IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 8, August 2023

DOI: 10.17148/IARJSET.2023.10830

SYMPTOMS-BASED DISEASE PREDICTION USING MACHINE LEARNING

Shreyash Kashyap[1], Dr. Ramesh Chundi[2]

Student (MCA), School of CSA, REVA University, Bangalore, India[1]

Assistant Professor, School of CSA, REVA University, Bangalore, India[2]

Abstract: The requirement for health information is altering knowledge-seeking behaviour, which should be noted globally. Many of us struggle with finding health information online about illnesses, diagnoses, and various treatments. It will save a lot of time if a suggestion system is frequently used for physicians and medications. Because the users of a system like this are laypeople, the user has several difficulties understanding the basic medical terminology. The customer is perplexed since there is an excessive amount of medical information available in various formats.

Keywords: Random Forest Algorithm, Naive Bayes, Support Vector Machine, Logistic regression and etc.

I. INTRODUCTION

A method called Disease Prediction using Machine Learning forecasts the disease based on the data the user provides. Additionally, it accurately forecasts the user's or the patient's disease based on the data or symptoms entered into the system and returns findings accordingly. If the customer simply wants to know the type of ailment the patient has experienced and the condition is not particularly significant. It is a system that gives users advice on how to keep their health systems in good shape and offers a technique to identify diseases using this prediction. Nowadays, the health industry plays a significant role in treating patients' illnesses, so this is frequently helpful for the industry to inform the user as well as helpful for the user in the event that he or she doesn't want to visit the hospital or other clinics. By entering the symptoms and all other relevant information, the user can understand the disease they are affected by, and as a result, the health industry can also benefit from this. Programming computers to perform better using example data or historical data is known as machine learning. The study of computer systems that learn from data and experience is known as machine learning. The training and testing tracks of a machine learning algorithm. illness diagnosis based on the symptoms and medical history of the patient Machine learning technology has improved throughout the years. The medical profession now has an incomparable platform thanks to machine learning technology, making it possible to handle healthcare challenges quickly. Machine learning is being used to keep entire hospital data. With the aid of machine learning technology, which enables constructing models to analyse data quickly and give answers more quickly, clinicians may more accurately diagnose and treat patients, which improves patient outcomes. The use of machine learning in the medical industry is best illustrated by the example of healthcare. The current work on unstructured and textual data will be done in order to increase the accuracy from enormous data. The current methods for disease prediction include linear, KNN, decision trees, logistic regression, and the Adaboost algorithm.

1.1. STATEMENT OF THE PROBLEM

Developing an accurate and reliable machine learning model for symptoms-based disease prediction is essential to enhance early diagnosis and treatment. This project aims to create a predictive model that can effectively analyze a patient's reported symptoms and provide accurate insights into potential diseases or medical conditions. By leveraging a dataset containing symptoms and corresponding diagnoses, the goal is to design a model that aids healthcare professionals in making informed decisions, improving patient outcomes, and optimizing healthcare resources.

1.2. BRIEF DESCRIPTION OF THE PROJECT

The project involves building a machine learning system that predicts diseases based on reported symptoms. By training on a dataset of symptoms and corresponding diagnoses, the system learns patterns in the data. When given a set of symptoms, it can then suggest potential diseases.

II. RELATED WORKS

The literature survey undertaken for this project, focusing on symptoms-based disease prediction using machine learning, involves a comprehensive exploration of existing research, methodologies, and advancements in the field to establish a well-rounded understanding of the current state of predictive models grounded in reported symptoms.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 8, August 2023

DOI: 10.17148/IARJSET.2023.10830

[1] Heart Disease:
In the context of symptoms-based disease prediction using machine learning, a specific emphasis is placed on the intricate interplay between reported symptoms and the accurate identification of heart diseases, reflecting a pivotal application area with significant potential for improving early detection and patient care.

[2] Diabetes Disease:
Within the realm of utilizing machine learning for symptoms-based disease prediction, a significant spotlight is directed toward the intricate dynamics of reported symptoms in the context of diabetes, underscoring the system's potential to advance timely detection and personalized management strategies for this prevalent metabolic disorder.

[3] Liver Disease:
In the domain of machine learning-driven symptoms-based disease prediction, a focal point emerges around the intricate correlation between reported symptoms and the accurate identification of liver diseases, presenting a critical avenue for enhancing early diagnosis and tailored interventions in this specific medical domain.

[4] Dengue Disease:
Within the framework of machine learning-enabled symptoms-based disease prediction, particular attention is directed toward unraveling the complex association between reported symptoms and the precise detection of dengue disease, offering a pivotal avenue for early intervention and targeted management strategies in the context of this mosquito-borne viral infection.

