# Disease Prediction System Using Data Processing

**Ankush S. Narkhede[1], Akash V. Talele[2], Dhanashree M.Chim[3], Neha V. Varade[4], Apurva P. Kharche[5]**
[1, 2, 3, 4, 5] Dept of Computer Science and Engineering
[1, 2, 3, 4, 5] Padm.Dr. V.B.Kolte College of Engineering,Malkapur

***Abstract-*** *In recent years, the global healthcare landscape has witnessed significant challenges due to the rising burden of chronic diseases, epidemics, and pandemics. Despite advances in medical technology, early disease detection remains a critical issue. Traditional methods of diagnosis are often time-consuming, expensive, and sometimes prone to human error. There is need of disease prediction system for helping users. For avoiding wrong treatment, medical field can be more efficient. With the increasing volume of medical data, there is a pressing need for AI-based systems that can analyses patterns and predict diseases before they become critical. One of the important real-world medical problems is the detection of diseases at its early stage. In this study, systematic efforts are made in designing a system which results in the prediction of disease like disease. During this work, three machine learning classification algorithms are studied and evaluated on various measures. This system can be helpful for doctors to treat their patients properly with accuracy. In future, the designed system with the machine learning and data processing approach can be used to predict or diagnose other diseases.*

***Keywords****:* Data, Disease, Data processing etc.

## I. INTRODUCTION

There are various applications of data mining such as telecommunication industry, commercial industry, biological data analysis and many more. With the advance changes happening in the technology, especially in the field of health organization a lot of data is produced day by day. Since there is need of analysis of data and the amount of data analyzed is in large amount, so there is need of excessive knowledge regarding the technology of data mining. For health diagnosis E-healthcare applies data mining and telecommunication techniques are utilized. E-health was primarily used for patient data analysis and disease diagnosis at various levels. There are some patients who require continuous checkup and might need doctor help immediately.

In this project, we have conducted a comprehensive study of techniques used in multinational companies. Our approach is to provide a scheme that will be able to satisfy the users need and requirements. To achieve such condition the usability and security features must be balanced. Also, in this topic we present a method to predict disease based on the symptoms observed in patient. Once again, we verify that the id can be used in authentication with strong advantages to what concerns to security and without significant entropy to the users.

In future development we can also add challenge response interaction. In challenge response interactions, server will present a challenge to the client and the client need to give response according to the symptoms given. If the response is correct then prediction is granted.

Diseases was a common problem among adult's specifically middle-aged people but due to changing lifestyles disease affects children too. Disease is unpreventable because of the various external environmental stimulants which result in the destruction of body's insulin producing cells. However, changing lifestyle to achieve the required body weight and obtain the physical activities can help to prevent disease to enlarge. Disease is a chronic health problem with devastating, yet preventable consequences. Therefore, predicting disease manually sometimes seems not to be objective and it consumes a lot of time and cost. Disease treatment focuses on controlling to prevent various symptoms and complications through diet and exercise.

"To make software fast in processing, with good user interface so that user can change it and it should be used for a long time without error and maintenance". The objective of this project is making system of treating patient error free and it provide proper record. The objectives behind preparing this software are:

1.More effective to user.
2.Increase the security and improve clinic wok.
3.Easily identification of the disease.

## II. LITERATURE SURVEY

Literature review and project methodology is important in a system that wants to develop. In this chapter, all task and modules Diseases was a common problem among adult's specifically middle-aged people but due to changing

lifestyles disease affects children too. Disease is unpreventable because of the various external environmental stimulants which result in the destruction of body's insulin producing cells. However, changing lifestyle to achieve the required body weight and obtain the physical activities can help to prevent disease to enlarge.

Disease is a chronic health problem with devastating, yet preventable consequences. Therefore, predicting disease manually sometimes seems not to be objective and it consumes a lot of time and cost. Disease treatment focuses on controlling to prevent various symptoms and complications through diet and exercise. Data processing is a relatively new concept used for retrieving information from a large set of data. Mining means using available data and processing it in such a way that it is useful for decision-making. Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems. Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use. Data mining thus has evolved based on human needs which can help humans in identifying relationship patterns and forecasts based on pre-set rules and stipulations built into the program (Eapen, 2004). Data mining helps in pattern identification and categorizing data records by conducting cluster analysis, identification of odd records also called detecting anomalies and association rule mining or dependencies.

Frawley and Piatetsky (1996) describes data mining as the process of extracting implicit and previously undisclosed important information about data sets that can be used for effective decision-making. The process is termed as Knowledge Discovery in Database, Such discovered knowledge can be very useful in many areas of sciences, and health care is no different having a Knowledge Discovery in Database would help in predicting trends of many kinds of diseases and illness.

