

Multiple Disease Prediction by Applying Machine Learning and Deep Learning Algorithms

M. Kalpana Chowdary*

Department of CSE
MLR Institute of Technology
Hyderabad, India

dr.kalpana@mlrinstitutions.ac.in

K.Anil Kumar

Department of ECE
Malla Reddy Engineering College
Hyderabad, India

kotthuru.anilkumar@gmail.com

C.Ganesh

Department of CCE
Sri Eshwar College of Engineering
Coimbatore, India

csganesh86@gmail.com

Rajsekhar Turaka

Department of ECE
Nalla Narsimha Reddy Education
Society's Group of Institutions
Hyderabad, India

dr.rajasekhar.turaka@gmail.com

B.Devananda rao

Department of CSE
MLR Institute of Technology
Hyderabad, India

dev.bolleddu@gmail.com

Sk.Lokesh Naik

Department of CSE
MLR Institute of Technology
Hyderabad, India

lokeshnaik@mlrinstitutions.ac.in

Abstract— A large number of machine-learning projects for healthcare inspection concentrate on just one disease at one time. One analysis is performed for each of the following conditions: diabetes, cancer, skin disease, and heart disease. A single analysis cannot forecast more than one disease using a same system. Numerous machine learning approaches can perform predictive analytics on colossal data sets. Although using predictive modeling in healthcare is challenging, it will eventually help practitioners make quick choices about the health and treatment of patients based on vast amounts of data. In this post, a system that uses the Flask API to forecast numerous diseases is proposed. Tensor flow, Flask API, and machine learning techniques were used to implement multiple illness analyses. The model behavior is saved using Python pickling, and the pickle file is loaded when needed using Python unpickling. The goal in this effort is to create a web application that can forecast numerous diseases, such as breast cancer, diabetes, heart disease, malaria, and pneumonia, using the principles of machine learning and deep learning. The finished model's behavior will be captured in a Python pickle file. The Flask API is created. The parameters of the disease must be sent together with the disease name when using this API. The relevant model will be called by the Flask API, which then returns the patient's state. In order to track patient condition and provide early warnings to patients in order to reduce death rates, it is important to identify the most common diseases.

Keywords— Machine Learning algorithms, Flask API (Application programming interface), Python pickling.

I. INTRODUCTION

Malaria, Pneumonia, Diabetes, Breast Cancer, Kidney, Liver, and heart diseases are the main causes of increased mortality rates in today's culture. Instead of "cardiovascular disease," the word "heart disease" is frequently used. The word "cardiovascular disease" typically refers to illnesses that damage your heart's rhythm valves or muscles, as well as conditions that cause heart attacks, angina (heart discomfort), or stroke. A type of cancer that appears in the breast tissue is called breast cancer. The common symptoms of breast cancer are inverted breasts, lumps, nipple discharge, and thickening or puckering of the skin. According to an article in "Times of India" dated August 19, 2020, the most common site for breast cancer cases in India is predicted to be home to about 2.4 lakh cases by 2025. Diabetes develops when your glucose

level is extremely high. The main source of energy comes from blood sugar, which is acquired from the food you intake. Insulin is a hormone secreted by the pancreas that helps fuel metabolic processes by removing glucose from the diet. According to the "International Diabetes Federation", diabetes kills 42 lakh people worldwide every year, and it costs over 760 billion USD (as a part of health expenditure). Epidemiology of Diabetes reports that over 10 lakh people in India lose their lives to diabetes each year, and the "Indian Heart Association" predicts that by 2035, approximately 11 crore people would have the disease. People can be infected with a life-threatening disease like malaria which is caused because of mosquitoes. This disease's symptoms include fever and headache. Malaria can cause death if people don't care about it.

It is treatable and preventable. There will probably be 241 million malaria cases worldwide by the year 2020. This article took into account data sets for diabetes analysis, heart disease detection, liver disease prediction, kidney illness, malaria, pneumonia, and breast cancer detection. In the future, numerous additional illnesses may be added, such as skin conditions, illnesses linked to fever, and many more. This adaptable analysis later encompassed numerous disorders for analysis. Whenever a new disease is found, model file is to be created that links to the study of a new disease. Python pickling is used to save the model behaviour. Using flask API, the pickled file can be loaded to extract the model behaviour. Whenever a person wants to access the patient's health this API can give accurate results. So, to decrease the death ratio.

II. LITERATURE SURVEY

In the paper [1&2], the author focuses on predicting three diseases which are Heart, Diabetes, Brain and Cancer the algorithms used by them are Random-forest, SVM, Logistic-regression, etc. The proposed system used Flask API to predict the diseases. This paper gave us accurate values for a particular disease that works very well. The only limitation is that they may have predicted more diseases using the best algorithms. Chest discomfort, arrhythmia, neck pain, and upper body are the symptoms of heart disease. Under the skin, a lump or thickening area might be felt, and changes in weight, such as unplanned loss or growth are the symptoms

of cancer. Leg pain, urinating more, and thirst are the symptoms of diabetes.