[5] Migraine Disease:
Amidst the realm of utilizing machine learning for symptoms-based disease prediction, a distinct focus centers on unraveling the intricate interplay between reported symptoms and the accurate identification of migraine disease, signifying a crucial avenue for enhancing early diagnosis and personalized treatment strategies in the context of this neurological disorder.

[6] Skin Disease:
In the landscape of harnessing machine learning for symptoms-based disease prediction, a specific spotlight is cast on deciphering the intricate nexus between reported symptoms and the precise identification of skin diseases, underscoring a pivotal pathway for advancing early diagnosis and tailored interventions in the realm of dermatological health.

[7] Joint-Pain:
Within the context of employing machine learning for symptoms-based disease prediction, a particular emphasis revolves around unraveling the intricate interplay between reported symptoms and the accurate identification of joint-pain diseases, illuminating a crucial avenue for enhancing early diagnosis and personalized interventions in the domain of musculoskeletal health.

III. METHODOLOGY

1. Data Collection:
Gather a comprehensive dataset that includes information about various diseases, their corresponding symptoms, and relevant patient data. You might obtain this data from medical databases, research papers, or healthcare institutions.

2. Data Preprocessing:
Clean and preprocess the data to ensure its quality. This involves handling missing values, removing duplicates, and converting categorical data into numerical representations. Symptoms might be in text format, so you'll need to process and encode them appropriately.

3. Feature Engineering:
Create relevant features from the data that can help the machine learning model make accurate predictions. For symptom-based disease prediction, these features might involve the presence or absence of specific symptoms.

4. Data Labeling:
Label the data with the correct disease outcomes. Each instance in the dataset should be associated with the disease it represents.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 8, August 2023

DOI: 10.17148/IARJSET.2023.10830

5. Data Splitting:
Divide the dataset into training, validation, and testing sets. A common split might be 70% for training, 15% for validation, and 15% for testing.

6. Model Selection:
Choose an appropriate machine learning algorithm for the task. Classification algorithms like Random Forest, Support Vector Machines, and Neural Networks are commonly used for disease prediction.

7. Model Training:
Train the selected model using the training dataset. The model will learn the patterns between symptoms and diseases during this phase.

8. Hyperparameter Tuning:
Optimize the hyperparameters of the chosen model using the validation dataset. This step helps improve the model's performance.

9. Model Evaluation:
Evaluate the trained model's performance using the testing dataset. Common evaluation metrics for classification tasks include accuracy, precision, recall, F1-score, and area under the ROC curve.

10. Interpretability (Optional):
Depending on the algorithm used, consider methods to interpret the model's decisions. This is crucial in a medical context to understand why the model is making certain predictions.

11. Deployment:
Once you're satisfied with the model's performance, deploy it to a real-world setting. This could be an application or platform that allows users to input symptoms and receive predictions about potential diseases.

12. Monitoring and Maintenance:
Continuously monitor the model's performance in the real world. As new data becomes available, retrain and update the model periodically to ensure it remains accurate.

IV. PROPOSED SYSTEM

This method is used to predict sickness using symptoms. This approach uses a decision tree classifier to assess the model. End users actively engage with this system, utilizing its functionalities to streamline their interactions and make informed decisions within their respective contexts. The technology will be able to predict sickness based on symptoms. This system makes use of machine learning technologies. To predict diseases, researchers employ the decision tree classifier approach. We refer to this technology as AI Therapist. This system is intended for those who are always concerned about their health, therefore we have added certain components that identify them and elevate their mood as well. As a consequence, the health awareness feature "Disease Predictor" may recognise illnesses based on their symptoms.

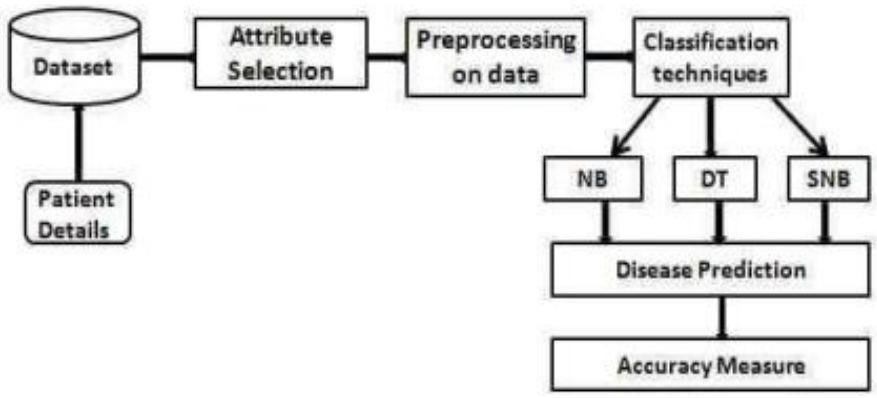

Fig.1: Diagram of Proposed System.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 8, August 2023

