**An Approach to forecasting multiple maladies utilizing machine learning algorithms**

P. Renukadevi

Associate Professor,

Department of Computer Science and Engineering,

Paavai Engineering College (Autonomous), Namakkal.

Varun Krishnan V, Nandhakumar M, Suriyasankar P

UG Student,

Department of Computer Science and Engineering,

Paavai Engineering College (Autonomous), Namakkal.

**Abstract**

Nowadays, machine learning is utilized across various industries, including healthcare, where it holds notable importance. One way to enhance sufferer satisfaction is by applying machine learning techniques to healthcare. The "An Approach to forecasting multiple maladies utilizing machine learning algorithms" system is based on predictive modelling that considers symptoms entered by the sufferer to forecast the disease. While many existing machine learning applications for health analytics focus on only one disease, such as diabetes, cancer, or skin diseases, there is a lack of a unified system capable of predicting multiple diseases. This article proposes a system capable of predicting diabetes and heart disease To achieve more accurate predictions of heart disease risk levels, the project proposes the use of a support vector machine (SVM) as the backbone of computational diagnostic tools. The SVM model is a promising classification method that can predict medication adherence in Cardiovascular disease (CVD) patients, thereby helping differentiate sufferers and making evidence-based decisions possible. To predict early diabetes risk, bagging and boosting methods using DTB algorithms were applied to experimental data. Besides this the admin basically the doctor can also able to train their own model using our web application which is created on top of Flask. Ultimately, this project can help improve the health of many people by allowing their condition to be monitored closely, and necessary precautions can be taken to increase their life expectancy.

*Keywords – Multiple maladies prediction, SVM, bagging and boosting, CVD*

**Introduction**

In modern times, healthcare has become an integral part of our lives, and technological advancements have made it possible to make significant progress in this field. One such area where technology has been immensely useful is the application of machine learning in predicting diseases. The ability to accurately and swiftly predict diseases is crucial in the healthcare industry as it can lead to improved patient outcomes and increase life expectancy. Hence, the objective of this project is to create a system capable of predicting multiple diseases, with a particular focus on diabetes and heart disease.

To achieve this, we have utilized two distinct approaches. For predicting the risk levels of heart disease, we have employed a support vector machine (SVM) model, which serves as the backbone of our computational diagnostic tool. SVM is a promising classification method and has shown effectiveness in predicting medication adherence in CVD patients, enabling differentiating between sufferers and facilitating evidence-based decision-making.

To predict early diabetes risk, we have applied bagging and boosting methods with decision tree-based algorithms on experimental data. Our goal is to expand the system to include other diseases such as fever analysis and other skin diseases in the future. The analysis of multiple diseases utilizes machine learning algorithms to examine all parameters related to the disease and identify the maximum effect caused by them.

The system we have developed can help enhance the health of many people by closely monitoring their condition and taking necessary precautions to increase their life expectancy. Overall, this project has the potential to significantly improve healthcare outcomes and enhance patient satisfaction.

**Existing system**

Existing maladies prediction systems focus on single diseases, causing inefficiencies and delays in treatment for patients with multiple health concerns. These single disease forecasting systems may not provide a holistic view of a patient's health, and can be cost-ineffective for healthcare providers. A multiple maladies forecasting system can address these limitations by providing a comprehensive approach to healthcare, taking into account all relevant health factors, and offering a more efficient and cost-effective solution for healthcare providers.

**Proposed system**

Our proposed system is a web application developed on Flask that enables doctors to train their own machine learning models for predicting various diseases, with a focus on heart disease and diabetes. The trained models are then integrated into the application, where users can input their symptoms and receive immediate predictions. The system analyses all relevant parameters related to each disease to provide accurate and reliable predictions. By offering predictions for multiple diseases in a single application, our system saves time and resources for both healthcare providers and patients. Overall, our system offers a more comprehensive and personalized approach to disease prediction and diagnosis in the healthcare industry.

**Data preparation**

Machine Learning projects rely heavily on data, and this project on predicting diabetes and heart disease is no exception. The data used in this project was obtained from Kaggle and included various features, including age, gender, blood pressure, and cholesterol levels, which were essential in the prediction process. To enhance the accuracy of the predictions, various techniques for feature selection and extraction were employed. For heart disease prediction, the Filter wrapper model was used for feature selection, while the confusion matrix was used for feature extraction. For diabetes prediction, the project utilized the principal component analysis (PCA) for feature selection, and the confusion matrix for feature extraction..

As per the standards training set and testing set are prepared. By using Scikit learn train\_test\_split method the data has been divided into two sets one is of 70% of the dataset for training and the other is of 30% of the dataset for testing. Example: diabetes\_feature\_training, diabetes\_feature \_testing, diabetes\_label\_ training, diabetes\_label \_testing = train\_test\_split (diabetes\_features, diabetes\_label, test\_size=0.3, random\_state=0)