MULTIPLE DISEASE PREDICTION USING MACHINE LEARNING AND STREAM-LIT

Haindavi Kothapeta*, Sowmya Lakkampelly**, Arun Mandari***, Mr. M. Sathyanarayana****, Dr. G. Vani*****

*1,2,3Computer Science Sreenidhi Institute Of Science And Technology (SNIST) Hyderabad, India.
****4,5Assistant Professor Department Of CSE Sreenidhi Institute Of Science And Technology (SNIST)Hyderabad, India.

ABSTRACT

Many of the existing machine learning models for health care analysis are concentrating on one disease per analysis. Like one analysis if for diabetes analysis, one for Heart analysis, one for Parkinson's disease like that. There is no common system where one analysis can perform more than one disease prediction. In this article proposing a system which used to predict multiple diseases by using Machine Learning and Stream-lit. In this article used to analyse Diabetes analysis, Parkinson's Disease, Heart Disease detection. Later other diseases like skin diseases, fever analysis and many more diseases can be included. To implement multiple disease analysis used machine learning algorithms, stream-lit and Python pickling is used to save the model behaviour and python unpick-lit is used to load the pickle file whenever required. The importance of this article analysis in while analyzing the diseases all the parameters which causes the disease is included so it possible to detect the maximum effect which the disease will cause. For example, for diabetes analysis in many existing systems considered few parameters like age, sex, Bmi, insulin, glucose, blood pressure, diabetes pedigree function, pregnancies, considered in addition to age, sex, Bmi, insulin, glucose, blood pressure, diabetes pedigree function, pregnancies included serum creatinine, potassium, Glasgow Coma Scale, heart rate/pulse Rate.

Keywords: Machine Learning, Python, Stream-Lit, Multiple Diseases, Algorithms.

I. INTRODUCTION

During analysis of many existing systems looked at only one disease at a time in health analysis. When any organization wants to analyze patient health reports, it has to implement many models. The current system approach is only useful for analyzing a specific disease. Now the daily death rate has increased because the exact disease has not been identified. Even a patient cured of one disease may suffer from another disease. Some existing systems used few parameters to analyze the disease. As a result, it may not be possible to detect diseases due to the effects of the disease. This project looked at diabetes analysis, heart disease and Parkinson's disease detection data. Many other diseases like skin diseases, fever related diseases and many others may be included in future. This analysis is flexible and included many diseases to analyze. When adding a new disease analysis to this existing model, the developer must add the model file associated with the new disease analysis. When developing a new disease, the developer must prepare a python extraction to preserve the behavior of the model. Using this Flow, a developer can upload a pickled file to recreate the behavior of the model. If the user wants to analyze the patient's health, either to predict a specific disease, or if the report contains parameters used to predict other diseases, this analysis allows the best possible detection of the respective diseases.

In our project, we use employee data available on Kaggle and train our machine learning models with SVM validation method using 80% 20% data splits. Our study uses the SVM machine learning algorithm so that we can choose the most accurate model among all and compare their accuracy.

II. LITERATURE SURVEY

Early diagnosis of Parkinson's disease is difficult to achieve with treatment and description of clinical features, including motor symptoms. Machine learning approaches to the classification of Parkinson's disease and healthy controls or matched clinical patients have been used to overcome these problems and improve the diagnosis and evaluation of Parkinson's disease. The main goal of the proposed system is to overcome the...
limitations of the current system and create a system that is accurate enough to diagnose Parkinson's disease in its early stages.

Hearing loss is complex and claims many lives each year. If the early side effects of heart disease are ignored, the patient can experience unusual consequences in a short period of time. The constant lifestyle and overwhelming worry in this day and age have made the situation worse. If diseases are detected early, it can be well monitored.

In this study, we excluded research articles in languages other than English and results published in the form of conference abstracts, posters and lectures. Despite the ongoing debate about the merits and importance of including conference abstracts in systematic reviews, conference abstracts often do not provide enough relevant information and therefore we had to exclude them. While high diagnostic accuracy for Parkinson's disease, diabetes and heart disease has been achieved in clinical settings, machine learning approaches have also achieved high accuracy, as shown in this study, while models incorporating support vector machine and neural networks are particularly useful.

(a) Diagnose Parkinson’s disease, heart disease and diabetes using information methods neglected in clinical decision making. b) identifying very significant features of these data.

1. According to the magazine, diabetes is one of the most dangerous diseases in the world, it can cause various disorders which includes darkness etc. In this paper, they used machine learning methods to detect diabetes as such, it is easy and flexible to predict whether a patient is sick or not. Their goal in this analysis was to develop a system that helps the patient identify a patient's diabetes with accurate results. Here they mainly used 4 main algorithms Decision Tree, Naive Bayes and SVM and compared their accuracy which is 85%, 77%, 77.3% respectively. They also used the ANN algorithm after training to see the reactions of the network whether the disease was correctly classified or not. Here we compared accuracy recovery and F1 score support and accuracy of all models[1].

