IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 3, March 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1203118 |

Multiple Disease Detection

B V N S Ganesh, Jothi Lakshmi

Department of Information Technology, R.M.D. Engineering College, Chennai, India

Department of Information Technology, R.M.D. Engineering College, Chennai, India

ABSTRACT: Globally, there is a substantial unmet need to diagnose various diseases effectively. The complexity of the different disease mechanisms and underlying symptoms of the patient population presents massive challenges to developing the early diagnosis tool and effective treatment. Machine Learning (ML) an area of Artificial Intelligence (AI), enables researchers, physicians, and patients to solve some of these issues. Based on relevant research, this review explains how Machine Learning (ML). Early detection and diagnosis of diseases such as diabetes, chronic kidney disease, liver disease, and breast cancer is crucial for improving patient outcomes and reducing healthcare costs. In recent years, machine learning has emerged as a promising tool for disease detection and diagnosis. In this project, we aim to develop a machine learning model for multiple disease detection, which can aid in early disease diagnosis and treatment. The proposed model will be trained on a large dataset of medical records, which includes patient demographics, medical history, symptoms, and diagnostic test results. The dataset will be carefully curated and pre-processed to ensure high data quality and completeness. We will also incorporate relevant features such as genetic markers and lifestyle factors, which have been shown to influence disease risk.

KEYWORDS: Machine Learning, Decision tree, Adaboost , Xgboost and Catboost and ML techniques, evaluation.

I. INTRODUCTION

Multiple Disease Detection is an emerging technology that aims to detect the presence of multiple diseases simultaneously. The technology utilizes advanced data analytics and machine learning algorithms to analyze data from multiple sources, such as medical records, lab reports, and imaging studies, to provide a comprehensive diagnosis of multiple diseases.The technology has shown great promise in the early detection and management of several chronic diseases, including Diabetes, Chronic Kidney Disease, Liver Disease, and Breast Cancer. These diseases are among the leading causes of morbidity and mortality worldwide, and early detection and intervention can significantly improve patient outcomes.Diabetes is a chronic disease characterized by high levels of blood sugar. Early detection of diabetes can prevent complications such as blindness, kidney disease, and nerve damage. Chronic Kidney Disease is a progressive disease that can lead to kidney failure if left untreated. Early detection and management can slow down the progression of the disease and prevent kidney failure.Liver Disease is a broad term used to describe any condition that affects the liver. Early detection of liver disease can prevent liver damage and improve liver function. As technology continues to advance, we can expect to see even more sophisticated tools and techniques for detecting and managing multiple diseases. Within the aging population, the frequency of cancer is increasing dramatically. In addition, multiple genetic and environmental factors lead to common multifactorial diseases, including cardiovascular disease, chronic kidney disease, chronic obstructive pulmonary disease, and metabolic-associated fatty liver disease. In recent years, there has been a growing awareness of the connection between cancer and multifactorial diseases, as well as how one can affect the other, resulting in a vicious cycle. Although the exact mechanistic explanations behind this remain to be fully explored, some progress has been made in uncovering the common pathologic mechanisms. In this review, we focus on the nature of the link between cancer and common multifactorial conditions, as well as specific shared mechanisms, some of which may represent either preventive or therapeutic targets. Rather than organ-specific interactions, we herein focus on the shared mechanisms among the multifactorial diseases, which may explain the increased cancer risk. More research on this subject will highlight the significance of developing new drugs that target multiple systems rather than just one disease.

II. LITERATURE SURVEY

It is facilitated to access the performance and information of attendance of a particular Student in a particular semester of study. The information is sorted by the teachers, instructors and advisors, as provided by the student for a particular

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1203118 |

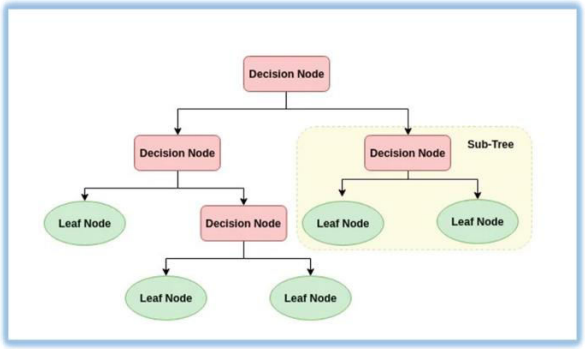
day throughout a complete semester.

This system will also enable the evaluation of student regular presence in various lectures which will determine the eligibility of the student to sit for a semester examination.

