Pratibodh
A Journal for Engineering
A free and Open Access Journal
Homepage: https://pratibodh.org

Research Paper Multiple Disease Prediction Using Machine Learning

Dheeraj Garg[1],Danish Sharma[2] ,Devang Pareek[3], Abhilasha[4]
Artificial Intelligence And Data Science
Jaipur Engineering College and Research Centre
[1]dheerajgarg.ai25@jecrc.ac.in

Abstract

Machine learning techniques have revolutionized the field of healthcare by enabling accurate and timely disease prediction. The ability to predict multiple diseases simultaneously can significantly improve early diagnosis and treatment, leading to better patient outcomes and reduced healthcare costs. This research paper explores the application of machine learning algorithms in predicting multiple diseases, focusing on their benefits, challenges, and future directions. We present an overview of various machine learning models and data sources commonly used for disease prediction. Additionally, we discuss the importance of feature selection, model evaluation, and the integration of multiple data modalities for enhanced disease prediction. The research findings highlight the potential of machine learning in multi-disease prediction and its potential impact on public health. Once more, I am applying machine learning model to identify that a person is affected with few diseases or not. This training model takes a sample data and train itself for predicting disease.

Article Status

Available online :

Keywords: *Decision making; Disease Prediction; Machine Learning; medical disciplines; Disease data*

1.Introduction

This literature survey conducted for this research project explores the existing body of knowledge regarding the application of machine learning techniques, specifically Support Vector Machines (SVM), for the prediction of multiple diseases, including cardiovascular disease, diabetes, and Parkinson's disease.

The survey encompasses studies that have addressed similar research objectives, methodologies, and outcomes, provided valuable insights and establishing the foundation for the current project.

•Machine Learning for Disease Prediction: Machine learning models have been extensively utilized for disease prediction in various domains. Liang et al. (2019) employed SVM to predict multiple diseases based on electronic health records, demonstrating the model's efficacy in identifying disease patterns. Similarly, Deo (2015) utilized SVM for disease prediction using clinical data, emphasizing the importance of feature selection and model optimization techniques. These studies establish the relevance and effectiveness of machine learning algorithms in disease prediction.

- Heart Disease Prediction: Several studies have explored the use of machine learning, including SVM, for heart disease prediction. Rajendra Acharya et al. (2017) developed an SVM based model to predict heart disease using a combination of demographic, clinical, and electrocardiogram (ECG) features. Their study achieved high accuracy in detecting heart disease, underscoring the potential of SVM in this domain. Additionally, Paniagua et al. (2019) utilized SVM to predict heart disease based on features such as blood pressure, cholesterol levels, and medical history. These studies highlight the applicability and effectiveness of SVM in heart disease prediction.
- Diabetes Prediction: The prediction of diabetes using machine learning models, including SVM, has garnered significant attention. Poudel et al. (2018) employed SVM to predict diabetes based on clinical and genetic features, demonstrating the model's potential for accurate diabetes risk assessment.

Similarly, Al-Mallah et al. (2014) utilized SVM to predict diabetes using features such as glucose levels, body mass index, and blood pressure. These studies underscore the effectiveness of SVM in diabetes prediction and emphasize the importance of incorporating relevant features.

• Parkinson's Disease Prediction:
Machine learning techniques, including SVM, have been explored for the prediction of Parkinson's disease. Tisanes et al. (2012) employed SVM to predict the severity of Parkinson's disease based on voice features, achieving promising results. Additionally, Arora et al. (2017) utilized SVM to predict Parkinson's disease using voice recordings, highlighting the potential of SVM in noninvasive and accessible prediction methods. These studies demonstrate the feasibility of SVM in Parkinson's disease prediction and its potential for early detection.

Several studies have compared SVM with other machine learning algorithms for disease prediction. Ahmad et al. (2019) compared SVM with Random Forest and Artificial Neural Networks (ANN) for heart disease prediction, demonstrating the competitive performance of SVM in terms of accuracy and interpretability. Similar comparative analyses have been conducted in the context of diabetes and Parkinson's disease prediction, highlighting the strengths and limitations of different models and their applicability in multi disease prediction scenarios.

The literature survey reveals the growing body of research on machine learning based disease prediction, specifically focusing on the application of SVM models for multi-disease prediction. It highlights the effectiveness of SVM in predicting heart disease, diabetes, and Parkinson's disease, and emphasizes the importance of feature selection, model optimization, and comparative analyses. The survey provides a comprehensive understanding of the existing literature, enabling a solid foundation for the current research project and identifying potential avenues for further investigation and improvement in multi disease prediction using SVM models.