2. The main purpose of the work is that the heart plays an important role in living organisms. This means that the diagnosis and prognosis of heart-related diseases must be complete and correct, because this is very important, which can lead to cardiac death. So machine learning and artificial intelligence support the prediction of all kinds of natural events. In this paper, they calculate the accuracy of machine learning to predict heart disease using k-nearest neighbor, decision trees, linear regression and SVM using the UCI archive data for training and testing. They also compared the algorithm and their accuracy SVM 83%, decision tree 79%, linear regression 78%, k-nearest neighbor 87%[2].

3. The system specifies that liver diseases cause many deaths in India and are also considered as a life-threatening disease in the world. Because liver diseases are difficult to detect in the early stages, use automatic programs using machine learning algorithms, we can accurately detect liver diseases. They used and compared decision trees using SVM and Random Forest algorithms, and measured precision, accuracy and recall metrics for quantitative measurement. Accuracy is 95%, 87% and 92%[3].

III. PROBLEM STATEMENT

Many of the existing machine learning models for health care analysis is focusing on one disease per analysis. For example first is for heart disease analysis, one for diabetes analysis, one for Parkinson’ diseases like that. If a user wants to predict more than one disease, he has to go through different platforms. There is no common system which can perform analysis on more than one disease on one platform. Some of the models have lower accuracy which can seriously affect patients’ health. When a doctor wants to analyse their patient’s health reports, they have to use many models which is turn into increase in cost as well as time.

IV. PROPOSED SYSTEM

When predicting multi-disease models, it is possible to predict more than one disease at a time. Thus, the user does not need to go through many models to predict diseases. It shortens the time, and due to the prediction of several diseases, there is an opportunity to reduce mortality. It takes less time than the existing system and has more advantages than the current system.

V. SYSTEM REQUIREMENT

The terms "systems analysis" and "requirements analysis" are interchangeable. It can also be used to help
someone (the decision maker) identify a better course of action and make a better decision than they would otherwise. This process requires brainstorming and breaking down the system to look at the scenario, assess the project goals and clarify what needs to be built and used to involve people to define clear requirements.

- **SOFTWARE REQUIREMENTS**
  - Operating System: Microsoft Windows / Linux / Mac-OS
  - Technology: Machine Learning
  - Tools: Jupyter Notebook or Spyder
  - Platform: Anaconda Distribution or Google Collab.

- **HARDWARE REQUIREMENTS**

- **PROCESSOR**
  - Any Intel or AMD x86 - 64bit processor is required as a minimum.
  - Any Intel or AMD CPU with at least 4 logical cores and hyper-threading support is recommended. Intel i5 8th generation equivalent or higher is preferred.

- **RAM**
  - Minimum: 8 GB DDR4 RAM
  - Recommended: 8 - 16 GB DDR4 RAM or 8 GB DDR5 RAM

VI. **SYSTEM DESIGN**

The process of defining the armature, factors, modules, interfaces, and data so that the system meets certain criteria is called system design. This is the operation of systems proposition to product development, in a nutshell. Object-acquainted design and analysis styles are snappily getting the most popular ways for erecting computer systems.

**Description:** System design is the process of defining system rudiments similar as modules, armature, factors and their interfaces and data grounded on the conditions defined for the system. It's the process of defining, developing and designing systems that meet the specific requirements and conditions of a business or association. Description A harmonious and well-performing system requires a methodical approach. A top-down or nethermost-up approach is needed to consider all applicable system variables. A developer uses modeling languages to express information and knowledge in a methodical structure defined by harmonious rules and delineations. Plans can be defined using visual or textual modeling languages.

Some exemplifications of graphical modeling languages are

a) Unified Modeling Language( UML) describing software both structurally and functionally with graphic memorandum.

b) Flowchart a schematic or step-by-step representation of an algorithm.


d) Design styles
  1) Architectural design describes the views, models, geste and structure of the system.
  2) Logical design represents the information inflow, inputs and labors of the system. illustration ER plates( reality relationship plates).
  3) Physical design defined as:
  a) how druggies add information to the system and how the system returns information to the stoner.
  b) How data is modeled and stored in the system.
  c) how data flows through the system, how data is validated, defended and or converted as it flows through and out of the system.

**ARCHITECTURE DESIGN**

Architectural design in software engineering is about decomposing the system into interacting factors. It's expressed as a block illustration defining an overview of the system structure, features of the factors, and how these factors communicate with each other to partake data. It identifies the factors that are necessary for
developing a computer-predicate system and communication between them i.e. relationship between these factors. It defines the structure and parcels of the factors that are involved in the system and also the non-intercourses between these factors. The architectural design process is about relating the factors i.e. subsystems that makeup the system and structure of the sub-system and they’re commerce. It’s an early stage of the system design phase. It acts as a link between specification conditions and the design process.

