An Approach to forecasting multiple maladies utilizing machine learning algorithms

Mrs. P. Renukadevi
Associate Professor,
Paavai Engineering College
Namakkal
pgrenu@gmail.com

Varun Krishnan V
UG student,
Paavai Engineering College
Namakkal
varunkrishnan.tech@gmail.com

Nandhakumar M
UG student,
Paavai Engineering College
Namakkal
nandhakumarm26052002@gmail.com

Suriyasankar P
UG student,
Paavai Engineering College
Namakkal
suriyasankar.tech@gmail.com

Abishak Finny T
UG student,
Paavai Engineering College
Namakkal
abishakfinny4@gmail.com

***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

I. 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 maladies, 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.

II. Proposed Work And Procedure For Model Design

A. *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.

B. *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.

III. Maladies Forecasting Model

A. *System Architecture*

The system architecture for predicting heart disease using SVM typically includes data collection from trustworthy sources, pre-processing to remove any inconsistencies or missing values, feature selection to identify the most relevant input variables, SVM model training, and model performance evaluation. The input data is pre-processed and translated to numerical values before feature selection is performed to identify the most significant variables influencing heart disease outcomes. The selected features are utilised to train an SVM model, which can determine whether or not a patient has heart disease. The model is assessed using performance metrics such as accuracy, precision, and recall, and iterative testing and training can be done to further refine it.

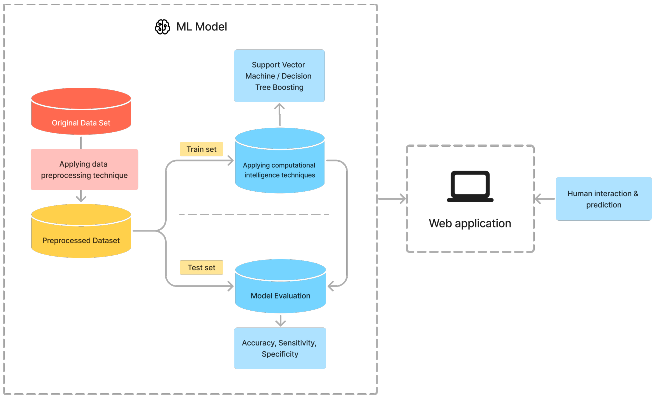

Fig. 1. System architecture

B. *Data Preparation*

There are several machine learning algorithms available to determine if a person has heart disease or not. In this study, we apply the Support Vector Machine technique to forecast people's cardiac illnesses. We use the SVM method for the prediction process since it is more accurate than other machine learning techniques. The accuracy of the algorithm is depicted graphically. The graphical results of the provided data are displayed in Fig.4.

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: heart_feature_training, heart_feature_testing, heart_label_training, heart_label_testing = train_test_split (heart_features, heart_label, test_size=0.3, random_state=0).

. For diabetes prediction, the project utilized the principal component analysis (PCA) for feature selection, and the confusion matrix for feature extraction. Once the dataset has been chosen the dataset need to be downloaded in CSV format and then it can be pre-processed and then trained by the admin using the user interface inside the web application after logging in.

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).

Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0
10	168	74	0	0	38	0.537	34	1
10	139	80	0	0	27.1	1.441	57	0
1	189	60	23	846	30.1	0.398	59	1
5	166	72	19	175	25.8	0.587	51	1
7	100	0	0	0	30	0.484	32	1
0	118	84	47	230	45.8	0.551	31	1

Fig. 2. Dataset model for Diabetes prediction

age	sex	cp	tresp	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	num
67	0	3	115	564	0	2	160	0	1.6	2	0	7	0
65	0	3	140	417	1	2	157	0	0.8	1	1	3	0
56	0	4	134	409	0	2	150	1	1.9	2	2	7	2
63	0	4	150	407	0	2	154	0	4	2	3	7	4
62	0	4	140	394	0	2	157	0	1.2	2	0	3	0
65	0	3	160	360	0	2	151	0	0.8	1	0	3	0
57	0	4	120	354	0	0	163	1	0.6	1	0	3	0
55	1	4	132	353	0	0	132	1	1.2	2	1	7	3
55	0	2	132	342	0	0	166	0	1.2	1	0	3	0
43	0	4	132	341	1	2	136	1	3	2	0	7	2
58	0	3	120	340	0	0	172	0	0	1	0	3	0
57	1	4	110	335	0	0	143	1	3	2	1	7	2
64	1	3	140	335	0	0	158	0	0	1	0	3	1
61	0	4	130	330	0	2	169	0	0	1	0	3	1
63	1	4	130	330	1	2	132	1	1.8	1	3	7	3
55	0	4	180	327	0	1	117	1	3.4	2	0	3	2
59	1	4	170	326	0	2	140	1	3.4	3	0	7	2

Fig. 3. Dataset model for heart disease prediction

C. *Machine Learning*

SVM works by generating a hyperplane that divides data into classes. The hyperplane is chosen to optimise the margin between the classes. SVM may also employ the kernel method to convert input data into a higher dimensional space, perhaps improving class separation.