The authors of [3] focused on three different diseases heart, Diabetes, and Breast cancer which can cause serious problems for humans. In today's generation, women are mostly suffering from breast cancer because they are unaware of how this disease is affecting them. They used three algorithms they are KNN, SVM, and Naïve Bayes. Heart disease got an accuracy which is 83% with KNN, Diabetes got an accuracy which is 76% with Naive Bayes, and Breast cancer with an accuracy of 83% with KNN. The limitations of this paper are that they did not use algorithms that may give higher accuracy than the algorithms they have used. But there are some limitations in this model like less accuracy and only limited dis-eases are predicted.

In the paper [4] the main focus is on diabetes, one of the most serious diseases in the world. Diabetes can lead to a variety of issues for people, including heart attacks, blindness, urinary organ issues, numerous illnesses, etc. The paper primarily focuses on diabetes, and to anticipate the disease, they have utilized a variety of machine learning techniques. Their primary goal was to accurately diagnose the condition. Decision Tree, Naive Bayes, SVM, and ANN algorithms were the four used algorithms. Here, they merely utilized a few algorithms and narrowed their emphasis to one ailment. The limitation is that they might have used more algorithms like a random forest which may give better results to predict diabetes.

The author of [5] highlights the importance of the heart in living organisms. As a result, it is critical that heart disease be adequately identified and anticipated since failure to do so can result in deaths. Heart-related illness diagnosis and prognosis require more precision since even little errors might result in deadly fatigue issues. As a result, artificial intelligence and machine learning can help anticipate all types of natural disasters. K-nearest neighbor, decision tree, linear regression, and SVM are some of the machine learning algorithms used to predict heart disease, in which the K-nearest neighbor got the highest in terms of accuracy. But the model has some limitations because using the Random-Forest algorithm heart disease can be predicted with higher accuracy.

The writer of [6] found that the second biggest cause of mortality for women globally is breast cancer, which is the main topic of this paper. Breast cancer is one of the most common cancers in women globally. The widespread issue exists in modern life. In the course of breast cancer, malignant cells may form in the liver, brain, bones, and lungs, which can result in bone pains, headaches, and other symptoms. SVM, Random-Forest, KNN, and Naive Bayes are the four algorithms utilized in this paper, which only focus on the disease of breast cancer. When it came to accuracy, SVM performed best.

The paper [7] mainly focused on diabetes conditions brought on by an increase in plasma sugar levels. A variety of computerized information systems were established for

predicting and diagnosing diabetes employing decision trees, SVM, Naive Bayes, and ANN algorithms as classifiers. The Adaboost algorithm, Support Vector Machine, Naive Bayes, and Decision Tree are the algorithms employed in this study. Where the Adaboost algorithm performed the best.

In the paper [8] the author found that the main reason why diabetic individuals experience vision impairment is diabetic retinopathy (DR). The outcomes of various eye testing techniques and the clinical diagnostic results provide a set of observations that help the doctor and therapist make decisions about diabetic retinopathy. Diabetic retinopathy constitutes the most common and devastating eye consequence, causing retinal damage and possibly blindness. The techniques they used in their paper are SVM, KNN, NB, and DT where SVM gave the highest accuracy.

The author of [9&10] focused on diseases like—diabetes, cerebral infarction, and heart disease—they conducted illness prediction experiments. Using algorithms including Naive Bayes, KNN, and decision trees algorithm to detect heart disease, diabetes, and cerebral infarction. Compared to Naive Bayes and KNN, the Decision Tree algorithm produces superior results. Also, they foretell if a patient will develop a cerebral infarction with a high or low risk. They used multi model illness risk prediction using text data based on CNN for cerebral infarction risk prediction. But the model has some limitations because it can only predict three diseases at a time and not more than that.

III. PROPOSED SYSTEM

As we see that there are multiple different models to predict different diseases but there is no one common system to predict all the diseases at one time in one go. So, the models can predict a different number of diseases in one common system. This can reduce the chances of getting a new disease. As we have seen various machine learning algorithms such as Random Forest Classifier, Decision Tree Classifier, Gradient Boosting Classifier, Logistic Regression, Xgboost, SVM, KNN, and Deep Learning Algorithms such as VGG19 CNN. This model is going to use all these algorithms for the prediction of the disease and then compare them and choose the best one for a particular disease. And this model is also going to give some preventive measures so that the disease may be cured in the early stage.

A. Block diagram

A Machine Learning based user-friendly prediction model of diseases is designed in such a way that it receives the input in the attributes, analyses the data, and predicts if the disease is present. The block diagram describes basically how the model is going to work how the data is taken as input how is divided into training and testing what features are required for particular disease prediction and what algorithms are being applied to the model. Later performance analysis is done to test the model like how the model is working and whether it is giving higher accuracy or not. After doing all this the disease is predicted.