DOI: 10.17148/IARJSET.2023.10830

V. FEATURES SELECTION

Everyone wants to have a decent healthcare system, and doctors are expected to be knowledgeable and kind in all situations. However, it is very improbable to review all the information, medical history, and documents required for each circumstance. Even if they have a vast amount of data and information, comparing and analysing all of the illnesses' symptoms in order to predict their outcomes is delicate. Therefore, incorporating facts into a case's unique profile and doing in-depth research are beyond of the purview of a croaker. As a result, the outcome is always described as a supported healthcare plan that was simply written for an existing. By evaluating actual data, prophetic analytics is the process of making predictions about the future.

Making the fashionable viewpoints available for health care would be possible in any situation. In predictive modelling, a vatic country is created using previously recorded data, trends, individualities, and situations. The model is set up such that a new existence can obtain a vatic nation without interruption. These predictive algorithms may be directly used by health and Medicare departments to determine when a case can be discharged without risk.

VI. RESULT

In a real project, you would present your results in a clear and organized manner. Here's how you might structure it:

Dataset Description:
Briefly introduce the dataset used for training and testing. Include the number of instances, features (symptoms), and the range of diseases covered.

Model Performance Metrics:
Display key performance metrics that assess the model's effectiveness in predicting diseases based on symptoms. Common metrics include accuracy, precision, recall, and F1-score.

Confusion Matrix:
Provide a confusion matrix to visually represent the model's predictions against actual outcomes. This matrix can help understand where the model excels and where it might struggle.

Receiver Operating Characteristic (ROC) Curve:
If applicable, show an ROC curve to illustrate the model's trade-off between true positive rate and false positive rate across different threshold values.

Feature Importance:
Discuss which symptoms or features played a crucial role in the model's predictions. This can provide insights into the relationships between symptoms and diseases.

Comparison to Baselines:
If you have baseline models or rule-based approaches, compare your machine learning model's performance against these benchmarks to demonstrate its added value.

Limitations:
Highlight any limitations of the model, such as data imbalance, potential biases, or specific diseases that the model struggles to predict accurately.

Future Directions:
Suggest potential improvements or future enhancements to the model. This could include using more data, incorporating additional features, or exploring advanced algorithms.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 8, August 2023