M. Durairaj, V. Ranjani presented a paper that aims to make a detailed study report of different types of data mining applications in the healthcare sector and to reduce the complexity of the study of the healthcare data transactions. Also presents a comparative study of different data mining applications, techniques and different methodologies applied for extracting knowledge from database generated in the healthcare industry. Finally, the existing data mining techniques with data mining algorithms and its application tools which are more valuable for healthcare services are discussed in detail [1].

Divya Tomar and Sonali Agarwal have presented a brief introduction of data mining techniques such as classification, clustering, association, regression in health domain and their advantages and disadvantages. This survey also highlights applications, challenges and future issues of Data Mining in healthcare [2].

R.Karthiyayini , J.Jayaprakash have presented a paper which analyses the various results generated by implementing the Apriori algorithm of Association technique. The focus of this paper is to provide precise information about chronic diseases for public [3].

Priyanka Vijay Pawar, Megha Sakharam Walunj, and Pallavi Chitte presents a methodology to predict diseases based on user input symptoms. They have built a prototype to demonstrate the efficiency of these methods which will inform users about the disease they are suffering from. It predicts probable diseases by mining data sets and provides suggested doctors and remedial solutions [4].

Gitanjali J, C.Ranichandra ,M.Pounambal has presented a method for identifying frequency of diseases in particular geographical location for a given period of time using Apriori data mining technique based on association rules is proposed[5].

Abdullah Saad Almalaise Alghamdi discussed about the importance of data mining using medical data then discussion of general data mining techniques has been presented. Furthermore, methodology describes the conceptual model for the extraction of rules on medical databases finally result can guide the relationship between the different attributes presented in the data. In this regard, they applied FP growth algorithm for extracting rules from the medical data [6].

## III. PROPOSED METHODOLOGYAND IMPLEMENTATION

The current system is a manual and file based one, we realize that system. We are going to build must give the solutions for wastage of time and space which affect the efficiency of the daily activities performed at the hospital. In previous system there is no location tracker for patient and doctors. There is no any feedback system in existing system for taking a feedback from patient. If the patient requires an instant diagnosis on their disease then they have to go doctor but it is not possible to everyone to identify disease at home instantly. Today's health prediction system is so much time consuming.

The system is designed to use intelligent data processing techniques to guess the most accurate illness based on patient's symptoms. If user's symptoms do not exactly match any disease in the database, then it is shows the diseases user could probably have based on his/her symptoms. It also consists of doctor address, contacts along with Feedback and administrator dashboard for system operations.

The purpose of developing this software is to overcome above mentioned problems. So, for solving the above problems we create workflow management for regulation of users activity. The health industry has been growing a lot from past few years .This technique has gained a lot of importance in medical areas. It has been calculated that a care hospital may generate five terabytes of data in the year. In our day to day life we have lot of other problems to deal with and we neglect our health problems. So in order to overcome such problem we have designed user friendly website which helps users to get diagnosed from their residence at any time. We also provide an option for booking an appointment with the doctor to discuss health related problems and get diagnosed properly.

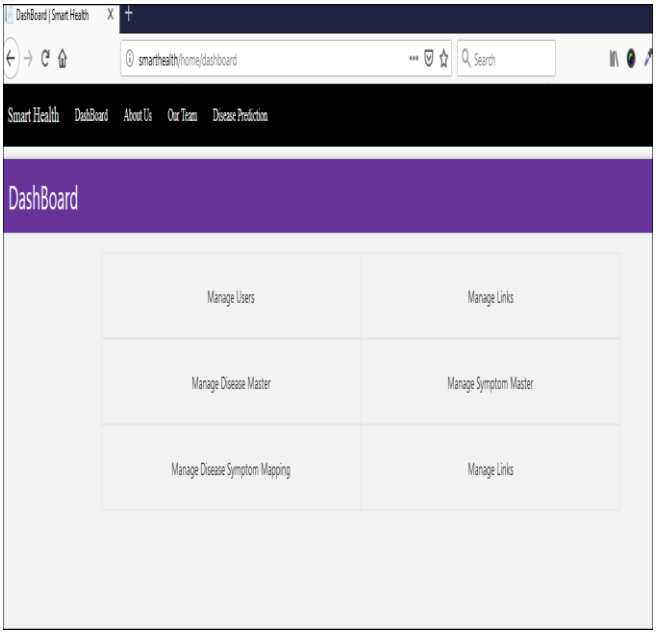
The proposed disease prediction system is built on cutting-edge machine learning algorithms, particularly deep learning. These algorithms process and analyze large datasets, including medical imaging, health records and patient-reported symptoms. The system incorporates decision support tools, predictive analytics, and real-time feedback to healthcare providers. Disease prediction systems using AI promise significant cost reductions in healthcare by enabling earlier diagnoses and personalized treatment plans. Early detection often leads to less invasive treatments, reducing the need for expensive emergency care and prolonged hospital stays.

