 A STUDY OF DIFFERENT DISEASE PREDICTION SYSTEMS IMPLEMENTED USING MACHINE LEARNING

**1Diya Pradhan, 2Bibidha Regmi,3Kouame Franck**

**4Sonali G. Karale**

*1,2,,3,Research Student, School of CS and IT , Jain (Deemed-to-be University), Bengaluru, India*

*4Assistant Professor, School of CS and IT, Jain (Deemed-to-be University), Bengaluru, India*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Abstract

*The Disease Prediction System is built on multiple prediction models that assist in predicting the user's disease depending on the symptoms that the user inputs as an input to the system. Predictive models utilize machine learning classification algorithms to examine the symptoms provided by the user and output the name and probability of the disease. The Naive Bayes Classifier, Decision Tree, and Random Forest Algorithms are used to predict disease. The Naive Bayes method aids in calculating the likelihood of the expected disease. The average predicted accuracy probability is 87%. The model employs a dataset including 132 symptoms from which the user can choose their symptoms.* *The user does not require a medical report to utilize this system because the forecast is based on the symptoms, saving money. The system also has an extremely simple user interface, allowing all users to utilize it to anticipate generic diseases.* *In this paper, we have studied some disease prediction systems implemented using machine learning and the disadvantage/limitations of those systems. We have found that the Bayes method is a suitable and accurate method for disease prediction.*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

# **Introduction**

There are instances when we need a doctor right away, but they are not always available for any reason, leaving us in danger. The solution we presented is user-friendly for receiving instant aid and guidance on health issues via the online healthcare system. Statistics and posterior distribution are now used to solve problems quickly and easily. As Bayesian statistics has a high success rate in the domains of economics, social science, and a few other fields, people have addressed several medical problems that are difficult to answer in conventional statistics by classification and can be solved easily in medical sectors.

Reverend Thomas Bayes established Naive Bayes as one of the fundamental common categorization strategies. The classification rules that aid in disease prediction are generated by samples that have been taught by themselves and aid in the solution of the problem. It is estimated that more than 70% of individuals in India are susceptible to various body diseases such as viruses, flu, cough, colds, and so on every two months. Because many people do not understand that general body diseases can be symptoms of something more dangerous, 25% of the population dies or suffers from serious medical problems as a result of ignoring early general body symptoms. This is a very serious situation that we are facing, and the problem can be proven to be a very dangerous situation for the population and can be alarming if people continue to ignore these diseases. As a result, diagnosing or anticipating the disease at an early stage is critical to avoiding difficulties and deaths.

When it comes to generalized disease, the currently available systems are either dedicated to a specific disease or are under development or study for solving the algorithms connected to the problem. The suggested system's major goal is to forecast regularly occurring diseases in their early stages because if they are not evaluated or inspected, they can progress into a more deadly diseases and even cause death.

In this paper, we have studied different algorithms and website applications which are mostly get used for disease prediction. Our study is based on the techniques, result/accuracy and limitations of these algorithms/websites.

1. **Literature review**

The study for the prediction of disease using machine learning was performed by Dr. C K Gomathy, and Mr. A. Rohith Naidu [1]. Most chronic diseases are predicted by their system. It accepts the structured type of data as input to the machine learning model. This system is used by end-users i.e. patients/any user. In this system, the user will enter all the symptoms from which he or she is suffering. Algorithms are then applied to which gives the best accuracy. Then System will be predicting disease on the basis of symptoms. This system uses Machine Learning Technology. The Naïve Bayes algorithm is used for predicting the disease by using symptoms, for classification KNN algorithm is used, Logistic regression is used for extracting features which are having most impact value, and the Decision tree is used to divide the big dataset into smaller parts. Finally, the disease will be predicted by this system.

Palle Pramod Reddy, Dirisinala Madhu Babu, Hardeep Kumar, and Dr. Shivi Sharma conducted a disease prediction study [2] in which they mixed structured and unstructured health data to determine disease risk. Reconstruct missing data in medical records obtained from online sources using latent factor models. We may also use statistical information to assess the most important chronic diseases in specific regions and populations. They consulted with hospital experts to learn more about useful functions when working with structured data. For unstructured text files, they use a random forest algorithm to automatically select features. Data collection is done from the internet to identify the disease here, i.e. collecting the actual symptoms of the disease. No dummy value was entered. The symptoms of the disease were collected from different health-related websites.