This user can receive alert / message from his teachers related to his attendance performance. It is obvious that Students with poor Attendance Management System attendance will see their attendance in a Red warning table that can make the student to be careful not to miss classes anymore. [1]Multiple disease detection model using machin learning. The importance of this article analysis in while analyzing the disease all the parameters which causes the disease is included so it possible to detect the maximum effects which the disease will cause.[2] Diabetes Disease Prediction Using Machine Learning and Flask. This paper deals with the prediction of diabetes disease by performing an analysis of five supervised machine learning algorithms[3] Disease prediction using machine learning. The wide adaptation of computer-based technology in the health care industry resulted in the accumulation of electronic data.medical doctors are analyze symptoms accurately and identify disease at an early stage.

Algorithms:

There are various Algorithm that can be used for multiple disease detection, but one of the most popular and effective methods is the Decision tree.Here is the block diagram for the method:



Decision tree is a flowchart-like tree structure where an internal node represents feature(or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree in recursively manner call recursive partitioning. This flowchart-like structure helps you in decision making. It's visualization like a flowchart diagram which easily mimics the human level thinking. That is why decision trees are easy to understand and interpret. The basic idea behind any decision tree algorithm is as follows:[1] Select the best attribute using Attribute Selection Measures (ASM) to split the records.[2]Make that attribute a decision node and breaks the dataset into smaller subsets.[3]Starts tree building by repeating this process recursively for each child until one of the conditions will match:

All the tuples belong to the same attribute value.

There are no more remaining attributes.

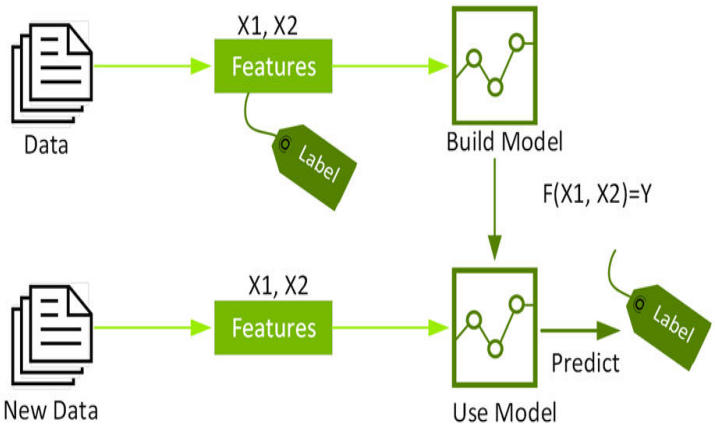
There are no more instances.

Extreme Gradient Boosting (XGBoost) is a scalable, distributed gradient-boosted decision tree (GBDT) machine learning framework. It is the top machine learning package for regression, classification, and ranking tasks, and it supports parallel tree boosting.To understand XGBoost, you must first understand the machine learning ideas and methods on which it is based: supervised machine learning, decision trees, ensemble learning, and gradient boosting.Supervised machine learning use algorithms to train a model to detect patterns in a dataset containing labels and features, and then employs the trained model to predict the labels on the features of a new dataset.

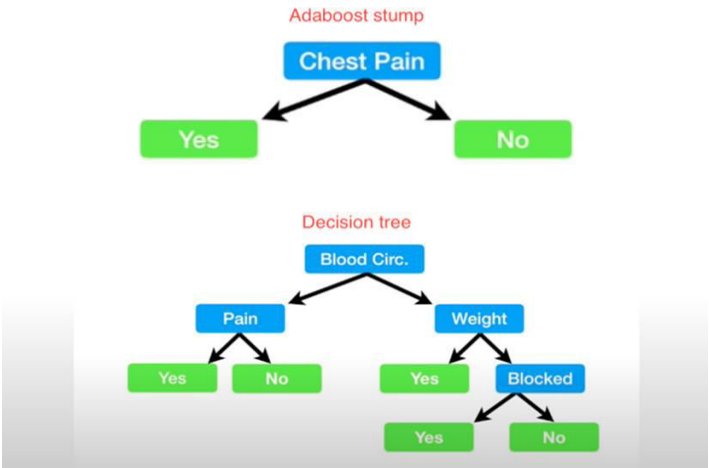
| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1203118 |



AdaBoost also called Adaptive Boosting is a technique in Machine Learning used as an Ensemble Method. The most common algorithm used with AdaBoost is decision trees with one level that means with Decision trees with only 1 split. These trees are also called Decision Stumps. AdaBoost algorithm, short for Adaptive Boosting, is a Boosting technique used as an Ensemble Method in Machine Learning. It is called Adaptive Boosting as the weights are re-assigned to each instance, with higher weights assigned to incorrectly classified instances. Boosting is used to reduce bias as well as variance for supervised learning. It works on the principle of learners growing sequentially. Except for the first, each subsequent learner is grown from previously grown learners. In simple words, weak learners are converted into strong ones. The AdaBoost algorithm works on the same principle as boosting with a slight difference. Let's discuss this difference in detail.