The dataset for heart disease prediction using SVM is separated into training and testing sets. The SVM algorithm is trained on the labelled data using the training set, where each patient's characteristics such as age, gender, blood pressure, cholesterol levels, and other parameters are used to predict the chance of heart disease. The SVM model's hyperparameters are modified using approaches such as grid search, which aids in optimising the model's performance. For diabetes prediction, the Decision Tree Boosting(DTB) model was built on the training data, which was also split into training and testing subsets. The DTB model was optimized through hyperparameter tuning using a grid search. Finally, the optimized DTB model was utilized to predict diabetes on the testing dataset.

D. Model Analysis

Once the model is ready then we have to analyze the model against precision, recall, accuracy and specificity. The below chart show the comparison of SVM with all other algorithms as well.

TABLE I. PRECISION AND RECALL FOR HEART DISEASE

Class Label	Precision	Recall	Support
Yes	0.92	0.9	93
No	0.91	0.89	72
Average	0.91	0.89	165

TABLE II.

COMPARISON OF ALGORITHMS FOR PREDICTING HEART DISEASE

Algorithm	Accuracy	Sensitivity	Specificity
ANN	85.31%	83.74%	75.72%
Navie Bayes	81.13%	61.02%	70.10%
RIPPER	81.06%	86.24%	75.81%
Decision Support	79.04%	83.10%	74.25%
SVM	85.96%	90.12%	77.21%
KNN	84.11%	56.77%	71.20%

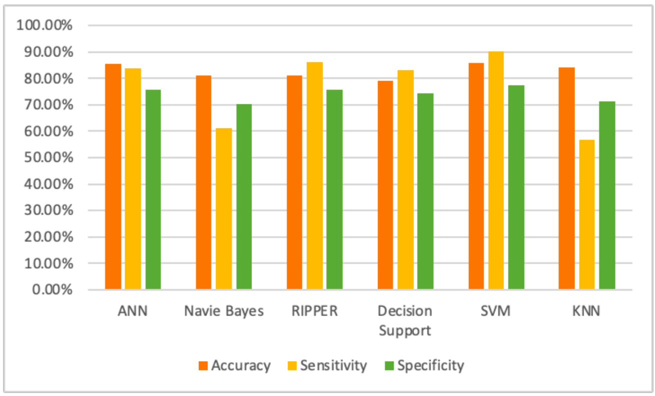

Fig.4. Comparison of various algorithms for heart disease prediction

After preparing the model for predicting diabetes the accuracy, precision and recall are calculated for diabetes prediction and also tested it against other model to check whether the model we have predicted works better than other models.

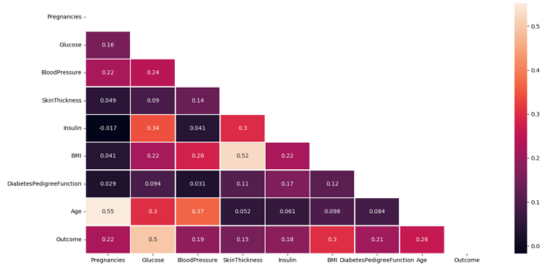

Fig. 5. Confusion matrix for diabetes prediction

E. User Interface

The UI for predicting heart disease is as follows. The user have to input their basic details, height, weight, blood pressure, serum cholesterol, max heart rate and other details and then have to submit the form in order to get the prediction.