MIN CHEN, proposed [3], 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. Machine learning techniques can use the raw data for the learning process and based on that learning they can predict the disease.There are two stages to the machine learning algorithm: 1) Testing and 2) training. Machine learning technology is still working through issues from decades past when attempting to anticipate the disease from a patient's symptoms and from the patient's history. Machine learning technology can be effectively used to solve healthcare challenges. They are using all available machine learning techniques to monitor patient health. This proposed system had an accuracy of 94.8%.

Ankita Dewan, recommended a disease prediction system [4] that uses data mining classification hybrid technique for predicting heart disease. Data mining is a method of discovering knowledge in which information is acquired by looking at data that may be concealed in very big sources, these sources are then studied from many angles using various approaches, and the retrieved information is then summarized into useable information. Data mining can be divided into two basic categories: supervised and unsupervised. Both methods are effective at analyzing and predicting diseases, but they have different applications. These methods are employed in the medical field in accordance with disease prediction and patient treatment decision-making. This system is using techniques like Neural Network, Decision Tree, and Naive Bayes. The accuracy of this system is 87%.

The study [5] for the best medical diagnosis mining technique was performed by K.M. Al-Aidaroos, A.A. Bakar, and Z. Othman. For this study, the authors compared Nave Baeyes to five other classifiers: LR, KStar (K\*), Decision Tree (DT), Neural Network (NN), and a basic rule-based algorithm (ZeroR). The efficiency of all algorithms was evaluated using 15 real-world medical problems from the UCI machine learning repository (Asuncion and Newman, 2007). In the experiment, NB outperformed the other algorithms in 8 of the 15 data sets, leading to the conclusion that the predictive accuracy results in Nave Baeyes are superior to other techniques.

A research paper [6] was written by Jyoti Soni, Ujma Ansari, Dipesh Sharma, and SunitaSoni provide a survey of existing techniques of information discovery in databases using data mining techniques that are used in today's medical research, specifically in Heart Disease Prediction. A number of experiments have been carried out to compare the performance of predictive data mining techniques on the same dataset, and the results show that Decision Tree outperforms, with Bayesian classification having comparable accuracy to Decision Tree in some cases, but other predictive approaches such as KNN, Neural Networks, and Classification based on Clustering underperform.

1. **SOME EXISTING WEBSITES**

**1. SickPredict**



Sickpredict [8] was inspired by COVID, but it is not COVID software. The software warns you that you're getting sick 24 to 48 hours before you notice any symptoms by analyzing data from your Apple Watch and iPhone. Although it is now only available for iOS, the creator is working on further versions that will allow the app to pair with other apps. The software assigns users a daily ill number from 1 to 10 on a scale of 1 to 10.

The higher the number, the more likely you are to catch a cold. More information, such as temperature and blood oxygen levels, can be entered to give the app more accurate readings. If you are in the upper range, you should consult your doctor.

sounds like magic, but it's more about science and math! Every day, your wearable devices track over 250,000 health and fitness metrics.

These metrics include but are not limited to: activity energy, body temperature, heart rate, heart rate variability, and sleep data. SickPredict uses its revolutionary machine learning algorithm to analyze this data and provide you with a daily SickNumber to determine how likely you are to be sick.

Health Logic Global, LLC, developers of the SickPredict app, were inspired by these difficult times to do something different. They were determined to find a more proactive way to determine if someone was sick before their temperature and other symptoms developed, which meant it was too late.

Over the past few years, several reputable universities and research institutes, including Stanford, Harvard, MIT, and Scripps, have conducted extensive research to test wearable devices that monitor and track health statistics to predict disease susceptibility.

More recently, other studies have shown that the accuracy of health data obtained from wearable devices (such as Apple Watch, Fitbit, Garmin, Oura, etc.) in predicting disease is very positive. The

SickPredict app uses this research and calculates your illness number in real time to help you predict any upcoming changes in your health before symptoms appear!

This is a revolutionary concept. Imagine the potential impact of this technology on reducing the spread of disease. Not only does SickPredict tell us when we're sick, but it also shows us important health information right from our Apple Watch, all in one place and presented in easy-to-read graphs.

Limitations of this software:

1. It is only available in IOS.

ii) It can only be used by users who have an Apple watch.

iii) They don’t refer doctors nearby.

