

A Project Synopsis Submitted for  
*Bachelor of Engineering*  
Project & Seminar (8KS07)



**PROJECT TITLE** : Analysis and Prediction of Common Health Issues using Machine Learning

**PROJECT GROUP NO.** : C-1

**NAME OF PROJECTEE'S** : 1) Yash J. Ingole  
2) Dhawal P. Ramdham  
3) Pranjal P. Hejib  
4) Pratik S. Rohankar

**NAME OF GUIDE** : Dr. R. R. Karwa

Department of Computer Science & Engineering,  
Prof. Ram Meghe Institute of Technology & Research,  
Badnera - Amravati  
2022-2023

# Abstract

Most of the medical diagnoses require going to the doctor and fixing appointments for a consultation and sometimes to get accurate disease indications, they have to wait for blood reports also they have to travel long distances to seek doctors consultation. When they are not feeling well the first thing they do is to check temperature to get an estimate or baseline idea of their fever so they can consult their doctor if the temperature is high enough similarly a medical disease prediction application can be used to get a baseline idea of disease and prediction model can indicate whether they should take immediate doctor consultation or not, or at least start some home remedies for the same to find temporary relief. Combining machine learning with an application interface to interact with users provides opportunities for easy interaction with the users with the machine learning model to get more accurate predictions.

# 1 Introduction

In terms of data collecting and processing, healthcare is one of the most worrisome industries. With the advent of the digital era and technological advancements, a vast quantity of multidimensional data on patients is created, including clinical factors, hospital resources, illness diagnostic information, patients' records, and medical equipment. The enormous, dense, and complex data must be processed and evaluated in order to extract knowledge for effective decision making. Medical data mining offers a lot of potential for uncovering hidden patterns in medical data sets [1].

By identifying significant patterns and detecting correlations and relationships among many variables in huge databases, the use of various data mining tools and machine learning approaches has changed healthcare organizations [2,3]. It serves as an important instrument in the medical sector, providing and comparing existing data for the future course of action. This technology combines multiple analytic methodologies with modern and complex algorithms, allowing for the exploration of massive amounts of data [4]. It is used in healthcare to gather, organize, and analyze patient data in a systematic manner. It may be used to identify inherent inefficiencies and best practices for providing better services, which may lead to improved diagnosis, better medicine, and more successful treatment, as well as a platform for a deeper knowledge of the mechanisms in practically all elements of the medical domain. Overall, it assists in the early detection and prevention of disease epidemics by searching medical databases for relevant information.

## 1.1 Motivation

There is a demand to make an ecosystem that will help end users to predict diseases on the basis of symptoms given in it without visiting hospitals. By doing so, it decreases the rush at OPD of hospitals and bring down the workload on medical staff. Not only this, this system reduces the costly treatment and panic moment at the end stages so that proper medication can be provided at the right time and which can lower down the death rate as well.

## 2 Literature Survey

This section outlines that what has been done so far in **Analysis and Prediction of Common Health Issues using Machine Learning**.

Kunal Takke et al.[16] To identify and predict patients suffering from more prevalent ailments using Multinomial Naive Bayes, Random Forest Classifier, K-Nearest Neighbours MNB :98% RFC : 98.2% KNN : 97%

Naganna Chetty et al.[17] developed a system that gives improved results for disease prediction and used a fuzzy approach. And used techniques like KNN classifier, Fuzzy c-means clustering, and Fuzzy KNN classifier. In this paper diabetes disease and liver, disorder prediction is done and the accuracy of Diabetes is 97.02% and Liver disorder is 96.13%

Dhiraj Dahiwade et al.[18] designed a model for prediction of the disease using approaches of machine learning and used techniques like KNN and CNN. This paper suggests disease prediction i.e. based on patient's symptoms. The accuracy of KNN is 95% and the accuracy of CNN is 98%.

Pahulpreet Singh Kohli et al.[19] suggested disease prediction by using applications and methods of machine learning and used techniques like Logistic Regression, Decision Tree, Support Vector Machine, Random Forest and Adaptive Boosting. This paper focuses on predicting Heart disease, Breast cancer, and Diabetes. The highest accuracies are obtained using Logistic Regression that is 95.71% for Breast cancer, 84.42% for Diabetes, and 87.12% for Heart disease.

