

# **Disease Prediction using Machine Learning**

**A Synopsis Submitted to**



**Rajiv Gandhi Proudhyogiki Vishwavidyalaya, Bhopal  
Towards Partial Fulfillment for the Award of**

**Bachelor of Technology  
(Computer Science and Engineering)**

**Under the Supervision of  
Professor Preeti Shukla**

**Submitted By**  
**Aman Patel (0827CS213D03)**  
**Bhomik Gahlot (0827CS213D08)**  
**Chandrans**  
**chouhan(0827CS201062)**  
**Anurag yadav (0827CS201039)**



**Department of Computer Science and Engineering  
Acropolis Institute of Technology & Research, Indore  
July-Dec 2022**

## **Abstract**

It is a system which provides the user the information and tricks to take care of the health system of the user and it provides how to search out the disease using this prediction. Now a day's health industry plays major role in curing the diseases of the patients so this is often also some quite help for the health industry to inform the user and also it's useful for the user just in case he/she doesn't want to travel to the hospital or the other clinics, so just by entering the symptoms and every one other useful information the user can get to grasp the disease he/she is affected by and also the health industry may also get enjoy this method by just asking the symptoms from the stoner and entering within the system and in only many seconds they'll tell the precise and over to some extent the accurate conditions. This Disease Prediction Using Machine Learning is totally through with the assistance of Machine Learning and Python programming language and also using the dataset that's available previously by the hospitals using that we are going to predict the diseases.

## **Introduction of the Project**

The purpose of constructing this project called "Disease Prediction Using Machine Learning" is to predict the accurate disease of the patient using all their general information and also the symptoms. If this Prediction is completed at the first stages of the disease with the assistance of this project and every other necessary measure, disease is cured and generally this prediction system can even be very useful in the health industry. The final purpose of this Disease prediction is to supply predictions for the assorted and customarily occurring diseases that when unchecked and sometimes ignored can turn into fatal disease and cause a lot of problems to the patient and moreover as their members of the family. This method will predict the foremost possible disease supporting the symptoms. The health industry is information yet and knowledge poor and this industry is an incredibly vast industry which has a lot of labor to be done. So, with the assistance of all those algorithms, techniques and methodologies we've done this project which is able to help the peoples who are within the need.

## **Objective**

Our primary objective is to make an application: - There is a need to study and make a system which will make it easy for an end users to predict the chronic diseases without visiting physician or doctor for diagnosis. To detect the Various Diseases through the examining Symptoms of patient's using different techniques of Machine Learning Models. To Handle Text data and Structured data is no Proper method. The Proposed system will consider both structure and unstructured data. The Predictions Accuracy will Increase using Machine Learning.

## **Scope**

Here the scope of the project is that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to

generate a knowledge-rich environment which can help to significantly improve the quality of clinical decision

## **Study of Existing System**

The system predicts the chronic diseases which is for particular region and for the particular community. The Prediction of Diseases is done only for particular diseases. In this System Big Data & CNN Algorithm is used for Diseases risk prediction. For S type data, system is using Machine Learning algorithm i.e K-nearest Neighbors, Decision Tree, Naïve Bayesian. The accuracy of the System is upto 94.8%. Existing paper, we streamline machine learning algorithms for effective prediction of chronic disease outbreak in disease-frequent communities. We experiment the modified prediction nm models over reallife hospital data collected from central China. We propose a new convolutional neural net-work based multimodal disease risk prediction (CNN-MDRP) algorithm using structured and unstructured data from hospital.

## ***Existing Problems***

By extracting characteristics and classifying them, the deep learning model achieves great results. However, the current system's model interpretability is inadequate. It is necessary to connect medical domains and deep learning interpretability of models. The interpretability of models refers to the degree to which humans can comprehend decision-making logic. Deep learning's problem is that it is entirely data-driven, with no regard for prior domain expertise or experience, as well as risk considerations. The deep learning model has been fully trained, and new data is being fed into it, with detection results being generated. The model, on the other hand, only presents the classification results depending on the input data and does not indicate how to detect or predict. The model's credibility is determined by its interpretability. As a result, future research should pay more attention to model interpretability.

## **Project Description :-**

- **DATA AUGMENTATION**

Data Augmentation is a technique for increasing training datasets without having to gather new images. Data augmentation alters the original images in some way. This is accomplished by using various processing techniques including as rotations, flips, zooming, and adding noise, among others. Large training datasets are significant in deep learning since they improve the training model's accuracy. It also aids in the avoidance of overfitting. The downsides of data augmentation include increased training time, transformation computation costs, and higher memory expenses.

## ● RELATED WORK

According to [5], a new strategy incorporating random forests (RF) and contour-based models is utilised to extract glioma features from multivariate volumetric MR images. They also use random forests algorithm as feature training kernels to analyse both geographic data, accuracy data from multiple images for tumour diagnosis utilising a feature representational technique for learning.

Discriminatory features and sparseness were included into the PCA model by [6]. Instead of the standard sparse PCA, which enforces sparseness on the loadings, sparse components are created to reflect the data.