CatBoost or Categorical Boosting is an open-source boosting library developed by Yandex. In addition to regression and classification, CatBoost can be used in ranking, recommendation systems, forecasting and even personal assistants.

MODULES DESCRIPTION:

[1]Data Collection. Data Source: - Liver patient CSV file.
Dataset Features: - Parameters taken for liver disease diagnosis are Age, Gender, Count of total bilirubin and direct bilurbin, Alkaline phosphotase, Alamine aminotransferase, Aspartate aminotransferase, Albumin, Albumin and globulin.
Learning Algorithms: - Decision tree,adaboost
Flask Platform: - Flask retrieves file using `request.files[]`.
User Interaction: - User login, select the module like liver disease prediction, CSV structure.
[2] Data Pre-Processing. Null Value Handling: - Use forward fill(),backward() to add the null values
Feature Selection: - Exclude less informative features like age, gender.
Variable Handling: - Numerical value using Label Encoder.
Binary Encoding: - Encodes the Status attribute into binary format (1,0)

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1203118 |

[3] Data Splitting. Data Splitting: Utilize `train_test_split()` from the sklearn library. Split data into Train and Test sets. Feature and Label Division: - Divide data into features x_train and y_train data.
Training Data Preparation: - Create X_train and y_train sets for model training. - Iterative process involving fitting and adjusting.
[4] Model Selecting And Traning. AdaBoost: Ensemble learning method that combines weak classifiers to create a strong classifier.To classify whether a given social network user or activity is fraudulent or not
DecisionTree: In this base node represents feature, branch represents a decision rule, and each leaf node represents the outcome. It learns to partition on the basis of the attribute value. It selects best attribute and split the records and make decision node and breaks dataset into smaller subsets. Repeates until no more attributes or no more instances.
SVM: SVM performs binary classification tasks, data from two classes can be separated by hyperplane. This algorithm is considered for highest classification accuracy.
[5] Data Visualization. Visualization: - Utilize Bar graphs, Scatter plot to show the proportion of instances.
Liver disease prediction : - Represent patterns and relationships within the data, helps in understanding key factors for predicting liver disease.
Data Exploration: - Analyzing various types of data to identify patterns and factors associated with liver diseases, helping in understanding the data before building predictive models.

III. RESULT

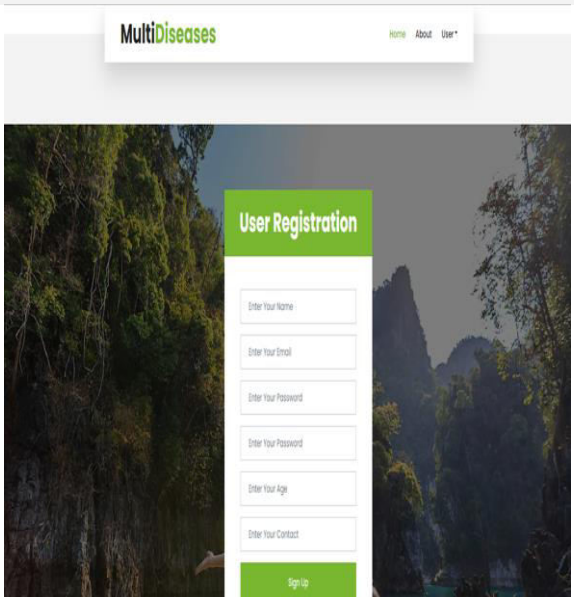
On the home page, you can assign participants to the disease detection system. It will detect the diseases.It will able to find the what disease you have.

