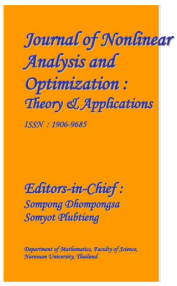
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1: 2024
ISSN :**1906-9685**



PREDICTIVE MODELLING FOR CONCURRENT DISEASES USINGMACHINE LEARNING

[#1]**Dr.G.PANDIT SAMUEL**, Assistant Professor, Department of Computer Science & Engineering, Vignan's Institute of Information Technology(A), Visakhapatnam, Andhra Pradesh, India

[#2]**MEDISETTI GEETHA**, Student Department of Computer Science & Engineering,
[#3]**PALLAM JHASWIKA SHRIYA**, Student Department of Computer Science & Engineering,
[#4]**PATRI LAVANYA**, Student Department of Computer Science & Engineering,
[#5]**MALLEPUDI V SAI SRUJAN**, Student Department of Computer Science & Engineering,
[#6]**MATAM HARIKA**, Student Department of Computer Science & Engineering,
Vignan's Institute of Information Technology(A), Visakhapatnam, Andhra Pradesh, India

ABSTRACT: Concurrent diseases prediction refers to predicting the probability of a patient's disease after examining the combinations of the patient's symptoms. Effective patient care can be facilitated by physicians keeping track of a patient's status and medical history. The accuracy of the data utilized by the previous researchers to train the model is low since it is not altered and is solely dependent on the symptoms. In order to forecast the likelihood of concurrent diseases, the model has to be modified using machine learning algorithms like Random Forest, Logistic Regression etc. for better efficiency and accuracy. The medical dataset from Kaggle is used in the proposed model.
Keywords: Heart disease prediction, Diabetes prediction, Lung cancer disease prediction, Symptoms, Machine Learning.

1. INTRODUCTION

Concurrent diseases also known as comorbidities, refer to the coexistence of multiple health conditions in an individual. Traditional statistical methods have limitations in handling the complexity of concurrent diseases. These concurrent diseases are prevalent and can complicate diagnoses, treatment plans, and health outcomes. Machine Learning is the domain that uses past data for predicting. It is used in various fields such as finance, business, medical etc. Healthcare is the prime example to introduce the machine learning in medical field. Healthcare industries generate huge amount of data regarding the patient's health. By using this data and machine learning techniques, a model has been developed. Multiple diseases, including diabetes, heart disease, and lung cancer, can be predicted using this approach through a single websitewhich helps the people to identify the disease at earlier stage and can get treated. Various machine learning algorithms are used here to detect the disease. To determine which method is best for prediction, the accuracy of each is verified and compared with one another. Furthermore, to attain the highest level of accuracy in the predicted results, multiple datasets are used (a single dataset for each disease). In the dataset, the target field consist 0 and 1 as result where 0 indicates there is no disease and 1 indicates there is disease.

2. REVIEW OF LITERATURE

In many different fields, machine learning models have been widely used to forecast disease. SVM was used by Liang et al. (2019) to forecast a number of diseases based on digital health information, showcasing the model's effectiveness in spotting disease trends. In a similar vein, Deo (2015) used SVM to predict diseases based on clinical data, highlighting the significance of feature selection and model tuning methods. These studies demonstrate the applicability and efficiency of machine learning algorithms for the prediction of disease. Models for forecasting a patient's disease based on symptoms collected from the patient are included in several research articles. The following are the most accurate and often used models:

Support Vector Machine (SVM) was employed in the Jianfang et al. method to classify diseases according to their symptoms. Although it takes longer to predict diseases, the SVM model is effective at doing so. The method's only shortcoming is that it uses a hyper-plane for object classification, which is not very efficient. Only two groups of sample data can be accurately classified using the hyper-plane. However, the current state of the medical field necessitates the identification of symptoms relating to more than two classifications(diseases). The Keniya et al. KNN algorithm. Using this technique, they allocated the data point to the class that most.

Kashvi et al. use the KNN approach. In other instances, like the prediction of diabetes and heart disease, they have also demonstrated excellent accuracy. There is the problem of classifying diseases based on a minimal amount of data.

Gomathy and Rohith Naidu have created a web-based program that allows for remote access to predict diseases. The data fed into the system determines how accurate the model is. The problem with the proposed model is that software for disease prediction needs to be developed with a more accurate dataset in order to improve accuracy. Naive Bayes classifier was utilized in Chhogyal and Nayak's technique. Their disease prediction accuracy has been subpar.