The disease prediction system is currently at a Technology Readiness Level (TRL) with successful trials and integration in select healthcare environments. However, further improvements are needed, particularly in expanding the diversity of datasets to enhance accuracy across different populations and disease types. Future developments will also focus on integrating the system with existing hospital management software for seamless operation and enhancing explain ability to build trust among healthcare professionals. Implementing the AI system will streamline healthcare operations, particularly in diagnostics and patient management. It will reduce the time taken for initial diagnosis, allowing clinicians to focus on treatment.

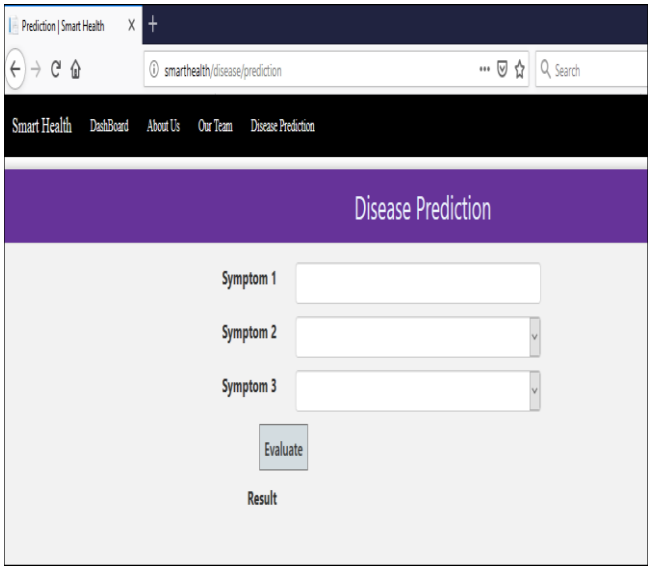
Initial trials of AI-driven disease prediction systems have shown promising results in various applications, from cancer detection to predicting cardiovascular risks. These systems rely on large datasets from medical records, patient history, imaging, and genetic data to train machine learning models capable of identifying early warning signs of diseases. Early validation has demonstrated high accuracy, sensitivity, and specificity in identifying conditions like diabetes, heart disease, and even certain cancers.

**1. Home page**

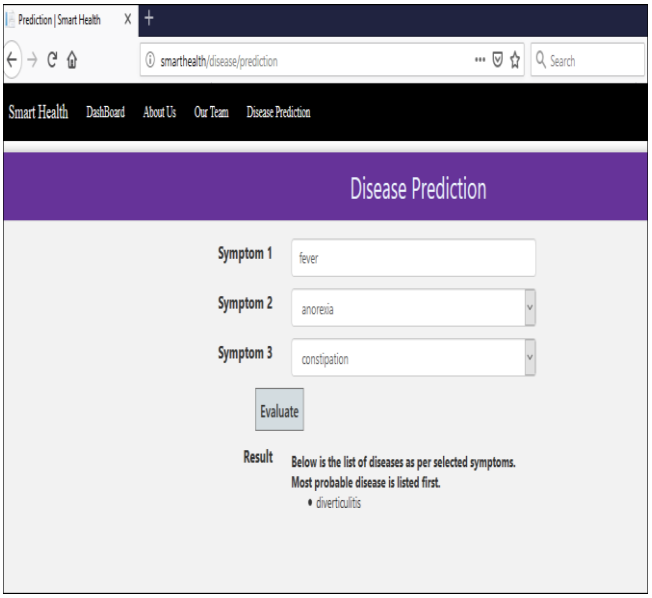

**2. Dashboard**



**3. Disease Prediction input form**




**4. Prediction Result**




Disease prediction system using data processing will be done in the following ways:

1. The admin will create disease master.
2. At the time of creating, the user will remember the symptoms mentioned in particular disease.
3. If symptoms are seen, then the disease is found. An Admin provides facility to generate the report of disease.
4. Users can be managed according to disease.
5. Finally, this software provides more secure way of handling diseases.