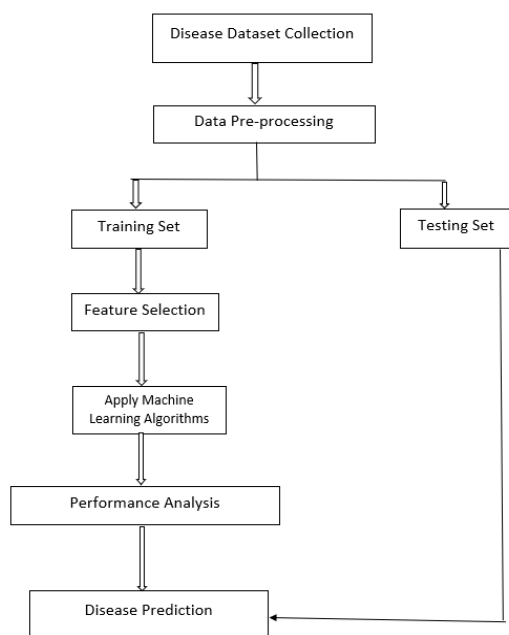


Fig. 1. Block diagram

1) Disease dataset collection

All the disease datasets are taken from Kaggle. The heart disease dataset is collected from UCI, Kaggle which has 13 medical prediction fields and one result/target field. The diabetes dataset is taken from the Pima Indians Diabetes dataset from Kaggle which has 8 medical prediction fields and one target field. Breast cancer is predicted by collecting Breast Cancer Wisconsin dataset which has 31 medical prediction fields and one target field. All datasets are available on the Kaggle site.

2) Data pre-processing

The modifications were done to the data before delivering it to the models known as pre-processing. Using pre-processing methods, the raw data are converted into a set of comprehensible data. To put it another way, anytime information is acquired from several sources in a format that makes analysis impossible. It explains the steps involved in cleaning, transforming, and combining data to prepare it for analysis. Data pre-processing is done to improve the quality of the data. Four steps take place in data pre-processing which are data cleaning, data integration, data transformation, and data reduction.

Data cleaning is used to remove any missing, incorrect, or useless data. This may entail getting rid of extraneous data, filling in blanks, and handling outliers. There are numerous potentials for data duplication or labelling errors when merging multiple data sources. Even if results and algorithms appear to be correct, they are unreliable if the data is inaccurate. However, it is essential to create a template for data cleaning procedure so that it carried out correctly each time.

Data integration involves combining information from several sources, such as databases, spreadsheets, and text files. Integration strives to generate a single, unified picture

of the data. Data transformation converts source data into target data. Without making significant changes to current applications or data structures, well-implemented data integration can lower IT costs, free up resources, enhance data quality, and promote creativity. This brings higher profits to firms.

Data transformation requires converting the data into a more suitable format for data mining. Encoding category data, creating dummy variables, and normalizing numerical information are all possibilities. Data is pulled into a single repository during the extraction phase from various sources or places, therefore it is in raw original form. The retrieved data must be converted into the desired format by going through several procedures to ensure its usability. This also resolves the issues related to missing values in the datasets.

Data reduction is a technique of lowering the quantity of data in a dataset by deleting or simplifying superfluous or unnecessary information. Data reduction's purpose is to simplify data analysis and make it easier to extract valuable insights from enormous datasets. A reduced version of datasets is obtained by applying reduction techniques that maintain integrity. By reducing the data efficiency is improved. When the data is small it is easier to apply the highly-priced algorithms. The reduction is based on rows and columns.

3) Splitting dataset into training and testing data

Training data make up 80% and testing data 20%, respectively, of the two datasets that make up the information. First, proper parameters must be used to train the model on the pre-processed datasets. The model can then be put to the test using test data. Train datasets are typically used to fit machine learning models, and test datasets are typically used to evaluate them. The key goal is to gauge the model's performance with fresh data. When there are enormous datasets, the models are trained and tested.

4) Feature Selection

It is a method of reducing inputs by only keeping useful data and discarding noisy/irrelevant data. A predictive model is built by the process of feature selection, which involves choosing a subset of pertinent characteristics. To decrease the number of features in the model while preserving or even enhancing its forecast accuracy, feature selection is used. Basically, in this, there are both useful data and noisy data from which only the useful data subset is taken in the model leaving the noisy one. Feature selection is performed by using filters method, by finding the missing value and information gain. The reduction in entropy caused by the transformation is determined by information gain. It may be used as a feature selection strategy by computing the information gain of each variable in relation to the target variable. The missing value ratio may be used to compare the feature set to the threshold value. The missing value ratio is calculated by dividing the total number of observations by the number of missing values in each column. If the variable has a value greater than the threshold, it can be removed.

5) Applying machine learning algorithms

After feature selection is done, various algorithms are applied to the models to predict the disease with high accuracy. Random Forest and SVM are the two machine learning algorithms that gained high accuracy in predicting all diseases. Whereas, VGG19 is being used for malaria and pneumonia diseases. The algorithms which are used in the model are very efficient and are giving higher accuracy. The algorithms improve their performance when available samples increase. Deep learning algorithms are a specialized form of machine learning algorithms.