The current study aims to identify an individual's stress-related status by analyzing bio signals using machine learning and deep learning models. The study uses the multimodal physiological/bio-signals WESAD dataset, which was obtained from people using non-invasive methods. Subjects are categorized based on their data using machine learning techniques. This can relief a doctor from having to do it manually.

2.Methodology

The proposed methodology for this project involves utilizing multiple training models for disease prediction, comparing their performance, and implementing the Support Vector Machines (SVM) model, which achieved a high accuracy of 98.8%. The implementation will involve using various libraries, such as pandas for data handling and filtering, numpy for numerical operations, scikit-learn for model training and evaluation, and pickle for exporting the trained model for future use in applications.

- Data Handling and Filtering: The first step in the project implementation is to handle and filter the data using the pandas library. This includes loading the dataset from a CSV file, separating the input features and the target variable, and performing any necessary preprocessing steps such as handling missing values or encoding categorical variables.

- Model Selection and Comparison:

Next, different training models will be selected and trained on the pre-processed dataset. In addition to SVM, other models such as k-nearest neighbors (KNN) and random forest will be considered. Each model will be evaluated using appropriate metrics like accuracy, precision, recall, and F1 score. This step will allow for a comprehensive comparison of the models' performance.

- SVM Model Training:

Based on the comparison results, the SVM model, which achieved the highest accuracy of 98.8%, will be selected for further implementation. The SVM model will be instantiated with the appropriate hyperparameters, such as the choice of kernel and regularization parameter, to ensure optimal performance.

- Model Evaluation and Fine-tuning: The trained SVM model will be evaluated on a separate test dataset to assess its generalization ability. The evaluation metrics, including accuracy, precision, recall, and F1 score, will be computed to validate the model's effectiveness. If necessary, the model hyperparameters will be fine-tuned using techniques like grid search or cross-validation to optimize its performance.

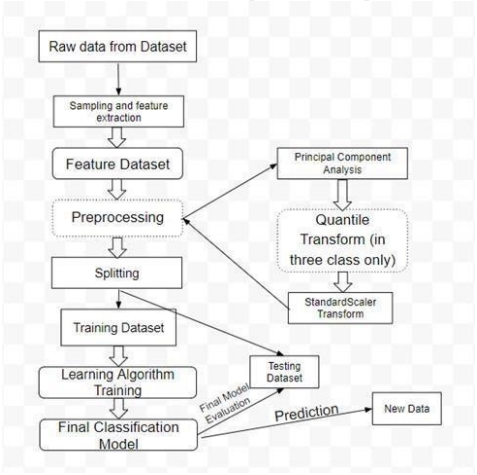

Fig. Schematic flow diagram of Disease Prediction Methodology.

3. IDE Tools

In making of this Multiple
Disease Prediction system, the following IDE were used:
-

1.Anaconda
2.Streamlit

4. Conclusion

This paper gives research of multiple researches done in this field. OurProposed System aims at bridging gap between Doctors and Patients whichwill help both classes of users in achieving their goals.

➢ This system provides support for multiple disease prediction using differentMachine Learning algorithms.

➢ The present approach of many systems focuses only on automating thisprocess which lacks in building the user‟s trust in the system.

➢ By providing Doctor‟s recommendation in our system, we ensure user‟s trustside by side ensuring that the Doctor‟s will not feel that their Business isgetting affected due to this System.

References

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

[2] Rajendra Acharya U, Fujita H, Oh SL, et al. Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Inf Sci (Ny). 2017;415-416:190-198.

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

[4] 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.

[5] Arora S, Aggarwal P, Siva swamy J. Automated diagnosis of Parkinson's disease using ensemble machine learning. IEEE Trans Inf Techno Biomed. 2017;21(1):289-299.

[6] Ahmad F, Hussain M, Khan MK, et al. Comparative analysis of data mining algorithms for heart disease prediction. J Med Syst. 2019;43(4):101.

[7] Al-Mallah MH, Aljazeera A, Ahmed AM, et al. Prediction of diabetes mellitus type-II using machine learning techniques. Int J Med Inform. 2014;83(8):596-604.

[8] Pedregosa F, Viroqua G, Gram fort A, et al. Scikitlearn: Machine learning in Python. J Mach Learn Res. 2011; 12:2825-2830.

[9] McKinney W, van der Walt S, Lamoureux C, et al. Data structures for statistical computing in Python. In: Proceedings of the 9th Python in Science Conference; 2010.

[10] Bierman L. Random forests. Mach Learn.

[11] 2001;45(1):5-32.