lesser than of 1 of weight contains the normal level of prediction and higher ID other than 1 comprise the higher risk levels. K-means clustering technique is used to study the pattern in the dataset. The algorithm clusters information's into k groups. Each point in the dataset is assigned to the closed cluster. Each cluster center is recomputed as the average of the points in that cluster.

(Nishar Banu, MA; Gomathy, B; 2013)

## III. PROPOSED WORK

The proposed application has more than a few customers namely Administrator, Doctor, Patient. The function of the administrator is to add or put off users. The Doctor's role consists of naming the ailment and attends to patient. The system will predict diseases based on the symptoms entered by the patient. The parameters can be in the form symptoms. Proposed system will be capable to take care of affected person queries from begin to end. It will reply with splendid solutions to queries when asked. The gadget would be light-weight with as many questions and responses aspossible.

- The key motive of our technique is that if consumers do no longer have ride of the scientific career and want to recognize about their fitness conditions, they will rapidly discover it besides a technological or medical person.
- Machine learning is the core principle in which the system offers greater unique forecasts.
- The SVM and RF is greater dependable on a similar scenario than the scientific subject estimation.

## IV.METHODOLOGY

In our system, there are three Modules: Admin, User (Patient), Doctor. Every new person has to get registered through admin. After profitable registration consumer wishes to join first before login. Users will want to sign up solely once. The disorder prediction gadget have three customers such as doctor, patient and admin.

- Each consumer of the gadget is authenticated by using the system.
- There is a position based totally get right of entry to to the system.
- The device lets in the affected person to provide signs and symptoms and according to these signs the device will predict a disease.
- The application suggests medical practitioner for envisioned disease.
- The device approves on line session for patients.
- The machine helps the sufferers to seek advice from the medical doctor at their comfort via sitting at home.

## V. ALGORITHM USED

As the name suggests, In our disease prediction system, We are using the Support Vector Machine (SVM) for classification and Random Forest (RF) for predicting the result.

A) Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning technique that is widely used in pattern recognition and classification problems. The SVM algorithm performs a classification by constructing a multidimensional hyperplane that optimally discriminates between two classes by maximizing the margin between two data clusters. This algorithm achieves high discriminative power by using special nonlinear functions called kernels to transform the input space into a multidimensional space

B) Random Forest Algorithm

Random Forest algorithm developed from trees algorithm and bagging algorithm is modelled. The developed the algorithm found that it can potentially improve classification accuracy. It is also work well with a data set with large number of input variables. The algorithm is started by creating a combination of trees which each will vote for a class .

## VI. SYSTEM TESTING

The proposed system was tested and results were compared with those results generated by existing systems. The various test performed were: unit test, integration test and user acceptance test.

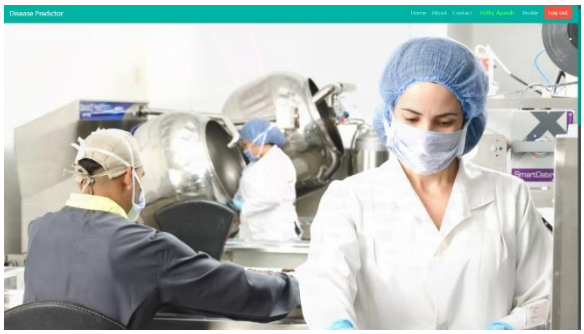
**1**. ***Unit Testing:*** This tests the modules and routines that are assembled and integrated to perform specific function. Unit testing focuses first on modules, independently of one another, to locate errors. This enables a system developer to detect errors in coding and logic that are contained within each module. This testing includes entering data and ascertaining if the value matches to the type and size supported by Visual Basic 6.0. The various controls are tested to ensure that each performs its action as required.

**2. *Integration Testing:*** Data can be lost across any interface, one module can have an adverse effect on another. Integration testing is a systematic testing to discover errors associated within the interface. The objective is to take unit tested modules and build a program structure. All the modules are combined and tested as a whole. During this phase, we integrated and tested the Server module and Client module options.

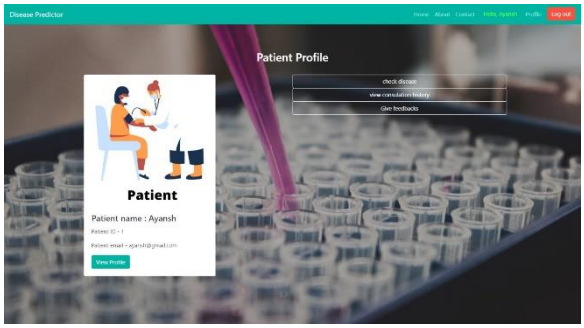
**3**. ***User Acceptance Testing:*** The proposed system was developed to meet user's acceptance by subjecting the system to various considerations during development.