6) Performance analysis

Performance analysis is a crucial component of machine learning since it aids in assessing the efficacy and efficiency of the built-in models. ROC is one of the techniques which is used in performance analysis in this model to show the accuracy of a graph. The ROC curve is a graph that shows how well a classification model performs. It illustrates how, for various threshold values, the true positive rate (sensitivity) and the false positive rate trade-off.

7) Disease Prediction

This is the last step, when the user enters all the attributes of a disease predictor and clicks the predict button, the results of the user will be shown accurately whether a person has a disease or not. It also gives some preventive measures to tackle diseases at the early stage.

Use-Case diagram

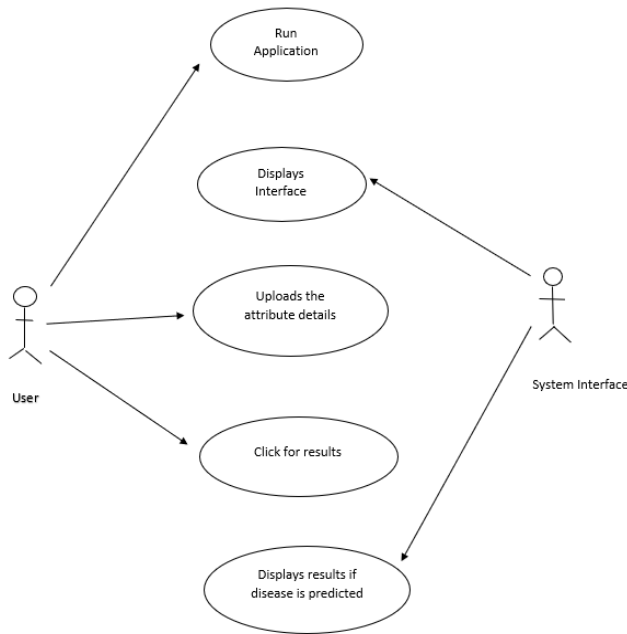


Fig. 2. Use-Case diagram

The figure 3 describes the user interface of the model.

There are two main actors in the use case diagram, they are:

1. User
2. System Interface

- Firstly, the user runs the application.
- Then, the system interface displays the interface.
- The user must now enter the attributes and click "predict" button to see the results.
- The system interface analyzes the attributes and gives the results accordingly, whether the patient is affected or not.

B. Dataset Description

1) Heart Disease Dataset

The "Heart Disease Dataset" by Kaggle is used to forecast the occurrence of heart disorders [11]. It contains 76 properties, including the one that was predicted, but none of the published studies indicate utilizing all of them; just a portion of 14 of them are mentioned. The "target" field gives results of the patient's heart disease. One represents if a patient is ill, whereas zero represents none. Sex, chest pain type (4 values), resting blood pressure, serum cholesterol in milligrams per deciliter, fasting blood sugar greater than 120 milligrams per deciliter, highest heart rate attained, exercise-induced angina, old peak = ST depression caused by exercise relative to rest, the slope of the peak exercise ST segment, and the number of major vessels (0–3) stained by fluoroscopy, and exercise-induced angina are some of the attributes.

2) Diabetics Dataset

"Pima Indian Diabetes Dataset" is used to predict diabetes in the model. This dataset offers information on the medical histories of Pima Indian women and whether or not they got diabetes. It comprises 768 observations. This dataset includes one target feature and eight medical predictor features. The traits include the total number of pregnancies, the plasma glucose concentration at two hours after an oral glucose tolerance test, the thickness of the triceps skin fold (in millimeters), the two-hour serum insulin concentration in micrograms per milliliter, and the body mass index (weight in kg divided by height in meters squared), a family history of diabetes, Years of age, variable class (0 or 1).

3) Breast Cancer Dataset

A breast cancer dataset is a collection of data on breast cancer patients that is used for research and analysis. The dataset includes information such as patient age, tumor size, tumor grade, cancer stage, and other clinical and demographic characteristics. Kaggle's "Breast Cancer Wisconsin Data Set" is used to forecast the occurrence of breast cancer illnesses [12]. This dataset contains 569 breast cancer biopsy cases. The objective is to determine if the cancer is benign or malignant. One target feature and 31 medical predictor features are included in this dataset. Several crucial qualities include the following. Identification, diameter, circumference, texture, area, perimeter, smoothness, compactness, and concavity points mean.

4) Liver Disease Dataset

The "Indian Liver Patient Records" dataset comprises 416 liver patient records and 167 non-liver patient records gathered in the north east of Andhra Pradesh, India. The "Dataset" column is a class label used to categories groups as liver patient (liver illness) or not (no disease). This data collection comprises 441 male patient records and 142 female medical records. Each patient whose age surpassed 89 is noted as being of age "90". This dataset comprises 583 patient records with 10 clinical variables such as age, gender, total bilirubin, direct bilirubin, alkaline phosphatase, alanine aminotransferase (ALT), aspartate aminotransferase (AST), total proteins, albumin, and albumin/globulin ratio. This dataset uses patient records to determine which patients have liver disease and which ones do not.

