Vivekanand Education Society's Institute of Technology

Department of Computer Engineering

Group No.: 23
Date :- 08/07/2024

Project Synopsis Template (2024-25) - Sem VII

Health+ : “Your Digital Health Guardian”

Dr. Dashrath Mane
Assistant Professor, Department of Computer Engineering

Vidisha Jadhwani
V.E.S.I.T
2021.vidisha.jadhwani@ves.ac.in

Riddhi Labde
V.E.S.I.T
2021.riddhi.labde@ves.ac.in

Priti Shamnani
V.E.S.I.T
2021.priti.shamnani@ves.ac.in

Nikhil Makhija
V.E.S.I.T
2021.nikhil.makhija@ves.ac.in

Abstract

Leveraging machine learning, our proposed disease prediction system aims to tackle the inefficiencies and challenges associated with hospital visits for minor health issues. Traditional visits can be time-consuming and costly, while managing appointment calls adds to the administrative burden on both patients and healthcare providers. Over the past decade, there has been a significant rise in the demand for disease prediction tools. This surge is driven by the increasing prevalence of various diseases and the growing disparity in the doctor-to-patient ratio. Our system is designed to address these issues by providing users with immediate and accurate health guidance based on the symptoms they report. Utilizing advanced machine learning algorithms, the system will cross-reference user symptoms with a comprehensive database and continually update it with new information as it becomes available. This approach ensures that the system remains current and effective in diagnosing a wide range of conditions.

In addition to predicting diseases, our system evaluates the severity of potential health issues, offering users detailed and actionable insights into their health status. To enhance accessibility and usability, we are planning to develop a web application that will provide easy portability, configuration, and remote access. This is especially important for users in regions with limited medical facilities. By delivering a convenient and efficient solution for disease prediction and guidance, our system aims to significantly improve overall health management. It will enable users to obtain timely health insights and recommendations without the need for immediate physical consultations, thus making healthcare more accessible and streamlined.

Introduction

When individuals experience symptoms of a disease, visiting a doctor can be both time-consuming and costly. This issue is compounded for those living in areas with limited medical facilities, where accessing healthcare can be challenging and diagnosing a disease may be difficult, potentially delaying crucial treatment. An automated disease diagnosis program offers a practical solution to these challenges, providing a more efficient and cost-effective means of identifying potential health issues. Such a program can significantly reduce the need for immediate physical consultations, making it easier for individuals to receive timely health assessments.

Health+ is a web-based application designed to offer disease predictions based on the symptoms reported by users. The application leverages extensive datasets gathered from a variety of sources to analyze the symptoms entered by users. This analysis helps estimate the likelihood of various diseases, providing users with valuable insights into their health. By offering this information, Health+ assists users in making informed decisions about whether further medical consultation is necessary, thereby streamlining the process of seeking appropriate healthcare

As reliance on the internet continues to grow, people increasingly turn to online resources for quick information and solutions. Health+ takes advantage of this widespread internet access by providing 24-hour availability for disease prediction and health guidance. This constant availability ensures that users can manage their health proactively, particularly in areas where traditional healthcare services are less accessible. By integrating comprehensive online resources with disease prediction capabilities, the system enhances overall health management, offering a convenient and efficient solution for users to monitor and address their health concerns.

Problem Statement

In modern healthcare, the timely and accurate diagnosis of diseases is critical for effective treatment and improved patient outcomes. However, many individuals face significant barriers to accessing medical advice when they first experience symptoms. These barriers include uncertainty about the severity of their symptoms, difficulty in accessing healthcare services due to geographic or financial constraints, and the tendency to self-diagnose or ignore symptoms. This delay in seeking professional medical advice can lead to the progression of potentially serious conditions, resulting in more complex and costly treatments and increased strain on healthcare systems.

The challenge is compounded by the overlapping nature of symptoms across different diseases. For example, symptoms such as fever, cough, and fatigue are common to a wide range of illnesses, from the common cold and influenza to more severe conditions like pneumonia and COVID-19. This symptom overlap makes it difficult for individuals to accurately assess their health status without professional medical input. Consequently, there is a pressing need for a solution that can help bridge this gap by providing preliminary disease predictions based on reported symptoms, thereby encouraging individuals to seek timely medical intervention.

Current self-diagnosis tools and online health information often lack the specificity and accuracy needed to provide reliable guidance. These tools typically do not account for individual health histories, variations in symptom presentation, or the probability of co-occurring conditions. Moreover, they often fail to provide actionable recommendations, leaving users uncertain about the next steps in their healthcare journey.

The core problem lies in the need for an accessible, accurate, and user-friendly tool that can assist individuals in assessing their health status based on initial symptoms. This tool must be capable of handling the complexity and variability of symptom presentation while providing personalized predictions that consider individual health factors. Additionally, it must ensure data privacy and security, maintaining user trust and compliance with relevant regulations.

