MULTI DISEASE PREDICTION USING MACHINE LEARNING

[1]G.SANGEETHA, [2]Ms. K.MEENATCHI,

[1]MCA, [2]ASSISTANT PROFESSOR,
A.R.J COLLEGE OF ENGINEERING AND TECHNOLOGY,MANNARGUDI.

Abstract-
Data mining for healthcare is an interdisciplinary field of study that has its origins in database statistics and may be used to analyses the success of medical treatments. Many present machine learning models for health care analysis focus on a particular condition at a time. For example, one analysis may be for diabetes, another for thyroid problems, and still another for cancer illnesses. There is no general method for forecasting many illnesses using a single analysis. This project provides a system that forecasts various illnesses using Python Flask API. This inquiry made use of diabetes analysis, thyroid analysis, and breast cancer analysis. To implement various sickness analysis, machine learning algorithms, tensor flow, and the Flask API were employed. Pickling in Python is used to store model behavior, whereas unpicking in Python is used to load the pickle file. The relevance of this research is that it evaluates disorders and includes all of the characteristics that cause the condition, allowing the disease's maximum impact to be detected. We undertake a thorough search of all known feature variables within the KAGGLE dataset to construct models for cardiovascular, prediabetes, and diabetes identification. Using several time-frames and feature sets for the data (based on laboratory data), the Support Vector Machine algorithm is used to forecast illnesses with greater accuracy.
.

I.INTRODUCTION

Multiple disease prediction using Support Vector Machine (SVM) algorithm is a popular and effective technique in the field of healthcare and medical informatics. SVM is a supervised machine learning algorithm that can be used for classification and regression tasks. In multiple disease prediction, SVM is used to predict the presence or absence of multiple diseases simultaneously. This approach can help healthcare professionals make more accurate and efficient diag Nnoses, as well as provide early detection and prevention of diseases. The SVM algorithm works by finding a hyper plane that maximally separates the data points into different classes. In the case of multiple disease prediction, the algorithm is trained on a dataset that includes information about the presence or absence of multiple diseases in a patient, as well as other relevant clinical features such as age, gender, and medical history. Once trained, the SVM model can be used to predict the likelihood of multiple diseases in a new patient based on their clinical data. The output of the SVM model is a probability score for each disease, indicating the likelihood of the patient having that particular disease. Multiple disease prediction using SVM algorithm has several advantages over traditional diagnostic methods, such as reducing the risk of misdiagnosis and improving the accuracy of diagnoses. It can also help identify high-risk patients for certain diseases and provide personalized treatment plans. Overall, this approach has the potential to greatly improve healthcare outcomes and reduce healthcare costs.

II.SYSTEM ARCHITECTURE

A system architecture or systems architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system. System architecture can comprise system components, the externally visible properties of those components, the relationships (e.g. the behavior) between them. It can provide a plan from which products can be procured, and systems developed, that will work together to implement the overall system. There have been efforts to formalize languages to describe system architecture; collectively these are called architecture description languages (ADLs).
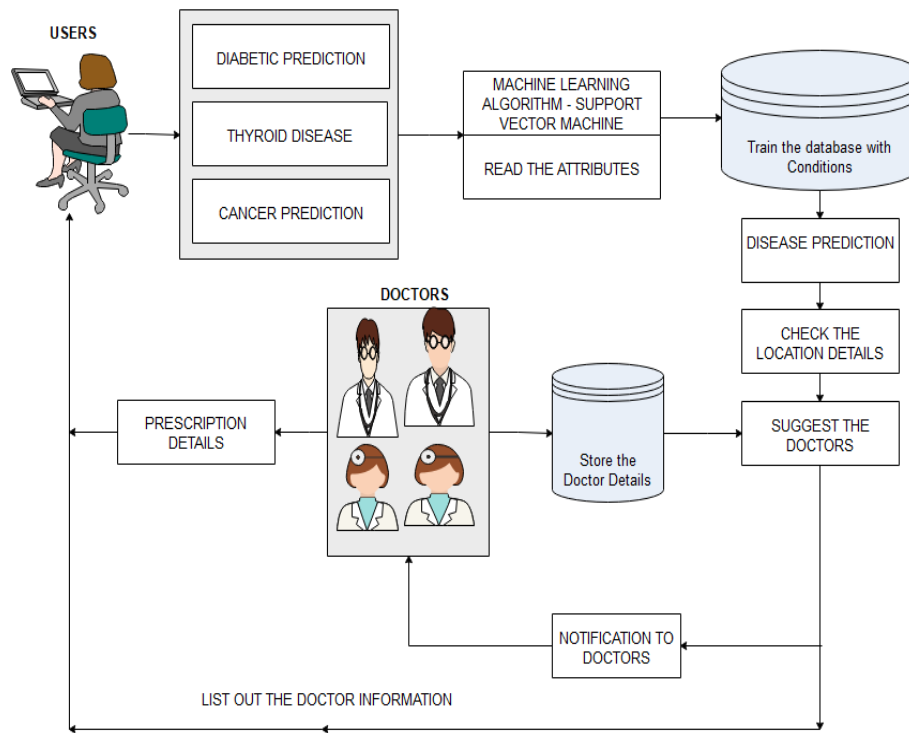
Fig.2a. **System architecture**

