

Web Application For Prediction And Classification Of Multiple Diseases Using Machine Learning And Deep Learning Techniques

Vishnupriya B

Assistant professor, Department of CSE
KPR Institute of Engineering and Technology
Coimbatore, India
vishnupriya.b@kpriet.ac.in

Abitha R

UG-Final year, Department of CSE
KPR Institute of Engineering and Technology
Coimbatore, India
abivaisu2221@gmail.com

Harshithaa A S

UG-Final year, Department of CSE
KPR Institute of Engineering and Technology
Coimbatore, India
sujisujithaa7@gmail.com

Agalya M

UG-Final year, Department of CSE
KPR Institute of Engineering and Technology
Coimbatore, India
agalya12172001@gmail.com

Gugapriya M

UG-Final year, Department of CSE
KPR Institute of Engineering and Technology
Coimbatore, India
gugapriyajuly@gmail.com

Abstract — Our proposal presents a disease prediction system using machine learning to predict diabetes, breast cancer, heart diseases, kidney diseases and deep learning to predict malaria and pneumonia. Users must go visit the hospital for a checkup for minor issues, which takes additional time. It's also quite stressful to answer phone calls for appointments. Such an issue can be fixed by employing illness prediction software and providing appropriate advice on leading a healthy lifestyle. Due to a range of diseases and a low doctor-to-patient ratio, the usage of particular disease prediction techniques and health-related concerns has expanded over the past ten years. Therefore, with this system, we are primarily focused on providing users with rapid and accurate illness prediction based on the symptoms they enter, together with the expected severity of the condition. The most appropriate algorithm and medical advice will be provided in this project. Different machine learning algorithms like random forest, decision tree, naïve bayes, KNN, support vector machine, logical regression and deep learning algorithms like CNN are employed to forecast illnesses to provide speedy and precise predictions. The symptoms entered will be compared to the database in one channel. Additionally, if the symptom is new, which is its primary function, it will be retained in the database, and the other channel will offer a forecast level of illness severity. A web/android application is set up for users to allow for simple portability, configuration, and remote access to areas that doctors find difficult to access. Since most users are unaware of all available treatments for a given ailment, this project also aims to offer pharmaceutical and prescription counseling for any anticipated diseases. As a result, this approach makes health management simpler. This web application was developed using Flask Web Framework. The models used to predict the diseases were trained on large Datasets. The web application can predict diabetes, breast cancer, heart disease, kidney disease, liver disease, malaria and pneumonia.

Keywords— *Diabetes, Breast Cancer, Heart Disease, Kidney Disease, Malaria, Pneumonia, Machine Learning, Deep Learning, Random Forest, Decision Tree, Support Vector Machine (SVM), Logistic Regression, Naive Bayes, Convolutional Neural Network (CNN), K-Nearest neighbour (KNN)*

I. INTRODUCTION

Heart disease is on the rise due to our busy, stressful lives nowadays. Heart disease affects people of all ages, therefore it's important to identify it early through symptoms or reports. When your kidneys become damaged and are unable to filter your blood as effectively, chronic kidney disease develops. When a disease reaches its final stage, it has two options. A technique for filtering blood is hemo dialysing; another is transplanting, which is regarded as being superior than purifying techniques. Diabetes mellitus is brought on by abnormally high blood sugar levels. It interferes with the body's capacity to generate the hormone insulin, improper glucose metabolism and elevated blood sugar levels result. The primary goal of such a system is to assist doctors in cross-verifying their diagnostic findings, which offers a viable remedy to the current death rates. In order to provide a novel platform and the most promising method for the early diagnosis of various diseases, we have put out this study. When medical data is insufficient or of low quality, it affects. Prediction of a disease based on the patient's symptoms and medical history For decades, machine learning has been a stumbling barrier. In the medical industry, Machine Learning technology provides a powerful venue for rapidly resolving healthcare challenges.

Severe liver disease [29] is characterised by the persistent degradation of liver functioning. Liver activities include the generation of clotting factors and other proteins, detoxification of toxic metabolic products, and bile excretion. This is a continual process of liver parenchymal inflammation, destruction, and regeneration that leads to fibrosis and cirrhosis. Cirrhosis is the last stage of chronic liver disease, characterised by disruption of hepatic architecture, the creation of extensive nodules, vascular rearrangement, neo-angiogenesis, and extracellular matrix deposition. The recruitment of stellate cells and fibroblasts that induce fibrosis is the basic process of fibrosis and cirrhosis at the cellular level, whereas parenchymal regeneration is dependent on hepatic stem cells.

Python pickling is used to store a model's behaviour. The significance of this system analysis is that in analysing illnesses, all of the characteristics that produce the condition are included, allowing for more efficient and accurate disease detection. The behaviour of the completed model will be preserved as a Python pickle file.