5) Kidney Disease Dataset

The Chronic Kidney Disease Dataset is accessible on the UCI Machine Learning Repository and comprises information on 400 patients with chronic kidney disease. The dataset comprises 24 variables such as age, blood pressure, serum creatinine, and albumin levels. These datasets may be used to train machine learning models for predicting kidney disease, which can aid in early identification and treatment. Nonetheless, it is vital to stress that employing these datasets for clinical decision-making should be done with caution and in cooperation with medical specialists.

6) Malaria Dataset

Malaria Cell Images Dataset consists of two folders which has infected and uninfected patient's data [14]. This dataset comprises 27,588 cell pictures, 13,780 of which are infected with malaria and 13,808 of which are uninfected. The photographs came from a variety of sources, including the National Institute of Health (NIH) and the Malaria Screener research initiative. This dataset helps to rescue humans by identifying and deploying Image Cells that contain or do not contain Malaria.

7) Pneumonia Dataset

The "Chest X-Ray Pictures (Pneumonia)" dataset is one of the most widely used pneumonia datasets [15]. This dataset comprises 5,856 JPG pictures of chest X-rays labelled as "normal" or "pneumonia". The dataset is organized into 3 folders (train, test, val). Chest X-ray pictures (anterior-posterior) were chosen from retrospective cohorts of pediatric patients aged one to five years old at Guangzhou Women and Children's Medical Center in Guangzhou. All chest X-ray imaging was done as part of the patients' usual clinical treatment. To begin the analysis of chest x-ray pictures, all chest radiographs were originally screened for quality control by deleting all low quality or unreadable scans. The diagnosis for the photographs were then assessed by two professional physicians before being approved for training the AI system. A third expert reviewed the assessment set to adjust for any grade issues.

C. Machine learning algorithms

Here is a brief explanation of the machine learning and deep learning techniques used in this article, which has as its main objective the development of a multi-disease prediction model. Diabetes analysis, heart disease prediction, liver disease, kidney illness, malaria, pneumonia, and breast cancer detection all use various machine learning and deep learning approaches.

1) Support Vector Machine (SVM)

Here the n-dimensional space is separated into different classes. The space is separated into different classes using a hyperplane. This hyperplane is a line drawn between 2 different classes to separate from each other. There are 2 parallel marginal lines to the hyperplane which are formed by selecting the corner points in the data. The distance between these two parallel lines is known as the marginal distance. Suppose, two marginal distances to a particular area then we choose the one that has a larger marginal distance. The construction of two parallel hyperplanes is made possible by support vectors of two kinds. Thus, the classifier's generalization error will be lower the wider the peripheral between the two hyperplanes. SVMs are implemented differently from other machine learning algorithms.

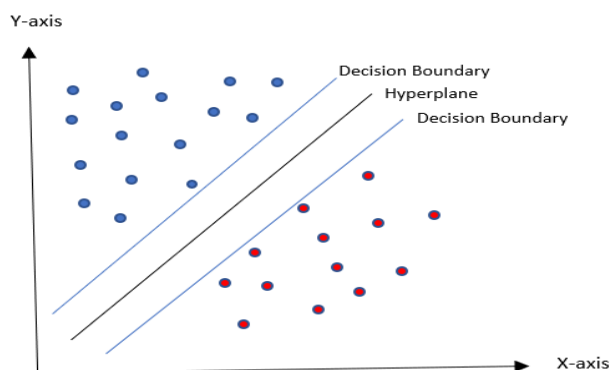


Fig. 3. SVM hyperplane

2) Random Forest Classifier

In essence, Random Decision Tree and Random Forest are synonyms. It is used to combine various classifiers to get accurate results while solving complex problems. The dataset is essentially divided into numerous Decision Trees by the Random Forest Classifier to get correct results. The dataset is divided into several Decision Trees, each of which uses a distinct row sampling and feature sampling strategy. The bulk of the decision trees is used as the output later on in the testing process. The bulk of the findings from diverse decision trees anticipates the ultimate conclusion. Decision trees have two key characteristics:

- 1) Low bias
- 2) Wide range

The large variance becomes minimal when the decision trees are combined.

The building of the random forest from the union of N decision trees and the prediction of each tree in the first stage are the two steps in the random forest's operation.

The random forest algorithm works like this:

1. Select P data points at random from the training set.
2. After selecting the P data points, start building the decision tree with those points.
3. Next, choose the N -th number of decision trees to build.
4. Repeat steps one and two.
5. After identifying the predictions made by each decision tree, we will award the winning category with the fresh data points.

Random forest classifier has gained a high accuracy in predicting heart disease, diabetes, and kidney diseases.