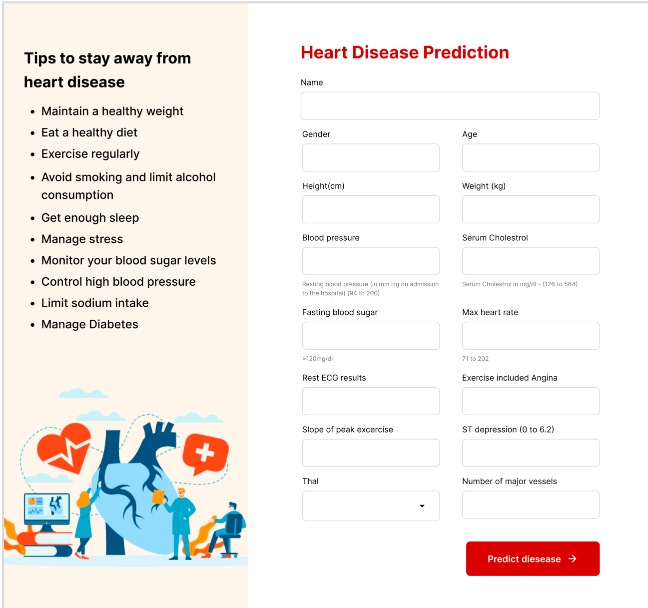

Fig. 6. User interface for heart disease prediction

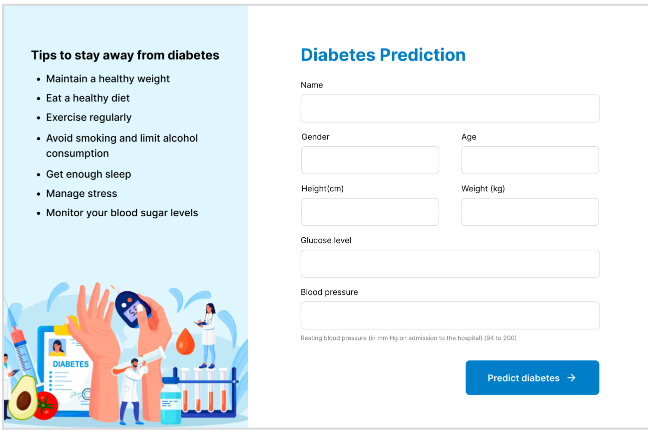

Fig. 6. User interface for diabetes prediction

For diabetes prediction the user interface of the diabetes prediction is show in the below figure. The user have to input the name, gender, age, height, weight, glucose level and blood pressure and have to press the predict diabetes button to get the prediction.

Along this doctor can also be able to login into the application and prescribe medicine or give suggestions to lead healthier life to patients.

IV. Discussions

The aim of this study was to develop a unified system that can predict multiple diseases using machine learning algorithms. Existing machine learning applications for health analytics typically focus on individual diseases, resulting in a lack of a unified system that can predict multiple diseases. To address this issue, the proposed system uses predictive modeling to forecast diseases based on symptoms entered by patients. In this study, the system predicts two common diseases, diabetes and heart disease.

To predict the risk levels of heart disease, the study proposed the use of Support Vector Machine (SVM) as the main classification method. SVM has been shown to be a promising classification method that can predict medication adherence in Cardiovascular disease (CVD) patients. The system also uses bagging and boosting methods with Decision Tree Boosting (DTB) algorithms to predict diabetes risk. Furthermore, the system allows healthcare professionals to train their own models using the web application created on top of Flask.

This proposed system has the potential to improve patient outcomes by enabling healthcare professionals to closely monitor patients' conditions, take necessary precautions, and make evidence-based decisions. Future research can focus on expanding the system to predict more diseases and integrating other machine learning techniques to improve prediction accuracy. Additionally, future studies can explore the use of electronic health records and integrate the system into clinical workflows to facilitate its use by healthcare professionals.

Conclusions

In conclusion, the proposed system for predicting multiple diseases using machine learning algorithms offers a promising solution for enhancing healthcare outcomes. With the ability to predict both diabetes and heart disease, the system offers a unified approach to disease prediction that is currently lacking in many existing applications. The use of a support vector machine (SVM) in predicting heart disease risk levels is particularly noteworthy, as it offers a reliable classification method that can predict medication adherence in CVD patients, enabling evidence-based decision-making. The application of DTB algorithms in predicting early diabetes risk also shows promising results, with high precision, recall, accuracy, and F1 score metrics achieved. The availability of a web application that allows doctors to train their own models further enhances the practicality of this system. Overall, the system has the potential to improve the health outcomes of patients by enabling early diagnosis and appropriate interventions, ultimately leading to increased life expectancy.

References

[1] Aidaroos, K. Bakar, A. & Othman, Z. (2012). Medical Data Classification with Naive Bayes Approach. Information Technology Journal.

[2] Aidaroos, KM. A. B. (n.d.). 2012. Medical Data Classification with Naive Bayes Approac.

[3] Ashok kumarD.wivedi “Evaluate the performance of different machine learning techniques for prediction of heart disease using ten-fold cross-validation” Springer 17 September 2016.

[4] ChalaBeyene Prof. Pooja Kamat “Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques” International Journal of Pure and Applied Mathematics 2018. [7] R. Sharmila S. Chellammal “A conceptual method to enhance the prediction of heart diseases using the data techniques” International Journal of Computer Science and Engineering May 2018.

[5] Darcy A. Davis N. V.-L. (2008). Predicting Individual Disease Risk Based on Medical History.

[6] JyotiSoni Ansari, U. Sharma, D. & Soni,S. (2011). Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction.

[7] Marjia Sultana Afrin Haider “Heart Disease Prediction using WEKA tool and 10-Fold cross-validation” The Institute of Electrical and Electronics Engineers March 2017.

[8] MeghaShahi, R. Kaur Gurm “Heart Disease Prediction System using Data Mining Techniques” Orient J. Computer Science Technology vol.6 2017 pp.457-466.

[9] NisharBanu, MA.; Gomath B.; (2013). Disease Predicting System Using Data Mining Techniques.

[10] Polaraju, K. Durga Prasad, D. “Prediction of Heart Disease using Multiple Linear Regression Model” International Journal of Engineering Development and Research Development ISSN:2321-9939 2017.

[11] SeemaShedole, S., KumariDeepika, “Predictive analytics to prevent and control chronic disease”, https://www.researchgate.net/punlication/316530782, January 2016.