Senthilkumar Mohan et al.[20] focused on hybrid techniques in machine learning that can be used for effectively predicting heart disease and used algorithms like Decision Tree, Support Vector Machine, Random Forest, Naïve Bayes, Neural Network and KNN. The accuracy of this system is 88.47%

Dhomse Kanchan B. et al.[21] studied special disease prediction utilizing principal component analysis using machine learning algorithms involving techniques like Naive Bayes classification, Decision Tree, and Support Vector Machine. The accuracy of this system is 34.89% for Diabetes and 53% for Heart disease.

Rati Shukla et al.[22] suggested prediction and detection for breast cancer by utilizing machine learning techniques like Decision Tree, Support Vector Machine, Random Forest, Naïve Bayes, Neural Network, and KNN. In this system, the Support Vector Machine gives more accurate results than all other algorithms.

Rashmi G Saboji et al.[23] tried to find a scalable solution that can predict heart disease utilizing classification mining and used Random Forest Algorithm. This system presents a comparison against Naïve-Bayes classifier but Random Forest gives more accurate results with accuracy 98%.

Anjan Nikhil Repaka et al.[24] designed and implemented a prediction model for heart disease using naive Bayesian. Any user can use this system using any smart-phone device and get the prediction results. The accuracy of this system is 89.77%. of CNN is 98%.

Lambodar Jena et al.[25] focused on risk prediction for chronic diseases by taking advantage of distributed machine learning classifiers and used techniques like Naive Bayes and Multilayer Perceptron. This paper tries to predict Chronic-Kidney-Disease and the accuracy of Naïve Bayes and Multilayer Perceptron is 95% and 99.7% respectively.

Ankita Dewan et al.[26] recommended a disease prediction system that uses data mining classification hybrid technique for predicting heart disease. This system is using techniques like Neural Network, Decision Tree, and Naive Bayes. The accuracy of this system is 87%.

Md. Ehtisham Farooqui et al.[27] Various models based on such algorithms, techniques is presented and their performance is analyzed. Researches have been conducted on various models of supervised learning algorithms.Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Decision Tree (DT), Naïve Bayes and Random Forest (RF)with Average accuracy : 95%.

**Table[2.1]: Purpose & Techniques of Conventional models**

Sr no.	Year	Author	Purpose	Technique Used	Accuracy
1	2022	Kunal Takke et al.[12]	To identify and predict patients suffering from more prevalent ailments	Multinomial Naive Bayes, Random Forest Classifier, K-Nearest Neighbors	MNB :98% RFC : 98.2% KNN : 97%
2	2021	Naganna Chetty et al.[13]	Developed a system that gives improved results for disease prediction and used a fuzzy approach	KNN classifier, Fuzzy c-means clustering, and Fuzzy KNN classifier	Diabetes: 97.02% Liver disorder: 96.13%
3	2019	Dhiraj Dahiwade et al.[14]	Designed a model for prediction of the disease using approaches of machine learning	K-Nearest neighbor (KNN) and Convolutional neural network (CNN)	KNN: 95% CNN: 98%
4	2018	Pahulpreet Singh Kohli et al.[15]	Suggested disease prediction by using applications and methods of machine learning	Logistic Regression, Decision Tree	Logistic Regression : Breast Cancer:95.71% Diabetes: 84.42% Heart Disease: 87.12% Decision Tree : Breast Cancer: 94.29% Diabetes: 74.03% Heart Disease: 70.97%
5	2019	Senthil Kumar Mohan et al.[16]	Focused on hybrid techniques in machine learning that can be used for effectively predicting heart disease	Decision Tree, Support Vector Machine, Random Forest, Naïve Bayes, Neural Network and KNN	88.4%