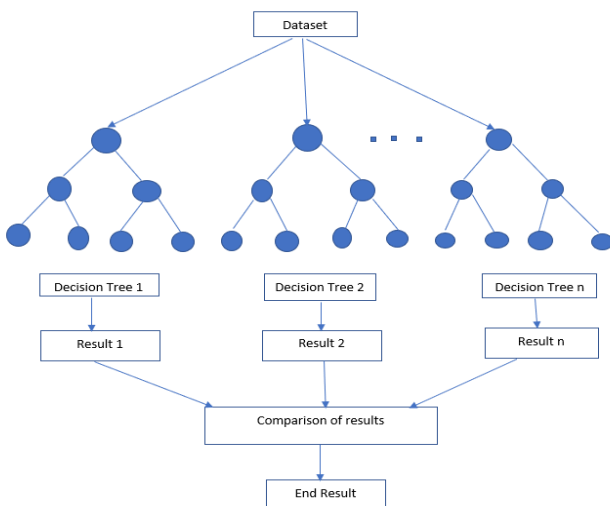


Fig. 4. Random Forest visual representation

D. Deep learning algorithms

CNN (Convolutional Neural Networks) is an artificial neural network that is widely used for image and object recognition. They are formed from neurons with learnable weights and biases. Every neuron gets many inputs, computes a weighted total, passes it through an activation function, and reacts with an output. Convolutional layers, pooling layers, and fully linked layers are the fundamental building components of a CNN. The convolutional layer extracts information like edges, forms, and textures from the input picture using a collection of filters. The pooling layer down samples the feature maps to keep just the most important information, reducing their spatial size. The fully connected layer takes the preceding layer's output and predicts it using a set of learned weights.

1) Transfer learning with VGG19

Visual Geometry Group 19 is used as an image-identifying technique and classification method. It has 19 layers. It is a convolutional neural network. It was trained using the ImageNet database, which has one million photos in a thousand different categories. Through transfer learning, an existing model can be leveraged to address several, connected issues. In essence, the goal is to use the knowledge gained

from one activity to help in generalization in a different one. To address the issue, the model's pre-trained weights are employed or model architecture. The classification tasks can be solved by using VGG16 and VGG19, different output layers. It saves training time and provides good accuracy in many instances. Kera's application has given users access to VGG19 architecture.

The VGG-19 and VGG-16 architectures are quite similar. With the VGG-16 network, there are 3 more convolutional layers. Following is the architecture:

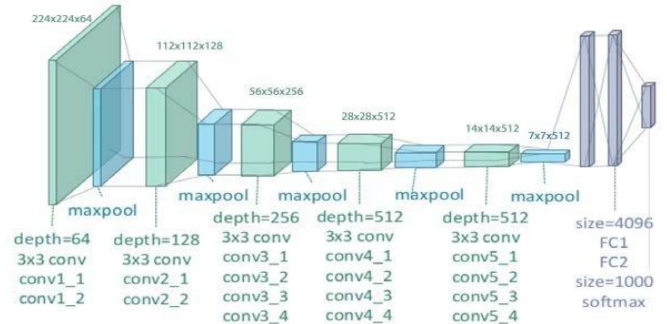


Fig. 5. Network architecture of VGG-19

Because of their depth, the VGG-16 and VGG-19 architectures create highly massive, slow-to-train models. Despite the diverse architectures that can be seen here, with a few lines of code, we can make a straightforward template for doing transfer learning from these models. To complete the necessary categorization work, use the following line of code.

IV. RESULTS

The following are the results for this model, the home page consists of a detailed description of the diseases that the model is predicting. The diseases that this model is predicting are heart, diabetes, kidney, liver, breast cancer, malaria, and pneumonia. The home page consists of various symptoms of the diseases. The home page also consists of the highest accuracies of each disease and also if a person wants to predict the disease, they need to click on the disease name and then they are translated to that page where a person needs to enter the parameters and then click the predict button after that they can get the result that the person is suffering from a particular disease or not.

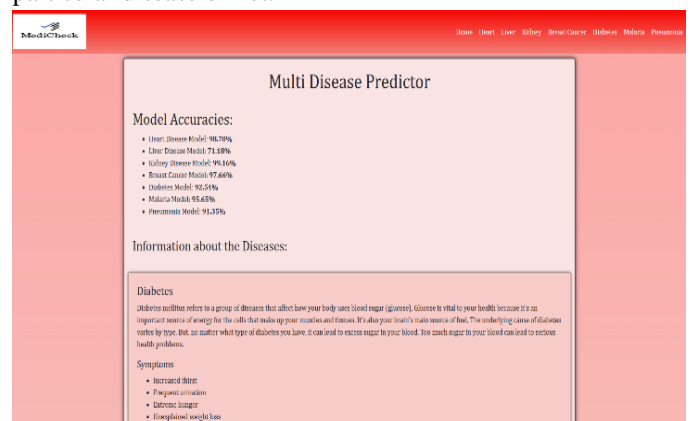


Fig. 6. User Interface

The above figure 6 represents the user interface that is the home page of the model. It contains information about all the diseases and the accuracies of each and every disease prediction model. It also contains the symptoms of the disease and preventive measures which can help a person to be fit and healthy.