A few problems with the previously described methods are their efficiency and accuracy, the small size of the data set required to train the model, and their consideration of only a few symptoms in order to make a diagnosis. In order to address all of these problems, a revised and precise model for predicting diseases in people must be proposed.

3. METHODOLOGY

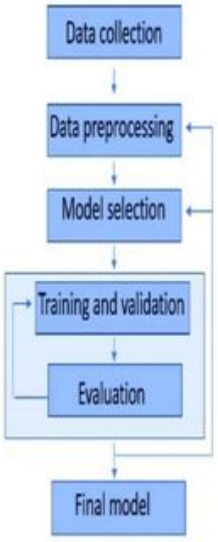


Figure demonstrates how multiple machine learning algorithms can help predict multiple diseases in order to close the gap between patientsand doctors so that each can pursue their own objectives.

Workflow:

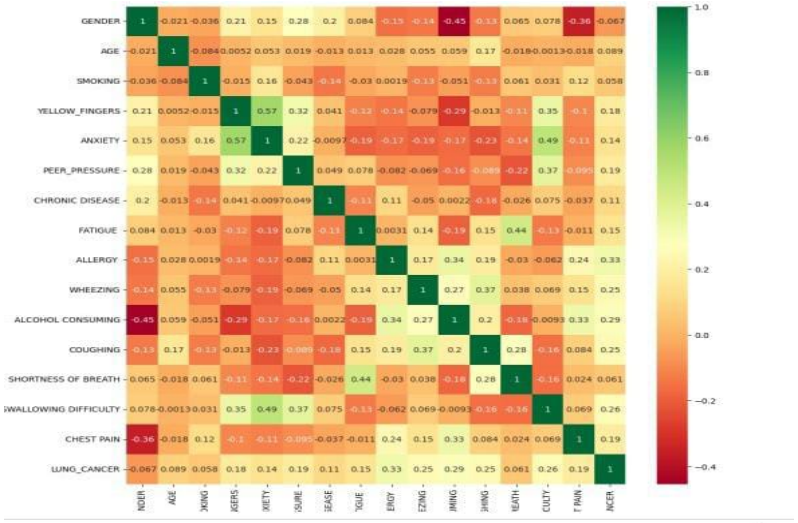
1) Collecting datasets :
Datasets from Kaggle have been gathered for this project, including a number of diseases like diabetes, lung cancer, and heart disease.
2) Performing data pre-processing :
Label encoding enables the transformation of categorical data, such as hunger and gender, into numerical data represented by 0s and 1s.
3) Creating models using various machine learning algorithms :
Different algorithms are applied for every disease to produce defined models.
4) Training the models :
Each dataset is divided into two parts i.e., testing set and training set.
5) Evaluating the models :
Accuracy is calculated after each model's evaluation.
6) Picking the best model :
The best model is chosen after its accuracy is compared to each other.

4. ALGORITHMS

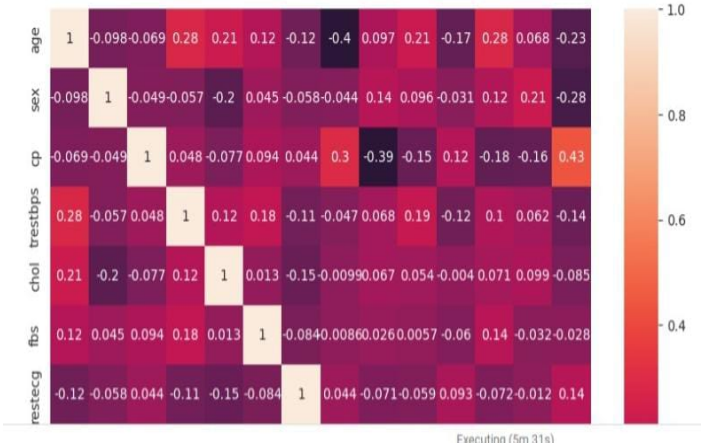
1) Support Vector Machine :
SVM is a procedural methodology used in regression and classification. It is mostly used to handle problems and challenges from everyday life. Every data point has a unique feature, and is represented as n-D space, where n is the total number of features. After the data has been divided using a hyper-plane, a classification approach is called.