## VII. SCREENSHOTS

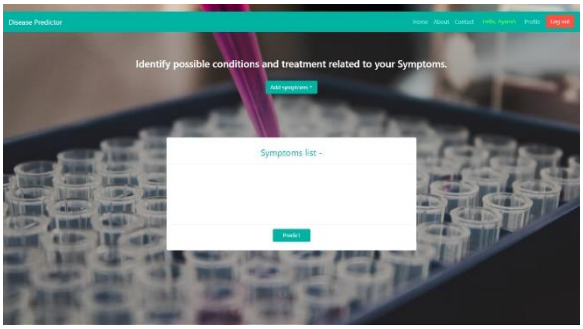
**Home Page**

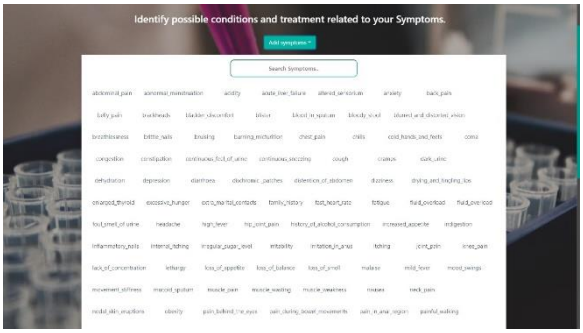


**Patient Profile**

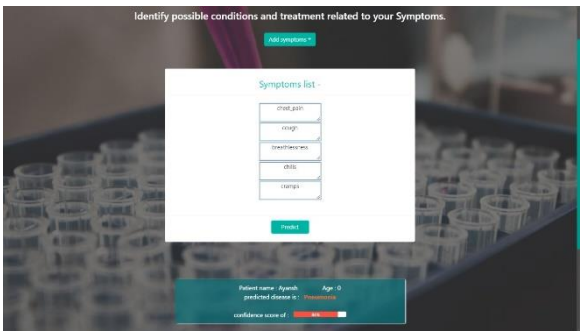



**Symptoms Page**




**Predicted Disease**




## VIII. CONCLUSION AND FUTURE SCOPE

This project aims to predict the disease on the basis of the symptoms. The project is designed in such a way that the system takes symptoms from the user as input and produces output i.e. predict disease. Average prediction accuracy probability of 85.6% is obtained. Disease Predictor was successfully implemented using machine learning.

The further work will mainly focus on the Medical Assistance and proper Medication to the patients as soon as possible soo as to build the best infrastructure and quick easiest way in the Medical sectors.

## IX. ACKNOWLEDGMENT

It gives us great pleasure in presenting this project report titled: “Different Diseases Prediction System”. On this momentous occasion, we wish to express our immense gratitude to the range of people who provided invaluable support in the completion of this project. Their guidance and encouragement has helped in making this project a great success.

We express our gratitude to our project guide Prof. Ronak Jain, who provided us with all the guidance and encouragement and making the lab available to us at any time. We also would like to deeply express our sincere gratitude to Project coordinators.

We are eager and glad to express our gratitude to the Head of the Information Technology Dept. Prof. Prashant Lakkadwala, for his approval of this project. We are also thankful to him for providing us the needed assistance, detailed suggestions and also encouragement to do the project.

## X.REFERENCES

[1] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, “Disease prediction by machine learning over big data from healthcare communities” IEEE Access, vol. 5, no. 1, pp. 8869–8879, 2017.

[2] Sayali Ambekar, Rashmi Phalnikar, “Disease Risk Prediction by Using Convolutional Neural Network” IEEE, 978-1-5386-5257-2/18, 2018.

[3]A.Davis, D., V.Chawla, N., Blumm, N., Christakis, N., & Barbasi, A. L. (2008). Predicting Individual Disease Risk Based On Medical History.

[4]Adam, S., & Parveen, A. (2012). Prediction System For Heart Disease Using Naive Bayes.

[5] Al-Aidaroos, K., Bakar, A., & Othman, Z. (2012). Medical Data Classification With Naive Bayes Approach. Information Technology Journal .

[6]Darcy A. Davis, N. V.-L. (2008). Predicting Individual Disease Risk Based On Medical History.

[7]JyotiSoni, Ansari, U., Sharma, D., & Soni, S. (2011). Predictive Data Mining for Medical Diagnosis: An Overview Of Heart Disease Prediction.

[8]K.M. Al-Aidaroos, A. B. (n.d.). K.M. Al-Aidaroos, A. B. (n.d.). 2012. Medical Data Classification With Naive Bayes Approach .

[9]NisharBanu, MA; Gomathy, B;. (2013). Disease Predicting System Using Data Mining Techniques