[2] On this page, you can create an account for different participants including the manufacturer, third-party hub, and customer, using local accounts. By creating these accounts, you can able to find the what disease you have diabetes are cancer. Each account can be tailored to the specific needs and access levels of each participant.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1203118 |



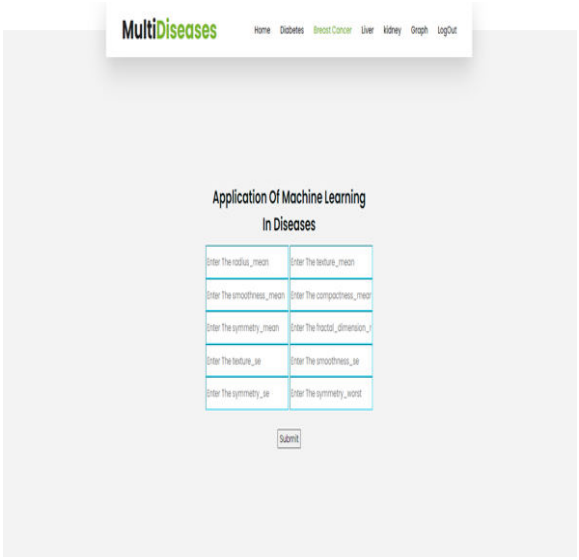
[3]Here we can read about our project.

[4]In this when you fill your information in this application form you will able to dectct chronic diseases or breast cancer etc.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1203118 |



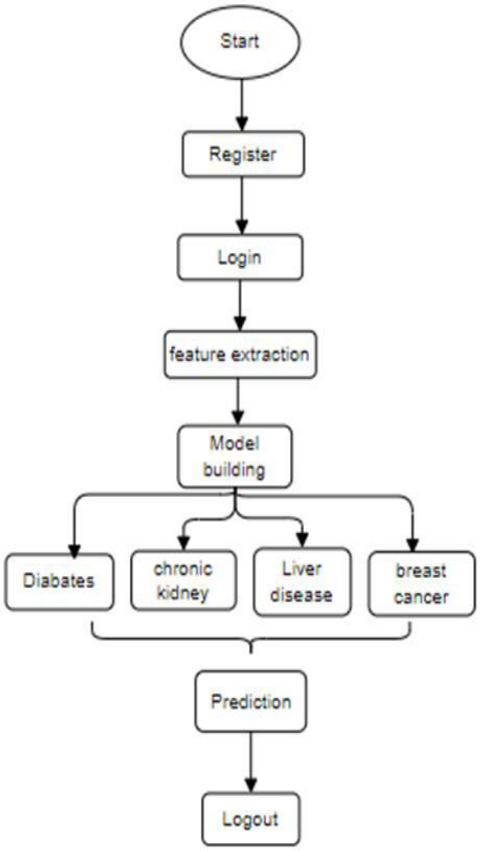
[5]Here we have graph about this disease

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 3, March 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1203118 |

IV. PROJECT FLOW CONTROL



V. CONCLUSION

The integration of advanced diagnostic technologies, including machine learning algorithms and medical data analysis, has shown significant promise in the early detection and management of multiple diseases such as diabetes, chronic kidney disease, liver disease, and breast cancer.

These innovative approaches offer the potential for timely interventions and personalized treatments, ultimately improving patient outcomes and reducing the burden on healthcare systems.

However, ongoing research, robust data collection, and stringent validation are essential to ensure the reliability and effectiveness of these detection methods in real-world clinical settings.

REFERENCES

[1] J. L. Scully, ‘‘What is a disease?’’ EMBO Rep., vol. 5, no. 7, pp. 650–653, 2004.

[2] R. Leaman, R. Islamaj Dogan, and Z. Lu, ‘‘DNorm: Disease name normalization with pairwise learning to rank,’’ Bioinformatics, vol. 29, no. 22, pp. 2909–2917, Nov. 2013.

[3] N. Armstrong and P. Hilton, ‘‘Doing diagnosis: Whether and how clinicians use a diagnostic tool of uncertain clinical utility,’’ Social Sci. Med., vol. 120, pp. 208–214, Nov. 2014.

[4] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, ‘‘Network medicine: A network-based approach to human disease,’’ Nature Rev. Genet., vol. 12, no. 1, pp. 56–68, Jan. 2011.

[5] R. H. Scheuermann, W. Ceusters, and B. Smith, ‘‘Toward an ontological treatment of disease and diagnosis,’’ Summit Transl. Bioinformat., vol. 2009, p. 116, Mar. 2009.

[6] P. Croft, D. G. Altman, and J. J. Deeks, ‘‘The science of clinical practice: Disease diagnosis or patient prognosis? Evidence about ‘what is likely to happen’ should shape clinical practice,’’ BMC Med., vol. 13, no. 1, p. 20, 2015.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details