Fig. 7. Heart Disease Prediction

The above figure 7 represents the heart disease prediction page which contains all the suitable attributes such as age, sex, chest pain type (4 values), resting blood pressure, serum cholesterol in mg/dl, fasting blood sugar > 120 mg/dl, etc. By using all this information, this model predicts whether a person is affected or not with the help of pre-trained models.

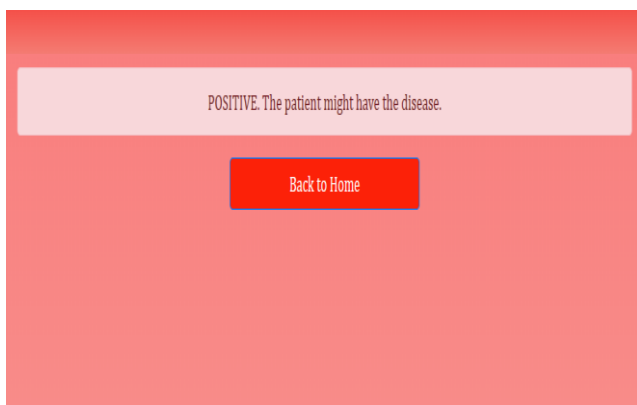


Fig. 8. Output of Heart disease prediction

Figure 8 represents the result of a patient’s heart condition. When a person clicks the predict button, it redirects to this page. It shows positive because the parameters are high than normal people have. Heart disease is usually caused when a person has high cholesterol, blood sugar, and blood pressure. With this model, this model can predict the disease at the early stage so that the mortality rate can be reduced.

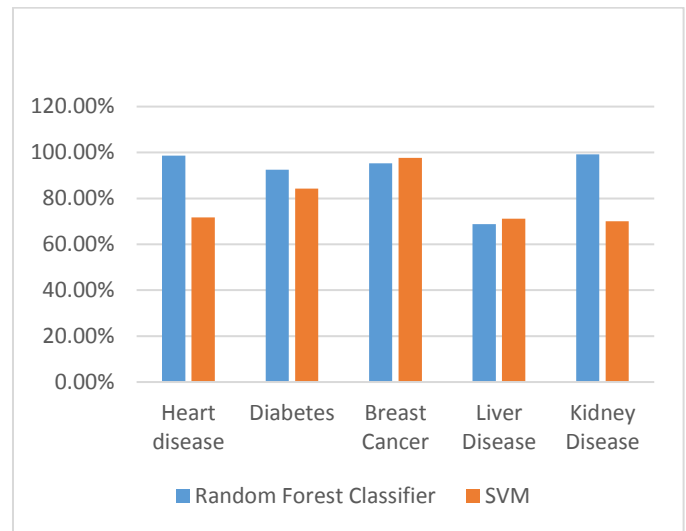


Fig. 9. Accuracy of diseases using machine learning algorithms

Basically, the model multi disease prediction gave us the results for multiple diseases. This model used various algorithms for various diseases and deployed the particular disease with the algorithm which gave us the highest accuracy. The above figure 9 gives us information about the accuracy of various diseases. Like heart disease got the highest accuracy which is 98.70% with random forest algorithm, and mean absolute error is 0.029%. Liver disease got the highest accuracy which is 71.18% with SVM algorithm, and mean absolute error is 1.024%. Kidney disease got the highest accuracy which is 99.16% with random forest algorithm, and mean absolute error is 0.043%. Breast cancer got the highest accuracy which is 97.66% with SVM algorithm, and mean absolute error is 0.035%. Diabetes got the highest accuracy which is 92.54% with random forest algorithm and mean absolute error is 0.22%. From this graph we can see that Random Forest Classifier and SVM have gained high accuracy in predicting various diseases.

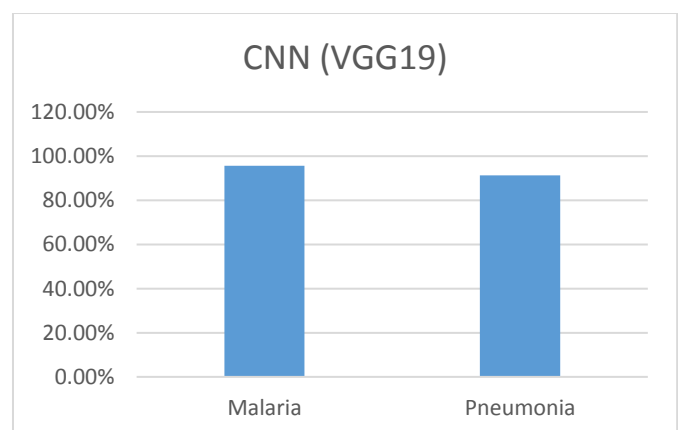


Fig. 10. Accuracy of diseases using deep learning algorithm

The above graph represents the accuracy of malaria and pneumonia using deep learning algorithms. Malaria got the highest accuracy which is 95.65% with VGG 19 algorithm and mean absolute error is 0.022%. Pneumonia got the highest accuracy which is 91.35% with VGG 19 algorithm and mean

absolute error is 0.032%. This algorithm is very suitable for image datasets. As malaria and pneumonia datasets are in the form of images, this model used the VGG19 algorithm to predict the disease. Because the VGG 19 algorithm has 19 layers of the network (16 convolutional layers, 3 fully linked layers), it can predict the disease with high accuracy.