A lot of research on existing health-care systems focused on only one ailment at a time. One approach, for example, is used to analyse diabetes, another to analyse diabetic retinopathy, and still another to predict heart disease. Most systems concentrate on a single ailment. When a company wishes to analyse their patients' health records, they must use a variety of models. The current system's methodology is only suitable for analysing certain disorders. A user can analyse more than one disease on a same webpage using the multiple diseases prediction method. The user is not need to travel to multiple locations in order to anticipate whether or not he or she has a certain condition. The user need a various disease prediction system. The goal is to develop a good Machine Learning algorithm that is efficient and accurate for illness prediction. In this study, the supervised Machine Learning concept is employed for forecasting the illnesses.

The key feature will be Machine Learning, in which we will use algorithms such as Decision Tree, Random Forest, Nave Bayes, and KNN to aid in accurate illness prediction and better patient care [25].

II. LITERATURE SURVEY

This section outlines the investigation of previously proposed models for illness prediction that are linked to our proposed work. Many research have been conducted in order to diagnose various illnesses. They used several data mining approaches to accurately forecast a range of illnesses.

[17] Prediction Of Diabetes Using Machine Learning Classification Approaches was proposed by G Naveen Kishore and a few other writers. Several classification techniques, such as SVM, Logistic Regression, Decision Tree, KNN, and Random Forest, are used on the 769 instances of the Pima dataset, which comprise characteristics such as Pregnancies, Blood Pressure, Body Mass Index, and so on. The classification method Random Forest achieved the maximum

accuracy of 74.4%, while the KNN achieved the lowest accuracy in this experiment.

[18] Dilip Singh Sisodia and Deepti Sisodia predicted diabetes using classification algorithms and reported an accuracy of roughly 76% on the Pima dataset.

The major goal of the study is, because the heart is crucial in living creatures. As a result, the diagnosis and prognosis of heart illness should be exact and correct since it is highly important and might cause death instances connected to heart. As a result, machine learning and artificial intelligence can help forecast any type of natural event. Hence, in this study, they analyse the accuracy of machine learning for predicting heart disease using k-nearest neighbour, decision tree, linear regression, and SVM, using the UCI repositor dataset used for training and testing. They also compared the algorithms' accuracy: SVM 83%, Decision tree 79%, Linear regression 78%, and k-nearest neighbour 87% [19].

The authors of [21] provided an automated method for addressing challenging questions about heart disease prediction. This intelligent system was built using the Naive Bayes approach to produce faster, better, and more accurate results. It might help doctors make clinical decisions regarding heart attacks. This system might be improved by adding SMS capabilities, developing Android and iOS mobile apps, and integrating a pacemaker in the order.

Diabetes and breast cancer were identified by adding the adaptivity feature into support vector machines [22]. The objective was to provide a quick, automated, and flexible diagnostic approach utilising adaptive SVM. The bias value in traditional SVM was modified to improve results. The proposed classifier generated output in the form of 'if-then' rules. The suggested technique was applied to diagnose diabetes and breast cancer, and it produced 100% correct classification rates for both illnesses. Future research should concentrate on developing more efficient methods for altering the bias value in traditional SVM.

[23] "Heart Disease Prediction System Using Machine Learning" has constructed a heart disease prediction system utilising NB algorithm that delivers 88.163% accuracy among others.

"Application of Machine Learning Predictive Models in the Chronic Disease " [24] focused on SVM and LR algorithms and evaluate the study models associated with diagnosis of chronic disease. These models are highly applicable in classification and diagnosis of CD.

III. PROPOSED SYSTEM

User-friendly Web application for prediction of disease using machine learning algorithm and deep learning .

1) Backend

a) Machine Learning

- i. Collecting a datasets of different disease
- ii. Data preprocessing
- iii. Training and Testing of datasets
- iv. Creation of model for different algorithms of KNN, SVM, Navie bayes, Decision Tree ,

Logistic Regression, Random forest .

- v. Analysis the high accuracy algorithm

b) Deep learning

- i. Using a image datasets
- ii. Image preprocessing
- iii. Creating a model using Convolution Neural Network

c) Model creation

- i. In machine learning , model is that has high accuracy is created as model for backend using pickle
- ii. In deep learning , model is saved by 'save_weights' in keras library.
- iii. Deploying the created model using flask library in python to connect to front end.

2) Frontend

a) Home page

- i. Consists of navigation to all other pages and contact information .

b) Prediction pages for different disaese

- i. Consists of 7 pages of prediction such that Kidney, Liver, Heart , Breast cancer, Diabetes , Malaria and pneumonia .
- ii. Getting the input from user in form.
- iii. After submitting the form , it will connect to the created model and based on input it predicts the disease.