DOI: 10.17148/IARJSET.2023.10830

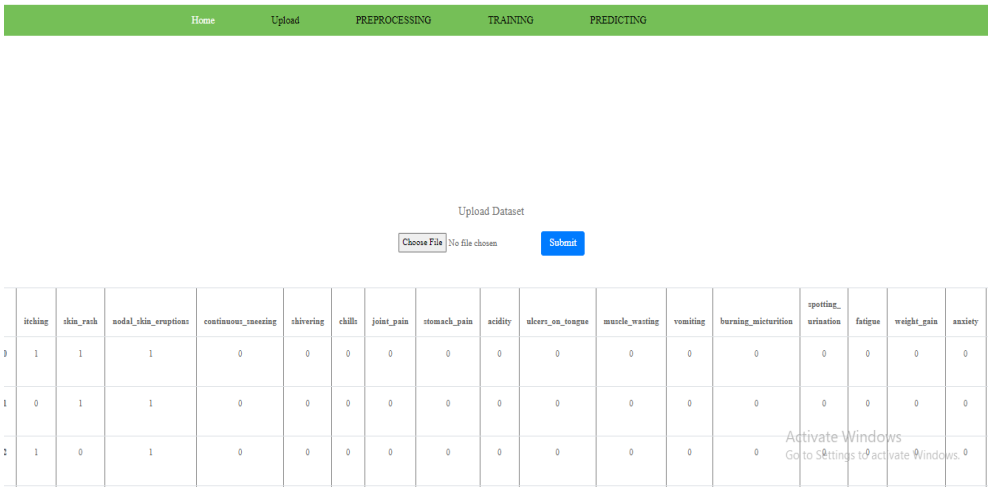

Fig.2: Datasets being used

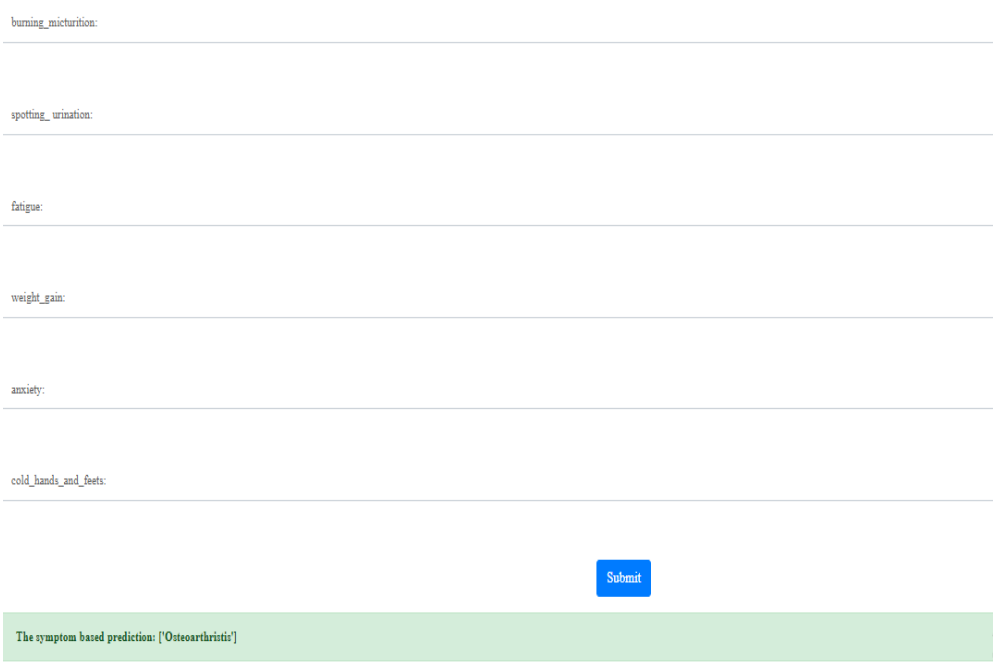

Fig.3: Figure showing results predicted

VII. CONCLUSION

In conclusion, my exploration leads me to the realization that machine learning holds the potential to revolutionize health monitoring, offering an efficient means to track our well-being. This approach allows us to conveniently monitor our health at regular intervals, without incurring any costs, ultimately contributing to maintaining our overall well-being. Following the development of the machine learning model, I implemented it using Flask, a Python web framework. As we progress, there's the prospect of transforming this domain into a website, rendering the resource freely accessible to a global audience. For users seeking personalized health insights, our algorithm only requires them to visit the relevant page and select 5 to 8 illnesses. Subsequently, the generated forecast will furnish valuable health insights, empowering users to take informed actions, including seeking medical attention if necessary. The reach of this endeavor extends to every individual across the planet, presenting a platform for positive health outcomes.

DOI: 10.17148/IARJSET.2023.10830

REFERENCES

[1]. Marshland, S. (2009) Machine Learnings an Algorithmic Perspectives. CRC Press, New Zealand, 6-7.
[2]. https://www.ijercse.com/viewabstract.php?id=16381&volume=Volume10&issue=Issue7 Author : Shreyash Kashyap 1 (2023) Prediction of Heart Disease Using Machine Learning (IJERCSE)
[3]. Fathima, A.S. and Manimeglai, D. (2016) Predictive Analysis for the Arbovirus Dengue using SVM Classifications.(IJET), 2, 521-527.
[4]. Iyer, A., Jeyalatha, S. and Sumbaly, R. (2015) Diagnosis of Diabetes Using Classification Mining Technique.(IJDKP), 5, 1-14.
[5]. Sarwar, A. and Sharma, V. (2012) Intelligent Naive Bayes Approaches to Diagnose Diabetes Type-2. Special Issues of International Journal of Computer Application (0975-8887) on Issues and Challenges in Networking, Intelligences and Computing Technologies-ICNICT 2012, 3, 14-16.
[6]. Vijayarani, S. and Dhayanand, S. (2015) Liver Diseases Prediction using SVM and Naive Bayes Algorithm. International Journals of Science, Engineering and Technology Researches (IJSETR), 4, 816-820.
[7]. Tarmizi, N.D.A et al., (2018) Malaysia Dengue Outbreaks Detection Using Data Mining Model.(JNIT), 4, 96-107.
[8]. Otoom et al., (2015) Effective Diagnosis and Monitoring of Heart Diseases. (IJSEA). 9, 143-156
[9]. Rambhajani, M et al., (2015) A Survey on Implementations of Machine Learning Technique for Dermatology Disease Classifications. International Journal of Advance in Engineering & Technology, 8, 194-195.