V. CONCLUSION

The proposed study integrates breast-cancer, diabetes, malaria, pneumonia, liver, kidney and heart disease on one platform by deploying the trained models using the straightforward flask API architecture. A multi-disease prediction system can be used to forecast many illnesses at once. Here, depending on user input, a disease prognosis is created. A multiple-disease prediction model has the advantage of estimating the probability that several diseases will reveal themselves while also reducing the mortality rate.

REFERENCES

- [1]. A. Yaganteeswarudu, "Multi Disease Prediction Model by using Machine Learning and Flask API," 2020 5th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2020, pp. 1242-1246, doi: 10.1109/ICCES48766.2020.9137896.
- [2]. Saravanan, B., Duraipandian, M., & Pandiaraj, V. (2022, November). An Effective Possibilistic Fuzzy Clustering Method for Tumor Segmentation in MRI brain Images. In 2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 823-827). IEEE.
- [3]. Indukuri Mohit, K. Santhosh Kumar, Uday Avula Kumar Reddy, Badhagouni Suresh Kumar, "An Approach to detect multiple diseases using machine learning algorithm", Journal of Physics: Conference Series, Volume 2089, 1st International Conference on Applied Mathematics, Modeling and Simulation in Engineering (AMSE) 2021 15-16 September 2021, India (Virtual), 2021. doi: 10.1088/1742-6596/2089/1/012009
- [4]. P. Sonar and K. JayaMalini, "Diabetes Prediction Using Different Machine Learning Approaches," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 367-371, doi: 10.1109/ICCMC.2019.8819841.
- [5]. A. Singh and R. Kumar, "Heart Disease Prediction Using Machine Learning Algorithms," 2020 International Conference on Electrical and Electronics Engineering (ICE3), Gorakhpur, India, 2020, pp. 452-457, doi: 10.1109/ICE348803.2020.9122958.
- [6]. Y. Khourdifi and M. Bahaj, "Applying Best Machine Learning Algorithms for Breast Cancer Prediction and Classification," 2018 International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS), Kenitra, Morocco, 2018, pp. 1-5, doi: 10.1109/ICECOCS.2018.8610632.
- [7]. V. V. Vijayan and C. Anjali, "Prediction and diagnosis of diabetes mellitus — A machine learning approach," 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS), Trivandrum, India, 2015, pp. 122-127, doi: 10.1109/RAICS.2015.7488400.
- [8]. R. Pal, J. Poray and M. Sen, "Application of machine learning algorithms on diabetic retinopathy," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2017, pp. 2046-2051, doi: 10.1109/RTEICT.2017.8256959.
- [9]. M. Chen, Y. Hao, K. Hwang, L. Wang and L. Wang, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities," in IEEE Access, vol. 5, pp. 8869-8879, 2017, doi: 10.1109/ACCESS.2017.2694446.
- [10]. Ahmed, L. J., Fathima, B. A., Mahaboob, M., & Gokulavasan, B. (2021, March). Biomedical Image Processing with Improved SPIHT Algorithm and optimized Curvelet Transform Technique. In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 1596-1602). IEEE.
- [11]. Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, Parneet Singh, "Prediction of Heart Disease Using a Combination of Machine Learning and Deep Learning", Computational Intelligence and Neuroscience, vol. 2021, Article ID 8387680, 11 pages, 2021. <https://doi.org/10.1155/2021/8387680>
- [12]. Yufeng Zheng, Clifford K. Yang, "Breast cancer screening using convolutional neural network and follow-up digital mammography", Proceedings Volume 10669, Computational Imaging III; 1066905 (2018) <https://doi.org/10.1117/12.2304564>.
- [13]. S. T. Himi, N. T. Monalisa, M. Whaiduzzaman, A. Barros and M. S. Uddin, "MedAi: A Smartwatch-Based Application Framework for the Prediction of Common Diseases Using Machine Learning," in IEEE Access, vol. 11, pp. 12342-12359, 2023, doi: 10.1109/ACCESS.2023.3236002.
- [14]. O. Iradukunda, H. Che, J. Uwineza, J. Y. Bayingana, M. S. Bin-Imam and I. Niyonzima, "Malaria Disease Prediction Based on Machine Learning," 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP), Chongqing, China, 2019, pp. 1-7, doi: 10.1109/ICSIDP47821.2019.9173011.
- [15]. U. Singh, A. Totla and D. P. Kumar, "Deep Learning Model to Predict Pneumonia Disease based on Observed Patterns in Lung X-rays," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2020, pp. 1315-1320, doi: 10.1109/ICECA49313.2020.9297388.