6	2018	Dhom se Kanchan B. et al.[17]	Studied special disease prediction utilizing principal component analysis using machine learning algorithms	Naive Bayes classification, Decision Tree and Support Vector Machine	Diabetes Disease: 34.89% Heart Disease: 53%
7	2019	Rati Shukla et al.[18]	Suggested prediction and detection for breast cancer by utilizing machine learning technique	Naive Bayes Classifier, Logistic Regression, Support Vector Machines(SVM),Artificial Nueral Networks and K-Nearest Neighbor	98.2%
8	2017	Rashmi G Saboji et al.[19]	Tried to find a scalable solution that can predict heart disease utilizing classification mining	Random Forest Algorithm	98%
9	2019	Anjan Nikhil Repak a et al.[20]	Designed and implemented a prediction model for heart disease using naive Bayesian	Naive Bayes	89.7%
10	2017	Lambo dar Jena et al.[21]	Focused on risk prediction for chronic diseases by taking advantage of distributed machine learning classifiers	Naive Bayes,Multilayer Perceptron	Naive Bayes : 95% MultiLayer Perceptron : 99.7%

### 3 Project Objectives & Scope

The following are the objectives of our system :

- To collect the data regarding different diseases and their corresponding symptoms.
- To manipulate the available data as per need.
- To fit the given data into a model.
- To use of ML algorithm which provide a maximum accuracy in disease prediction.
- To provide Storage for the name of the disease of the patient in the Database which can be used as past record and will help in future for prediction of diseases.
- To give user-friendly GUI.

#### **Scope:**

- Providing a chatbot service so, can get a instant solution on user problems.
- According to estimated disease system is going to suggest the name of medical practitioners to get a proper treatment.
- Providing the suggestion of diet chart, medicines, therapist etc. on the basis of predicted results.



## 4 Requirement for the proposed work

These are some of the hardware and software requirements of our project. Hardware requirements includes system specifications that are necessary to run software which are required.

### 4.1 Hardware Requirements

- 1 Core i5/i7 processor
- 2 At least 8 GB RAM
- 3 At least 60 GB of Usable Hard Disk Space

### 4.2 Software Requirements

- 1 Software: Anaconda which includes Jupyter Notebook, Python.exe, Spyder etc.
- 2 The project would be developed in Python language using data science libraries like NumPy, Pandas, Matplotlib, Seaborn, Scikit, Sklearn, TensorFlow, Keras.
- 3 We will give the input in CSV file to the system. After taking input all the algorithms can be applied to the dataset. The proposed system will add some more parameters significant to heart attack with their weight, age and the priority levels are by consulting expertise doctors and the medical experts.
- 4 For lung cancer detection we would use DICOM images of cancer affected lungs and develop neural network for cancer prediction using algorithms like CNN.
- 5 The diabetes prediction is proposed using PIMA diabetes dataset to detect and predict using machine learning techniques.

## References

- [1] R. Manne, S.C. Kantheti, Application of artificial intelligence in healthcare: chances and challenges, *Curr. J. Appl. Sci. Technol.* 40 (6) (2021) 78–89, <https://doi.org/10.9734/cjast/2021/v40i631320>.
- [2] M. Sivakami, P. Prabhu. Classification of algorithms supported factual knowledge recovery from cardiac data set, *Int. J. Curr. Res. Rev.* 13(6) 161- 166. ISSN: 2231-2196 (Print) ISSN: 0975-5241
- [3] M. Sivakami, P. Prabhu. A Comparative Review of Recent Data Mining Techniques for Prediction of Cardiovascular Disease from Electronic Health Records. In: Hemanth D., Shakya S., Baig Z. (eds) *Intelligent Data Communication Technologies and Internet of Things. ICICI 2019. Lecture Notes on Data Engineering and Communications Technologies*, vol 38. Springer, Cham 477-484. ISSN 2367-4512 ISSN 2367-4520 (electronic), ISBN 978-3-030-34079-7 ISBN 978-3-030-34080-3
- [4] P. Prabhu, S. Selvabharathi. Deep Belief Neural Network Model for Prediction of Diabetes Mellitus. In 2019 3rd International Conference on Imaging, Signal Processing and Communication, *ICISPC 2019* (pp. 138–142) Institute of Electrical and Electronics Engineers Inc. ISBN:9781728136639. 2019.
- [5] N. Jothi, N.A. Rashid, W. Husain, Data mining in health-care – A review, *Procedia Comput. Sci.* 72 (2015)–313. <https://www.sciencedirect.com/science/article/pii/S1877050915036066>
- [6] H. Polat, H. Danaei Mehr, A. Cetin. Diagnosis of chronic kidney disease based on support vector machine by feature selection methods, *J. Med. Syst.* 41(4) 2017 55.
- [7] K.B. Waghlikar, V. Sundararajan, A.W. Deshpande, Modeling paradigms for medical diagnostic decision support: a survey and future directions, *J. Med. Syst.* 36 (5) (2012) 3029–3049. <https://link.springer.com/article/10.1007/s10916-011-9780-4>
- [8] E. Gurbuz, E. Kılıc, A new adaptive support vector machine for diagnosis of diseases, *Expert Syst.* 31 (5) (2014) 389–397. [http://refhub.elsevier.com/S2214-7853\(21\)05220-2/h0040](http://refhub.elsevier.com/S2214-7853(21)05220-2/h0040)
- [9] M. Seera, C.P. Lim, A hybrid intelligent system for medical data classification, *Expert Syst. Appl.* 41 (5) (2014) 2239–2249. [http://refhub.elsevier.com/S2214-7853\(21\)05220-2/h0045](http://refhub.elsevier.com/S2214-7853(21)05220-2/h0045)