**2. Medipredict**



Medipredict [9] provides a comprehensive assessment of health status by integrating data from multiple sources, including detailed physical examinations, medical imaging, microbiome analysis, metabolomics, genomics, and medical history processing.

It aims to help people by predicting disease risk, providing early diagnosis, and providing actionable insights to maintain optimal health using cutting-edge technologies and evidence-based science.

The steps in this Medipredict are:

Step 1: Registration and health questionnaire

Step 2: Consultation with the medical team

Step 3: Comprehensive examination

Step 4: Data generation and analysis

Step 5: Health report and consultation

Step 6: Follow-up and health coaching

Registering with Medipredict involves signing various legal documents regarding our services. All information is encrypted, stored securely, and processed privately.

They then send us detailed questionnaires about our health, focusing on all areas of our lives, to give us an overview of our current body condition. This is the first step to achieving optimal health with the help of media predict. An online medical consultation is scheduled with your doctor to discuss your current health and past medical history and prepare for your visit.

Also, let us know if you have any questions or special requests. Based on this conversation, they individually provide a step-by-step overview of the entire process.

Limitations of this software:

1. It is little complex to understand.

ii) People needs guidance to use this software. Especially old people.

iii) Even though it gives somewhat accurate result, it is a long process and consumes time.

1. **PredictionHealth**

****

PredictionHealth[10] is here to help you because you went into healthcare to benefit people. It automates documentation and brings machine learning into daily clinical practice.

Clinicians and administrators in healthcare spend an unusual amount of time collecting and interpreting information from patient charts. It is also overly tough to extract insights from that data as efficiently as it should.

Their tools relieve physicians of the documentation load and administrative stresses from management, allowing everyone to focus on what they do best: caring for the patient. By employing AI, PredictionHealth enables each therapist to obtain insight into the strengths and weaknesses of every single note. This tool can assist save staff time and avoid denials and penalties by identifying billing problems and unnecessary mistakes.

The PredictionHealth system is a project for end-user support and online consultation. This method enables customers to receive immediate assistance for their health difficulties via an intelligent health care system online.

The system stores information about numerous symptoms as well as the disease/illness linked with those symptoms.

Limitations of this software:

1. You need to wait until the person contacts you.

ii) It is not instant. So the patient has to wait even if it is an emergency.

1. **Florence**



The chatbot [7] is basically a personal nurse, and works on Facebook Messenger, Skype or Kik. “She” can remind patients to take their pills, which might be a handy feature for older patients. You just write the name of the medicine in chat, the number of times a day you must take it and at what time. Then, Florence sends you a message in chat every time you must take the pill. The proposed idea of this is to make it easier for people to check on their health as compared to the conventional way of standing in a queue for hours before they could get their medication done. In order to build a chatbot, this research aims to apply the use of the RASA framework. As any person, the chatbot can connect with people and take on the user’s symptoms. It will then identify the most likely disease and predict it along with the treatment recommended. This will help people get a quick answer to all their queries without any hassle. Moreover, Florence can track the user’s health, for example, body weight, mood or period, helping them to reach their goals. The chatbot also has the skills to find the nearest pharmacy or doctor’s office in case you need it. It keeps track of your health and helps you to reach your goals. For example, you can track your body weight, mood or your period with Florence.

Limitation of this software:

1. It has limited responses.
2. They're not often able to answer multi-part questions or questions that require decisions
3. **Infermedica**



Infermedica [11] assists machine learning technology to power the symptom-checker chatbot, Symptomate. The platform runs online and on mobile phones as a chatbot or voice-based application. It assesses the user’s health status and based on the symptoms, it sets up a possible diagnosis and gives actionable recommendations. The company uses artificial intelligence and machine learning to assess symptoms and find patterns in data. Thanks to algorithms, Infermedica gets smarter over time. Additionally, the team of physicians verifies every piece of information that is added to the medical database to ensure that patients get safe and reliable recommendations. To date, physicians involved in the project have spent over 30,000 hours on reviews.