**Fig-1: Architecture Design**

**Data Flow model**

A data inflow illustration takes business processes and conditioning and uses them to produce a clear illustration of how data flows through a system. DFDs represent the inflow of data from external realities into a single system by moving and storing data from one process to another. Through the use of data inflow plates, a system can be perished into subsystems, and subsystems can be farther perished into lower-position subsystems. Each subsystem represents a process or exertion in which data is reused. Once the smallest position is reached, processes can no longer be perished. Data inflow modeling can be used to identify a variety of different effects, similar as Information that’s entered from or transferred to other individualities, associations, or other computer systems. Areas within a system where information is stored and the overflows of information within the system are being modeled. The processes of a system that act upon information entered and produce the performing labors.

**Fig-2: Data Flow Model**

**VII. IMPLEMENTATION**

**SUPPORT VECTOR MACHINE (SVM)**

Support Vector Machine or SVM is one of the most popular supervised literacy algorithms used to break bracket and retrogression problems. still, it’s substantially used to break machine literacy bracket problems. The purpose of the SVM algorithm is to produce the stylish line or decision boundary that can separate the n-dimensional space into classes so that we can fluently place a new data point into the correct class in the future.
This stylish decision boundary is called the hyperplane. SVM selects the extreme points vectors that help produce the hyperplane. These extreme cases are called support vectors and thus the algorithm is called a support vector machine. Consider the illustration below with two different classes classified using a decision boundary or hyperplane:

**Fig-3: Support Vector Machine Algorithm**

First, support vector machines (SVM) were used for bracket. still, SVMs have presently been used in numerous problems, including pattern recognition, bioinformatics, and textbook bracket. thus, SVM bracket is performed by applying a direct or non-linear separation face. But it can be observed that training SVM requires working a quadratic optimization problem. SMO classifier with SVM is used to achieve the asked result. Platte proposed this algorithm called SMO to design an SVM classifier.

**SVM CAN BE OF TWO TYPES**

**LINEAR SVM**
Linear SVM is used for linearly divisible data, which means if a set of data can be classified into two classes using a single straight line, similar data is called linearly divisible data and the classifier is called a direct SVM classifier.

**NON-LINEAR SVM**
Non-linear SVM is used for non-linearly distributed data, which means that if a set of dates can not be classified in a straight line, similar data is called non-linear data and the classifier used is an non-linear SVM classifier.

**HYPERPLANE**
There can be multiple rows decision boundaries in an n-dimensional space to separate classes, but we need to find the stylish decision boundary to help classify the data points.

- This stylish bound is known as the SVM hyperplane.
- The confines of the hyperplane depend on the features in the data set, which means that if there are 2 features (as shown), the hyperplane is a straight line.
- And if there are 3 parcels, also the hyperplane is a 2-dimensional aeroplane

**SUPPORT VECTORS**
The data points closest to the hyperplane or the vectors that affect the position of the hyperplane are called support vectors. Since these vectors support the hyperplane, they’re thus called support vectors.
LINEAR SVM

Linear SVM: Linear SVM is used for linearly divisible data, which means if a set of data can be classified into two classes using a single straight line, similar data is called linearly divisible data and the classifier is called a direct SVM classifier. The working of the SVM algorithm can be understood with the help of an illustration. Suppose we've a set of dates with two markers (green and blue) and the set of dates has two parcels x1 and x2. We want a classifier that can classify a brace of equals (x1, x2) as either green or blue. Consider the image below.

[Image: Fig-4: Linear Support Vector Machine]

NON-LINEAR SVM:

Non-linear SVM: Non-linear SVM is used for non-linearly distributed data, which means that if a set of dates can not be classified in a straight line, similar data is called non-linear data and the classifier used is non-linear. For direct data, direct SVM classifier. still, we can separate them with a straight line, but for non-linear data, if the data is arranged linearly. To separate these data points, we need to add another dimension. For direct data we used two confines x and y, so for non-linear data we add a third dimension z. This can be calculated as

FORMULA:

\[ Z = (x^2) + (y^2) \]

[Image: Fig-5: Non-Linear Support Vector Machine]

By adding the third dimension, the sample space will become as below image. So now, SVM will divide the datasets into classes in the following way.

[Image: Fig-6]
VIII. RESULTS

In the system we have used SVM algorithm for the prediction. When the patient will input the values in system according to that it will show whether the patient has a disease or not. The parameters will show the range of the values needed. The entered value is not between the ranges or is not valid or is empty it will show the warning sign that add a correct value.