According to [7], A suggested strategy based solely on 3D convolutional neural networks (CNN) provides effective performance on the publicly available dataset for lung nodule identification and malignancy classification. While methods for detecting nodules are frequently developed and improved separately, the relationship between component identification and classification is crucial.

According to [8], For lung nodule recognition and categorization, an enhanced multidimensional Region-based Fully Convolutional Network (mRFCN) based automated system was used. The mRFCN is being used to investigate the multi-Layer fusion Region Proposal Network (mLRPN) using position-sensitive score maps (PSSM) as an image classifier for extracting features. Then, using the suggested mLRPN, a median intensity projection was employed to take benefit of 3D information from CT scans, and then a de-convolutional layer was added to the architecture to autonomously choose possible zones of interest.

## ● IMPLEMENTATION

The project malady Prediction mistreatment Machine Learning is developed to beat general malady in earlier stages as we tend to all recognize in competitive surroundings of economic development the human race has concerned thus much that he/she isn't involved regarding health per analysis there area unit four-hundredth peoples however Ignores regarding general malady that ends up in harmful malady later. Even the interface of this project is completed mistreatment python's library interface referred to as Tkinter. Here 1st the user must register into the system so as to use the prediction, user must register with username, email-id, phone, agenda parole. of these values area unit keep into the filing system severally, then user has choice to move forward or leave, then user must login to the system mistreatment the username and parole that he/she provided throughout the time of registration. If he/she enter incorrect username and proper parole then the error message can prompt stating incorrect username and he/she enters incorrect parole and proper when work within the user must the name and desires to pick out the symptoms from given change posture menu, for additional correct result the user must enter all the given symptoms, then the system can give the correct result. This prediction is essentially through with the assistance of three algorithms of machine learning like call Tree, Random Forest and Naïve mathematician. once user enter all the symptoms then he must press the buttons of various rule, for instance there area unit three buttons for three algorithms, if user

enters all symptoms and presses solely Random Forest button then the result are going to be provided solely shrewd mistreatment that rule, like this we've got used three algorithms to produce additional clear image of the results and user must be happy along with his expected result.

## Expected Outcome

The result for this prediction system displays a convenient user interface consisting of details like name, symptoms and the algorithm that we use to predict as a button and the results will be predicted based on the implemented algorithm. It also displays the accuracy percentage on which algorithm has the best accuracy so based on the accuracy of the decision tree, random forest and naive bayes algorithm random forest has the better accuracy percentage of 0.96. It is a best suited algorithm for this model.

**Disease Predictor**

Name of the Patient:

Symptom 1:

Symptom 2:

Symptom 3:

Symptom 4:

Symptom 5:

**RandomForest**

**ExtraTree**

**NaiveBayes**

**SVM**

**RandomForest**

**ExtraTree**

**NaiveBayes**

**SVM**

## Resources

Developer's Requirement: -

Hardware:

- A computer with Vs code
- Must be optimized for running Machine learning Software:
- Permission for user Data
- Permission to user data for augmented reality features
- Permission to use internet

User's Requirement: -

Hardware:

- An android phone
- An internet connection

Software:

- Android version 5.0 (Lollipop)

## Limitations

- Requires an internet connection
- Users need to choose symptoms names
- Smaller disease or basic pain data is unavailable
- Lag time may increase due to augmented reality feature

## Conclusion

The main aim of this disease prediction system is to predict the disease on the basis of the symptoms. This system takes the symptoms of the user from which he or she suffers as input and generates final output as a prediction of disease. Average prediction accuracy probability of 100% is obtained. Disease Predictor was successfully implemented using the grails framework. This system gives a user-friendly environment and easy to use. As the system is based on the web application, the user can use this system from anywhere and at any time. In conclusion, for disease risk modeling, the accuracy of risk prediction depends on the diversity feature of the hospital data. This systematic review aims to determine the performance, limitations, and future use of Software in health care. Findings may help inform future developers of Disease Predictability Software and promote personalized patient care. The program predicts Patient Diseases. Disease Prediction is done through User Symbols.

In this System Decision tree, Unplanned Forest, the Naïve Bayes Algorithm is used to predict diseases. For the data format, the system uses the Machine Learning algorithm Process Data on Database Data namely, Random Forest, Decision Tree, Naive Bayes. System accuracy reaches 98.3%. machine learning skills are designed to successfully predict outbreaks.

## References

- [1] Philippe Meister, Jack Miller, Kexin Wang, Michael C. Dorneich, Eliot Winer, "Using Three Dimensional Augmented Reality to Enhance General Aviation Weather Training".
- [2] Aruna R., Vasuki S., Dhanya M. R., Divya V., Gopika G. S., Harini M., Haritha P., "Weather Forecasting and Its Visualizations Using AI".
- [3] Marko Heinrich, Bruce H. Thomas, Stefan Mueller, Christian Sandor, "An Augmented Reality Weather System".
- [4] Stephan Rasp, Peter D. Dueben, Sebastian Scher, Jonathan A. Weyn, Soukayna Mouatadid, Nils Thuerey, "WeatherBench: A benchmark dataset for data-driven weather forecasting".