Proposed Solution

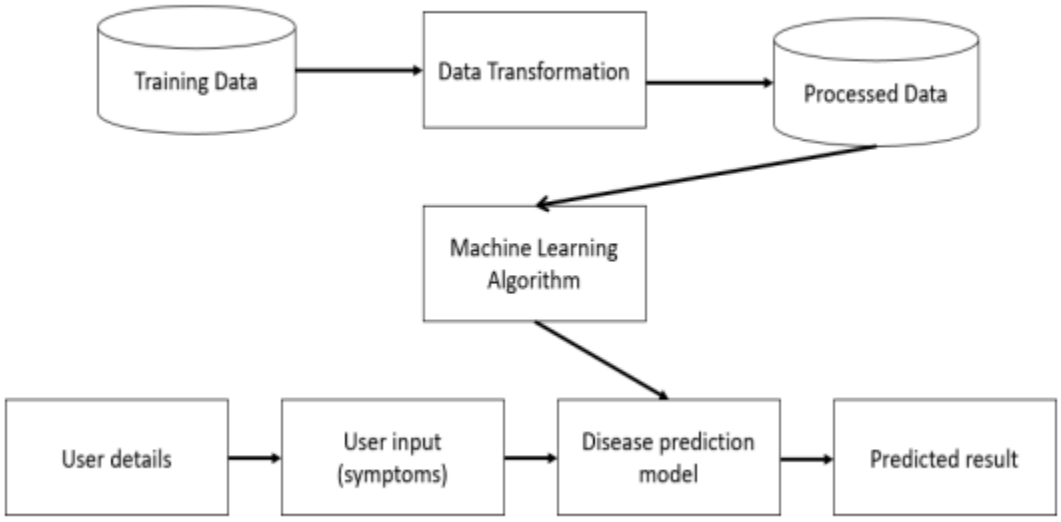
To address the critical issue of timely and accurate disease diagnosis based on initial symptoms, we propose the development of an Health+. This innovative solution aims to bridge the gap between symptom onset and professional medical diagnosis, providing users with preliminary disease predictions that encourage early medical consultation and proactive healthcare management.

The core of this system is an advanced machine learning model trained on a comprehensive dataset of symptoms and associated diseases. The system will be capable of analyzing user-reported symptoms and providing probable disease predictions with associated confidence levels. The diseases covered will include approximately 15 common and impactful conditions: common cold, influenza, COVID-19, pneumonia, strep throat, migraine, diabetes (Type 1 and Type 2), hypertension, heart attack, asthma, and depression. These diseases were selected due to their prevalence, the overlap of their symptoms, and the significant benefits of early detection.

A key feature of this system is its user-friendly interface, which will allow users to easily input their symptoms, including severity and duration.Additionally, the system will account for individual health factors such as age, gender, medical history, and lifestyle, providing personalized predictions that enhance the accuracy of the diagnosis.The prediction engine will utilize sophisticated algorithms to match reported symptoms with potential diseases, providing users with a list of probable conditions ranked by probability scores. The system will provide detailed information about each predicted disease, including symptoms, causes, and recommended next steps, thereby guiding users towards appropriate medical consultation.

Data privacy and security are paramount in this solution. The system will ensure that all user data is handled securely. Users will have the option to input symptoms anonymously if they prefer, ensuring their personal information remains protected.To continually improve the accuracy and reliability of the predictions, the system will incorporate a feedback loop where users can report the outcomes of their medical consultations. This feedback will be used to refine and update the machine learning models, ensuring they stay current with the latest medical research and user experiences. By integrating user feedback and continuously updating the prediction algorithms, the system will evolve to become increasingly accurate and reliable over time.

Methodology / Block Diagram



Hardware Requirements

- Processor : Core i3/i5/i7
- RAM : 2-4GB
- HDD : 500 GB

Software Requirements

- Platform : Windows Xp/7/8/10
- Coding Language : Python

Proposed Evaluation Measures

To evaluate the effectiveness of a disease prediction model based on symptoms, we need to consider various metrics that can provide insights into its accuracy, reliability, and usability. Here are the key evaluation measures:

1. Accuracy:
   - Measures the percentage of correct predictions made by the model out of all predictions made.
   - Formula: Accuracy = (TP + TN) / (TP + TN + FP + FN)

2. Precision:
   - Measures the percentage of true positive predictions out of all positive predictions made by the model.
   - Formula:Precision = TP / (TP + FP)
3. Recall (Sensitivity):
   - Measures the percentage of true positive predictions out of all actual positive cases.
   - Formula: Recall = TP / (TP + FN)
4. F1 Score:
   - The harmonic mean of precision and recall, providing a single measure of a model's performance.
   - Formula: F1-Score = 2 * (Precision * Recall) / (Precision + Recall)
5. ROC-AUC Score (Receiver Operating Characteristic - Area Under Curve):
   - Measures the ability of the model to distinguish between classes. Higher values indicate better model performance.
   - The curve is plotted with True Positive Rate (Recall) against False Positive Rate.
6. Confusion Matrix:
   - A table used to describe the performance of a classification model, showing the true positives, false positives, true negatives, and false negatives.
7. Specificity:
   - Measures the percentage of true negative predictions out of all actual negative cases.
   - Formula: Specificity = TN / (TN + FP)
8. Logarithmic Loss (Log Loss):
   - Measures the uncertainty of the model by comparing the predicted probability distributions of the outcomes with the true outcomes.
9. Cross-Validation:
   - Evaluates the model by dividing the dataset into k subsets and training/testing the model k times, each time using a different subset as the test set.
10. Confusion Matrix:
   - Provides detailed information about true positives, false positives, true negatives, and false negatives, allowing for deeper insight into the types of errors the model is making.