- [10] Y. Kazemi, S.A. Mirroshandel, A novel method for predicting kidney stone type using ensemble learning, *Artif. Intell. Med.* 84 (2018) 117–126. [http://refhub.elsevier.com/S2214-7853\(21\)05220-2/h0050](http://refhub.elsevier.com/S2214-7853(21)05220-2/h0050)
- [11] Amit Kumar Das, Aman Kedia, “Data mining techniques in Indian Healthcare: A Short Review”, 2015 International Conference on Man and Machine Interfacing (MAMI), 978-1-5090-0225-2/15
- [12] Kunal Takke, “Medical Disease Prediction using Machine Learning Algorithm”, *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* ISSN: 2321-9653 IC Value: 45.98; SJ Impact Factor: 7.538 Volume 10 Issue V May 2022
- [13] Naganna Chetty, Kunwar Singh Vaisla and Nagamma Patil, “An Improved Method for Disease Prediction using Fuzzy Approach” *IEEE*, DOI 10.1109/ICACCE.2021.67, pp. 569-572, 2021.
- [14] Dhiraj Dahiwade, Gajanan Patle and Ektaa Meshram, “Designing Disease Prediction Model Using Machine Learning Approach” *IEEE Xplore* Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4, pp. 1211-1215, 2019.
- [15] Pahulpreet Singh Kohli and Shriya Arora, “Application of Machine Learning in Disease Prediction” *IEEE*, 978-1-5386-6947-1/18, pp. 1-4, 2018.
- [16] Senthilkumar Mohan, Chandrasegar Thirumalai and Gautam Srivastava, “Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques” *IEEE Access*, DOI 10.1109/ACCESS.2019.2923707, pp. 81542-81554, 2019.
- [17] Dhomse Kanchan B. and Mahale Kishor M., “Study of Machine Learning Algorithms for Special Disease Prediction using Principal of Component Analysis” *IEEE*, 978-1-5090-0467-6/18, pp. 5-10, 2018.
- [18] Rati Shukla, Vikash Yadav, Parashu Ram Pal and Pankaj Pathak, “Machine Learning Techniques for Detecting and Predicting Breast Cancer” *IJITEE*, ISSN: 2278-3075, Volume-8, pp. 2658-2662, 2019.
- [19] Rashmi G Saboji and Prem Kumar Ramesh, “A Scalable Solution for Heart Disease Prediction using Classification Mining Technique” *IEEE*, 978-1-5386-1887-5/17, pp. 1780-1785, 2017.
- [20] Anjan Nikhil Repaka, Sai Deepak Ravikanti and Ramya G Franklin, “Design And Implementing Heart Disease Prediction Using Naives Bayesian” *IEEE Xplore* Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8, pp. 292-297, 2019.
- [21] Lambodar Jena and Ramakrushna Swain, “Chronic Disease Risk Prediction using Distributed Machine Learning Classifiers” *IEEE*, 978-1-5386-2924-6/17, pp. 170-173, 2017.

Students Name

Signature

1. Yash J. Ingole
2. Dhawal P. Ramdham
3. Pranjali P. Hejib
4. Pratik S. Rohankar

**Signature of Guide**

Dr. R. R. Karwa