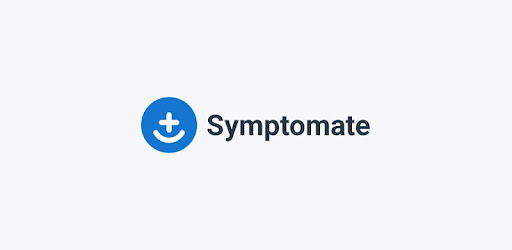
Infermedica claims to have more than 8 million health checks to date, supports 19 languages in their products. During the pandemic’s first year, the company launched a COVID-19 risk assessment tool, used by over half a million patients and implemented by 300 organizations, including two national governments. Over the last year, Infermedica grew into a team of over 100 people, and its products were translated into 17 languages, making them available in countries such as China and the United Arab Emirates. The latest investment round of $10.25M will help to expand market presence in the United States and Europe, and accelerate the development of the technology.

Infermedica was founded with the goal of transforming healthcare outcomes by making primary care accessible to all, and throughout 2022 we made significant progress towards achieving that goal. It has also celebrated the company's 10th anniversary and reached a milestone of conducting more than 12 million successful health checks for patients around the world to support better treatment.

Limitation of this software:

i) It has same answer for most of the Query

1. **Symptomate**



Symptomate [12] is a registered Class I medical device in the European Union. Symptomate is a medical device regulated by the FDA as a general wellness product in the US. It is not yet a licensed medical device in other countries, in particular in Canada nor Australia. For more information about territorial use, please see our Terms of Service. Symptomate does not provide a medical diagnosis, and should not replace the judgment of a licensed healthcare practitioner. It provides information to help guide your decision making based on readily available information about symptoms. If you have any questions or concerns about the output of Symptomate, consult your healthcare practitioner. Symptomate is an innovative tool that will help you to assess your symptoms. It's smart, anonymous and completely free. Symptom checkers provide individuals with a fast way to identify possible causes for their medical symptoms. Users enter their symptoms and respond to targeted follow-up questions that help narrow the possible causes. Symptomate provides you with a fast assessment of your symptoms.

Step to get fast assessment of your symptoms:

Step 1: Enter your symptoms

Step 2: Answer some simple questions

Step 3: Done! Your assessment will reveal:

* the possible causes for your symptoms
* a summary of present and absent symptoms
* a triage recommendation

Limitation of this software:

i) Doesn’t provide accurate result

ii) It doesn’t refer to any doctor or suggest any medicine

1. **Buoy Health Symptoms Checker**

****

Buoy Health [13] is a self-triage tool that patients can access for free online. When using the tool, patients type in their symptoms and answer questions the AI assistant asks. The tool then provides options for patients on how to seek care, including telehealth or virtual cares. Buoy Health is an example of the growing popularity of self-triage tools. This one has a machine learning algorithm that takes online symptom checking to the next level. The tool's machine learning algorithm learns from each patient encounter to improve on future encounters. Self-triage and symptom checker tools have been around for decades, but they are often built on rudimentary flow charts that operate more like a "choose-your-own-adventure novel".

Limitation:

1. The user can only make use of the website to check symptoms
2. Diagnostic accuracy of 52%.
3. Android and iOS mobile application are not available
4. Provided list of questions for symptoms checking is very long thus it can discourage the user to complete the check up
5. There is no user profile available on the website to track health record

Techniques:

The website works much like a text message chat between two people. The founder Andrew Le and his developer team made the symptoms checker more interactive between the user and the website. It has no Chabot integrated, instead the user must answer a very long list of questions from predefined symptoms available on the website.

1. **Isabelhealthcare**

****

Isabel Healthcare [14] is a global brand providing professional clinical reasoning tools for doctors and the acclaimed Isabel Symptom Checker for patients. But the company actually started out as a small charity. And the inspiration behind that charity, and indeed the company’s name, was a 3 year-old girl called Isabel.

Features:

Unlike any other symptom checker, Isabel is unique in being able to handle even large and complex medical presentations using multiple AI and enhanced machine learning algorithms while remaining quick and easy to use. It’s been crafted, tuned, tested and used by doctors around the world for almost 20 years.

Core Functionality: Utilization management, Clinical workflow, Health and wellness management, Screening/ assessment tools

Target Populations: Adolescents/ children, Elderly, Frequent emergency department users, Uninsured individuals, Medicaid beneficiaries

Users: Family/ caregivers, Patients/ consumers, Medical providers, Peers and community health workers

Limitation:

1. Diagnostic accuracy of 68%
2. Mobile application is not available to the user
3. It can only be accessed on the browser

Technique:

Isabel Symptom Checker is based on Isabel’s curated machine learning and natural language processing, artificial intelligence (AI) clinical engine — covering 6,000 conditions and all specialties.