It is based on how accurate outputs are retrieved through support vector machine algorithm. Support Vector Machine Algorithm mainly useful for linear data.

In this project we divided the data into two datasets:

- Training data
- Testing data

The delicacy handed by SVM is about 88 on training data and 87 on test data. Once the Flask API is created. The model can be worn in the front. Secured by designing a template website. When the stoner clicks Get Heart Disease Patient Statue, it shows whether the case has the complaint or not.

1. Homepage

2. Diabetes Disease
3. Heart Disease

A multi-disease prediction model is used to prognosticate multiple conditions contemporaneously. Then complaint is prognosticated grounded on stoner input. The choice is given to the user. However, the corresponding complaint model is erected and prognosticated grounded on the inputs entered by the stoner. If the stoner wants to prognosticate a complaint or if the stoner doesn't enter any complaint. The advantage of multi-disease prediction models is that it's possible to prognosticate the probability of circumstance of different conditions and reduce the death rate. The use of different ML algorithms enabled the early discovery of numerous conditions similar as heart, order, casket and brain conditions. Algorithms from the SVM, RF, and LR literature were substantially used because prediction delicacy was the most used performance measure. The CNN model has proven to be the most suitable for prognosticating common conditions. In addition, the SVM model showed advanced delicacy substantially for high-dimensional, semi-structured and unshaped order complaint and PD trust-ability data. bone cancer prophet RF showed a better probability of rightly classifying conditions due to its veritably large dataset and preceptivity to avoid over-fitting. Eventually, the LR algorithm proved to be the most dependable in prognosticating heart complaint.

4. Parkinson's Disease

IX. CONCLUSION

A multi-disease prediction model is used to prognosticate multiple conditions contemporaneously. Then complaint is prognosticated grounded on stoner input. The choice is given to the user. However, the corresponding complaint model is erected and prognosticated grounded on the inputs entered by the stoner. If the stoner wants to prognosticate a complaint or if the stoner doesn't enter any complaint. The advantage of multi-disease prediction models is that it's possible to prognosticate the probability of circumstance of different conditions and reduce the death rate. The use of different ML algorithms enabled the early discovery of numerous conditions similar as heart, order, casket and brain conditions. Algorithms from the SVM, RF, and LR literature were substantially used because prediction delicacy was the most used performance measure. The CNN model has proven to be the most suitable for prognosticating common conditions. In addition, the SVM model showed advanced delicacy substantially for high-dimensional, semi-structured and unshaped order complaint and PD trust-ability data. bone cancer prophet RF showed a better probability of rightly classifying conditions due to its veritably large dataset and preceptivity to avoid over-fitting. Eventually, the LR algorithm proved to be the most dependable in prognosticating heart complaint.
X. FUTURE SCOPE

Farther exploration could explore the possibility of combining a support vector machine algorithm to prize
prophetic features with unsupervised literacy. Likewise, only healthy cases were used in this analysis. A
machine literacy model with analogous perfection, delicacy and recall may be possible for a model that
considers multiple affiliated judgments. More complex ML algorithms will be created in the future, much is
demanded to ameliorate complaint prediction.

In addition, learning models should be calibrated more frequently after the training phase to achieve better
results. In addition, datasets should be expanded to include different population groups to avoid over-fitting
and ameliorate the data delicacy of applied models. Eventually, more important selection styles should be used
to ameliorate performance. In the future, we can add to the being complaint API. To reduce mortality, we can
try to ameliorate the delicacy of the cast. Try that the system is friendly and offers standard converse.

ACKNOWLEDGMENT

We sincerely thanks to our college Sreenidhi Institute of Science and Technology, Hyderabad for giving us a
platform to prepare a project on the topic "Multiple Disease Prediction using machine Learning and stream-lit".
We are sincerely grateful for Mr.M.SathyaNarayana as our guide and Dr.G.Vani for providing help during our
research, which would have seemed difficult without their motivation, constant support, and valuable
suggestion.

XI. REFERENCES

learning,” in 2018 Second International Conference on Electronics, Communication and Aerospace

processing and machine learning,” in 2017 International Conference on Intelligent Sustainable Systems
(ICISS), 2017, pp. 1047–1051.

algorithms for disease prediction,” BMC Medical Informatics and Decision Making, vol. 19, no. 1, pp. 1–
16, 2019.

2020 International Conference on Electronics and Sustainable Communication Systems (ICESC),
2020, pp. 302–305.


disease prediction using support vector machine,” in 2018 9th International Conference on Computing,

2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2016,
pp. 1903–1907.

disease diagnosis,” in 2015 IEEE Canada International Humanitarian Technology Conference
(IHTC2015), 2015, pp. 1–3.

approach,” Proceedings of the 3rd International Conference on Computing Methodologies and