## IV. CONCLUSION

One of the important real-world medical problems is the detection of diseases at its early stage. In this study, systematic efforts are made in designing a system which results in the prediction of disease like disease. During this work, three machine learning classification algorithms are studied and evaluated on various measures. In future, the designed system with the used machine learning classification algorithms can be used to predict or diagnose other diseases. The work can be extended and improved for the automation of disease analysis including some other machine learning algorithms.

The system would drastically reduce the human effort, reduce the cost and time constraint in terms of human resources and expertise, and increase the diagnostic accuracy. The prediction of diseases using Data Mining applications is a challenging and risky task as the data found are noisy, irrelevant and massive too. In this scenario, data mining tools come in handy in exploring of knowledge of the medical data and it is quite interesting.

It is a famous and powerful technology which is of high interest in computer world. It is a sub field of computer science that uses already existing data in different databases to transform it into new researches and results. It makes use of Artificial Intelligence, machine learning and database management techniques to extract new patterns from large data sets and the knowledge associated with these patterns.

This is very cost effective software. It costs about 50000 thousand including system requirement and implementation charges. After doing trials with local hospital and common people, software undergoes these changes.

The ingenuity of the disease prediction system lies in its ability to aggregate diverse sources of medical data ranging from clinical records to genomics into actionable insights. By using advanced machine learning techniques, the system goes beyond simple data analysis to offer predictive insights that can anticipate disease progression or detect anomalies at an early stage. Furthermore, the ability to adapt and learn over time enhances its long-term value, creating a dynamic solution that evolves with emerging healthcare trends and medical research.

In future, we can utilize this software for veterinary and agriculture purpose also by modifying databases. This can be huge utility task regarding disease prediction which directly serves society.

## REFERENCES


[1] Nur Ghaniaviyanto Ramadhan Warih Maharani ,Adiwijaya ,(Member, Ieee), ,Andalfian Akbar Gozali, Department of Data Science, School of Computing, Telkom University, Bandung 40257, Indone "Chronic Diseases Prediction Using Machine Learning With Data Preprocessing Handling: A Critical Review"

[2] Holzinger, Andreas, Rocker, Carsten, Ziefle, Martina, "Smart Health", 2015

[3] Longhi, Sauro, Freddi, Allesandro,'" Human Monitoring, Smart Health And Assisted Living", 2017.

[4] Koehler, "Smart Health: What Today's Doctors Aren't Telling You", 2016

[5] Thareja, Reema, "Data Warehousing", 2009.

[6] Mannila, Heikki, Hand, David. J, Smyth, Padhraic, "Principles of Data Mining", 2001.

[7] Dean, Jared, "Big Data, Data Mining and Machine Learning", 2014.

[8] Aditya Tomar, " An Approach to Devise an Interactive Software Solution for Smart Health Prediction using Data Mining," Vol. 5, Issue 7, July 2016 .

[9] Vikramaditya R. Jakkula1, Diane J. Cook2, Gaurav Jain3 Prediction Models for a Smart Home based Health Care System WA 99164.

[10] Aakash Khatavkar1 Piyush Potpose2 Pankajkumar Pandey3" Smart Health Prediction System" Vol. 5, Issue 02, 2017.

[11] Prashant Tiwari1, Aman Jaiswal2, Narendra Vishwakarma3, Pushpanjali Patel4" SMART HEALTH CARE" Volume: 04 Issue: 04 | Apr -2017.

[12] Mohammed, Zaki. J, Meira, Wagner,"Data Mining and Analysis: Fundamental Concepts and Algorithm", 2014.

[13] Zhou, S.-M., Fernandez-Gutierrez, F., Kennedy, J., Cooksey, R., Atkinson, M., Denaxas, S., Siebert, S., Dixon, W.G., O'Neill, T.W. and Choy, E., "Defining disease phenotypes in primary care electronic health records by a machine learning approach: [14] A case study in identifying rheumatoid arthritis", PloS One, Vol. 11, No. 5, (2016), https://doi.org/10.1371/journal.pone.0154515 2. e0154515. Littell, C.L., "Innovation in medical technology: Reading the indicators", Health Affairs, Vol. 13, No. 3, (1994), 226-235. https://doi.org/10.1377/hlthaff.13.3.226 3. 4. 5. 6. 7. 8. 9.

[14] Milella, F., Minelli, E.A., Strozzi, F. and Croce, D., "Change and innovation in healthcare: Findings from literature", ClinicoEconomics and Outcomes Research, (2021), 395-408. doi: 10.2147/CEOR.S301169.