Input integrity controls include a number of methods to eliminate common input errors by end-users. They also include checks on the value of individual fields; both for format and the completeness of all inputs. Audit trails for data entry and other system operations are created using transaction logs which gives a record of all changes introduced in the database to provide security and means of recovery in case of any failure. The design of output is the most important task of any system. During output design, developers identify the type of outputs needed, and consider the necessary output controls and prototype report layouts. Manufacturers create and design external outputs for printers. External outputs enable the system to leave the trigger actions on the part of their recipients or confirm actions to their recipients. Some of the external outputs are designed as turnaround outputs, which are implemented as a form and re-enter the system as an input. Output integrity controls include routing codes to identify the receiving system, and verification messages to confirm successful receipt of messages that are handled by network protocol. Printed or screen-format reports should include a date/time for report printing and the data. Multipage reports contain report title or description, and pagination. Pre-printed forms usually include a version number and effective date

III. SYSTEM IMPLEMENTATION

Implementation is the stage in the project where the theoretical design is turned into a working system. The most critical stage is achieving a successful system and in giving confidence on the new system for the users, what it will work efficient and effectively. It involves careful planning, investing of the current system, and its constraints on implementation, design of methods to achieve the changeover methods. The implementation process begins with preparing a plan for the implementation of the system. According to this plan, the activities are to be carried out in these plans; discussion has been made regarding the equipment, resources and how to test activities. The coding step translates a detail design representation into a programming language

realization.Programming languages are vehicles for communication between human and computers programming language characteristics and coding style can profoundly affect software quality and maintainability. The coding is done with the following characteristics in mind. Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

IV.FRAMEWORK CONSTRUCTION

In this module, create the login for admin and user login. Admin can upload the datasets related to heart and diabetic disease. A data set (or dataset, although this spelling is not present in many contemporary dictionaries like Merriam-Webster) is a collection of data. Most commonly a data set corresponds to the contents of a single database table, or a single statistical data matrix, where every column of the table represents a particular variable, and each row corresponds to a given member of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set. Each value is known as a datum. The data set may comprise data for one or more members, corresponding to the number of rows. The term data set may also be used more loosely, to refer to the data in a collection of closely related tables, corresponding to a particular experiment or event. In this module, we can upload the datasets related to diabetic, thyroid and cancer diseases which includes the attributes such as age, gender, height, weight, systolic blood pressure, diastolic blood pressure, cholesterol, glucose, smoke, alcohol, active status, cardio labels. Data pre-processing is an important step in the [data mining] process. The phrase "garbage in, garbage out" is particularly applicable to data mining and machine learning projects. Data-gathering methods are often loosely controlled, resulting in out-of-range values, impossible data combinations, missing values, etc. Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data is first and foremost before running an analysis. If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult. Data preparation and filtering steps can take considerable amount of processing time. In this module, we can eliminate the irrelevant values and also estimate the missing values of data. Finally provide structured datasets.

V.CONCLUSION

The use of data mining in medical data analysis is an excellent way to consider the existing correlations between variables. We've proven that mining can help us find useful connection even when the traits we're looking for aren't direct indications of the class we're trying to forecast. In our research, we attempted to forecast the likelihood of developing system for predicting diabetic, thyroid and cancer disease datasets, and we show that the proposed system provide improved accuracy rate in disease prediction. This type of classifier can aid in the early detection of a diabetic patient's and future prediction. Patients can be forewarned to adjust their lifestyle this way. This will result in preventing from multiple diseases there by resulting in low mortality rates as well as reduced cost on health for the state. SVM's have proven to be a classification technique with excellent predictive performance and also been investigated with the help of ROC curve for both training and testing data. Hence this SVM model can be recommended for the classification of the diseases and recommend the doctors based on disease predictions. And also, extend the framework to implement multiple diseases and recommend the diagnosis information like doctor suggestion, prescription and so on.

REFERENCES:
1. Van Rossum, Guido, and Fred L. Drake. The python language reference manual. Network Theory Ltd., 2011.
2. Van Rossum, Guido, and Fred L. Drake. The python language reference manual. Network Theory Ltd., 2011.
3. Dierbach, Charles. Introduction to Computer Science using Python: A Computational Problem-Solving Focus. Wiley Publishing, 2012.
4. James, Mike. Programmer's Python: Everything is an Object Something Completely Different. I/O Press, 2018.
5. Reges, Stuart, Marty Stepp, and Allison Obourn. Building Python Programs. Pearson, 2018.