Key Terms:

- TP: True Positive (correctly predicted positive)
- TN: True Negative (correctly predicted negative)
- FP: False Positive (incorrectly predicted positive)
- FN: False Negative (incorrectly predicted negative)

Conclusion

Health+ effectively tackles the challenge of accessing timely and affordable healthcare, particularly for individuals residing in remote or underserved areas. Traditional doctor visits can be time-consuming and costly, and this issue is exacerbated for those in locations with limited medical facilities. By offering an automated disease diagnosis through a web-based platform, Health+ provides a practical solution, enabling users to evaluate their health based on reported symptoms without the immediate need for an in-person consultation.

Utilizing extensive datasets, Health+ analyzes user-reported symptoms to estimate the likelihood of various diseases. This analysis equips users with essential insights into their health status, aiding them in deciding whether further medical evaluation is necessary. This preliminary assessment enhances the efficiency and cost-effectiveness of disease diagnosis, particularly in regions where access to traditional healthcare services is limited.

Health+ leverages the widespread use of online resources by providing 24-hour availability, allowing users to manage their health proactively. This constant access ensures that users can receive timely health guidance and monitor their health status conveniently, overcoming barriers related to healthcare access. The ability to access health insights anytime and from anywhere significantly improves health management and decision-making.

In summary, Health+ represents a significant advancement in digital health technology. By combining comprehensive disease prediction capabilities with the convenience of online access, Health+ offers a more efficient and user-friendly approach to health monitoring. This innovative system not only enhances individual health management but also supports more informed healthcare decisions and better overall health outcomes.

References

[1] A. Singh et al., "Heart Disease Prediction Using Machine Learning Algorithms", 2020 International Conference on Electrical and Electronics Engineering (ICE3), pp. 452-457, February 2020.

[2] A Narin, C Kaya and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks", Mar 2020.

[3] Rajesh. Ranjan, "Predictions for COVID-19 outbreak in India using Epidemiological models", 2020.

[4] Mohan Senthilkumar, Chandrasegar Thirumalai and Gautam Srivastava, "Effective heart disease prediction using hybrid machine learning techniques", IEEE Access, vol. 7, pp. 81542-81554, 2019.

[5] M. Bayati, S. Bhaskar and A. Montanari, "Statistical analysis of a low cost method for multiple disease prediction", Statistical Methods Med. Res., vol. 27, no. 8, pp. 2312-2328, 2018.

[6]A. K. Shrivas and S. Kumar Sahu, "Classification of Chronic Kidney Disease using Feature Selection Techniques", IJCSE, vol. 6, no. 5, pp. 649- 653, 2018. 75

[7] N Chaithra and B Madhu, "Classification Models on Cardiovascular Disease Prediction using Data Mining Techniques", Journal of Cardiovascular Diseases & Diagnosis, vol. 6, pp. 1-4, 2018.

[8] R. Katarya and P. Srinivas, “Predicting heart disease at early stages using machine learning: A survey,” in 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC),2020,pp. 302–305

[9] Poudel RP, Lamichhane S, Kumar A, et al.Predicting the risk of type 2 diabetes mellitus using data mining techniques. J Diabetes Res.2018;2018:1686023.

[10] S. Ismaeel, A. Miri, and D. Chourishi, “Using the extreme learning ma-chine (elm) technique for heart disease diagnosis,” in 2015 IEEE Canada International Humanitarian Technology Conference
(IHTC2015), 2015,pp. 1–3.

[11] A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, “Prediction of heart disease using machine learning,” in 2018 Second International Conference on Electronics, Communication and Aerospace Technology(ICECA), 2018, pp. 1275–1278.

[12] Deo RC. Machine learning in medicine. Circulation. 2015;132(20):1920-1930.

[13] Breiman L, Friedman JH, Olshen RA, Stone CJ.Classification and Regression Trees. Wadsworth and Brooks; 1984.

[14] Parashar A, Gupta A, Gupta A. Machine learning techniques for diabetes prediction. Int J Emerg Technol Adv Eng. 2014;4(3):672-675.

[15] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Wadsworth and Brooks; 1984.

[16] Paniagua JA, Molina-Antonio JD, Lopez-Martinez F, et al. Heart disease prediction using random forests. J Med Syst. 2019;43(10):329.