2) Logistic Regression :
The statistical algorithm evaluates the relation between two data elements. It provides a probability value between 0 and 1 based on inputs that are independent variables.
3) Random Forest :
The way it works is by creating multiple decision trees based on distinct data subsets, then combining the results of each decision tree to produce the final result. The Bagging Principle is used by Random Forest. First, a method known as bootstrapping is used to generate a variety of subsets. Each model now operates on these subsets and generates output; the random forest then aggregates the outputs from the different models and generates output depending on voting.

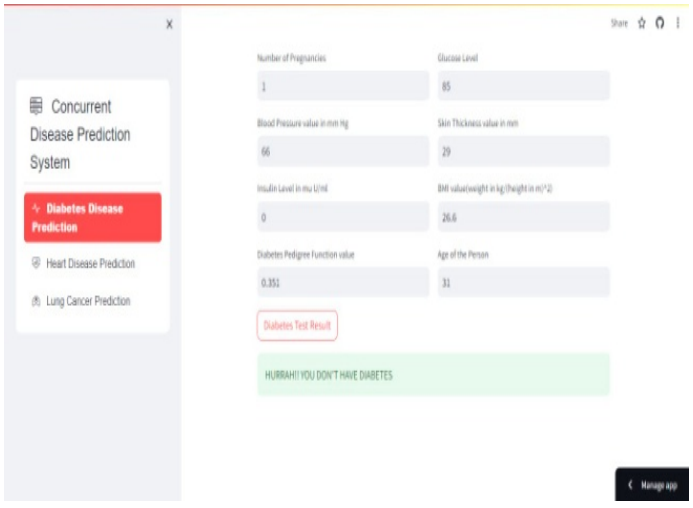
5. RESULTS



CORRELATION ANALYSIS OF LUNG CANCER PREDICTION



CORRELATION ANALYSIS OF DIABETES DISEASE PREDICTION



DIABETES PREDICTION PAGE

6. CONCLUSION

The proposed approach represents a remarkable development in supporting the health care industry by:

1. Decrease in healthcare expenses:

This model can assist in lowering healthcare costs and enhancing the general effectiveness of the healthcare system by enhancing patient outcomes and lowering the need for needless hospital visits.

2. Better patient outcomes:

Disease prediction apps can assist enhance patient outcomes by enabling earlier and more efficient interventions by giving medical professionals insightful information about a patient's disease risk.

3. Early diagnosis:

This model can assist healthcare providers in making an early diagnosis, which is essential for enhancing patient outcomes. It does this by evaluating patient data and detecting risk factors for particular diseases. Finally, we draw the conclusion that our model can offer improved precision and a trustworthy model for illness prediction based on symptoms.

7. COMPARATIVE ANALYSIS

SVM algorithm is used to develop the model for diabetes disease prediction which has higher accuracy than the decision tree algorithm.

Logistic Regression algorithm used to develop the model for heart disease prediction which has higher accuracy than the KNN and random forest algorithm.

Logistic Regression algorithm used to develop the model for lung cancer disease prediction which has higher accuracy than the KNN and SVM algorithm.

	KNN	LR	RF/DT	SVM
Diabetes Disease	-	-	68%	78%
Heart Disease	72%	88%	82%	-
Lung Cancer Disease	87%	90%	-	88%

8. FUTURE SCOPE

The model has the possibility to provide a more detailed and precise framework in the future, thereby contributing to the development of an improved human disease prediction model. In addition to the current diseases (diabetes disease prediction, heart disease prediction, lung cancer disease prediction), we may also include a few additional diseases for prediction.

REFERENCES:

1. www.ieeexplore.ieee.org
2. www.sciencedirect.com
3. https://docs.streamlit.io/
4. Breiman L. Random forests. MachLearn.

5. https://docs.python.org
6. https://dl.acm.org/
7. https://www.irejournals.com
8. https://streamlit.io/
9. Corinna Cortes and Vladimir Vapnik (1995). Support-vector networks. Machine Learning, 20(3), 273-297.
10. Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). Applied Logistic Regression (3rd ed.). John Wiley & Sons.
11. Zhang, Y., & Ghorbani, A. (2019). A review on machine learning algorithms for diagnosis of heart disease. IEEE Access, 7, 112751-112760.
12. www.kaggle.com