Isabel Symptom Checker enables patients to enter symptoms in everyday language and helps them to both understand their symptoms and get appropriate care. Symptoms are then analyzed by the Isabel clinical engine and possible conditions are presented along with additional information so that patients can research their symptoms further. A ‘Where now?’ triage feature is also available to help patients decide where to seek care.

1. **WebMD**

**C:\Users\franc\OneDrive\Pictures\Screenshots\Screenshot (83).png**

WebMD [15] is an American corporation known primarily as an online publisher of news and information pertaining to human health and well-being. The site includes information pertaining to drugs. It is one of the top healthcare websites. It was founded in 1998 by internet entrepreneur Jeff Arnold. URAC, the Utilization Review Accreditation Commission, has accredited WebMD's operations continuously since 2001 regarding everything from proper disclosures and health content to security and privacy. WebMD offers services to physicians and private clients.

Features:

Key Features: Symptom Checker – Choose your symptoms, learn about potential conditions or issues, and lookup treatment and care options. Symptom Tracker – Track ongoing symptoms & conditions over time. Allergy Tracker - Get daily alerts when allergy levels are high in your area.

This symptom checker now includes the ability to select symptoms by body location. The tool also allows the user to select multiple symptoms quickly.

Limitation:

1. The website doesn’t provide medical advice
2. Diagnostic accuracy of 59%.
3. It has no mobile application available
4. The user has no option available on the website to create a profile for record tracking

Technique:

When a body location is selected, the “most common symptoms” are displayed first, but the user can also switch tabs to see “All” symptoms. The user can also use the category-specific search box to search for all symptoms in that category.

1. **CONCLUSION**

A literature review shows that this system provides accurate results. We are working with large data sets that ensures better performance than before. A symptom checker using artificial intelligence can provide diagnostic and triage recommendations with a level of accuracy and safety close to that of a human physician. Such a system has the potential to reduce costs and improve access to health care worldwide, but achieving this requires a higher level of trust from the medical community and the general public. Key to this trust is a better understanding of the strengths and weaknesses of human physicians, who may not always agree on the causes of a patient's symptoms or the most appropriate triage outcome, and the accuracy of AI-powered and safety-aware systems. This system performs all the necessary steps using algorithms and makes it user-friendly. Users can get relevant responses displayed on the Android app and refer to it for analysis.

**REFERENCE**

1. Dr C K Gomathy, Mr. A. Rohith Naidu (2021). THE PREDICTION OF DISEASE USING MACHINE LEARNING (IJSREM).
2. Palle Pramod Reddy, Dirisinala Madhu Babu, Hardeep Kumar and Dr.Shivi Sharma (2021). Disease Prediction using Machine Learning (IJCRT).
3. Dong Jin Park, Min Woo Park, Homin Lee, Young-Jin Kim, Yeongsic Kim & Young Hoon Park (2021). Development of machine learning model for diagnostic disease prediction based on laboratory tests.
4. Muhammad Nabeel,Shumaila Majeed,Mazhar Javed Awan,Hooria Muslih-Ud-Din (2021). Review on Effective Disease Prediction through Data Mining Techniques.
5. Al-Aidaroos, K., Bakar, A., & Othman, Z. (2012). Medical Data Classification with Naive Bayes Approach. Information Technology Journal.
6. Jyoti Soni, Ansari, U., Sharma, D., & Soni, S. (2011). Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction.
7. J. Gupta, V. Singh and I. Kumar, "Florence- A Health Care Chatbot," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2021, pp. 504-508, doi: 10.1109/ICACCS51430.2021.9442006.
8. SickPredict: <https://sickpredictapp.com/> [Accessed on 20.02.2023]
9. Medipredict: <https://medipredict.com/> [Accessed on 20.02.2023]
10. Predict Health: <https://www.predictionhealth.com/> [Accessed on 20.02.2023]
11. Infermedica: <https://infermedica.com/> [Accessed on 20.02.2023]
12. Symptomate: <https://symptomate.com/> [Accessed on 20.02.2023]
13. Buoy Health Checker: <https://www.buoyhealth.com/symptom-checker/> [Accessed on 20.02.2023]
14. Isabel Health: <https://www.isabelhealthcare.com/> [Accessed on 20.02.2023]
15. WebMD: <https://www.webmd.com/> [Accessed on 20.02.2023]