IV. METHODOLOGIES

1) MACHINE LEARNING

a) DATASET

Datasets are collected from Kaggle, and datasets are analasied and preprocessed for prediction

- i. Diabetes datasets

Name of Dataset : Pima Indians Diabetes Database.

Predict the onset of diabetes based on diagnostic measures [Fig. 1](#)

Data - 768

Features- 9

```
: Pregnancies      0
  Glucose          0
  BloodPressure    0
  SkinThickness    0
  Insulin          0
  BMI              0
  DiabetesPedigreeFunction  0
  Age              0
  Outcome          0
dtype: int64
```

Figure 1. Features for diabetes prediction

- ii. Heart Datasets

Data – 304 [Fig. 2](#)

Features – 14

```
: age      0
  sex      0
  cp       0
  trestbps 0
  chol     0
  fbs      0
  restecg  0
  thalach  0
  exang    0
  oldpeak  0
  slope    0
  ca       0
  thal     0
  target   0
dtype: int64
```

Figure 2. Features for heart disease prediction

- iii. Kidney Datasets

Data – 400 [Fig. 3](#)

Features - 26

```
age      9
bp       12
sg       47
al       46
su       49
rbc      152
pc       65
pcc      4
ba       4
bgr      44
bu       19
sc       17
sod      87
pot      88
hemo     52
pcv      70
wc      105
rc       130
htn      2
dm       2
cad      2
appet    1
pe       1
ane      1
classification  0
dtype: int64
```

Figure 3. Features for Kidney disease prediction

- iv. Liver Datasets

Data – 583 [Fig. 4](#)

Feature – 11

```
: Age      0
  Gender   0
  Total_Bilirubin  0
  Direct_Bilirubin  0
  Alkaline_Phosphotase  0
  Alamine_Aminotransferase  0
  Aspartate_Aminotransferase  0
  Total_Protiens  0
  Albumin  0
  Albumin_and_Globulin_Ratio  4
  Dataset  0
dtype: int64
```

Figure 4 . Features for Liver disease prediction

v. Breast cancer Datasets

Data – 569 [Fig. 5](#)
 Features 33
 scikit-learn pandas.core.frame.DataFrame
 RangeIndex: 569 entries, 0 to 568
 Data columns (total 33 columns):
 id 569 non-null int64
 diagnosis 569 non-null object
 radius_mean 569 non-null float64
 texture_mean 569 non-null float64
 perimeter_mean 569 non-null float64
 area_mean 569 non-null float64
 smoothness_mean 569 non-null float64
 compactness_mean 569 non-null float64
 concavity_mean 569 non-null float64
 concave_points_mean 569 non-null float64
 symmetry_mean 569 non-null float64
 fractal_dimension_mean 569 non-null float64
 radius_se 569 non-null float64
 texture_se 569 non-null float64
 perimeter_se 569 non-null float64
 area_se 569 non-null float64
 smoothness_se 569 non-null float64
 compactness_se 569 non-null float64
 concavity_se 569 non-null float64
 concave_points_se 569 non-null float64
 symmetry_se 569 non-null float64
 fractal_dimension_se 569 non-null float64
 radius_worst 569 non-null float64
 texture_worst 569 non-null float64
 perimeter_worst 569 non-null float64
 area_worst 569 non-null float64
 smoothness_worst 569 non-null float64
 compactness_worst 569 non-null float64
 concavity_worst 569 non-null float64
 concave_points_worst 569 non-null float64
 symmetry_worst 569 non-null float64
 fractal_dimension_worst 569 non-null float64
 Unnamed: 32 0 non-null float64
 dtypes: float64(31), int64(1), object(1)
 memory usage: 146.8+ KB

Figure 5. Features for Breast Cancer prediction

b) MACHINE LEARNING ALGORITHM

i. Decision Tree

A supervised technique for solving classification-related issues is a decision tree [8]. There is a widely used pre-defined target variable included. The decision tree may be used to input and output variables that are continuous or specified. This method may be used to tackle classification and regression issues because Based on the splitter that matters the most for the input variable, it divides the population or sample into two or more homogeneous groups called subpopulations. Starting at the root, a decision tree is built using a top-down method that includes data segmentation. Below are the formulas for calculating [Fig. 6](#) entropy and the Gini index [Fig. 7](#).

$$Entropy = - \sum_j p_j \log_2 p_j$$

Figure. 6 Entropy formula

$$Gini = 1 - \sum_j p_j^2$$

Figure. 7 Gini Formula

ii. Naive Bayes

The Naive Bayes [\[10\]](#) method uses the Bayes theorem to train a classifier. In other words, it is a probabilistic classifier that was taught using the Naive Bayes approach. It computes a probability distribution for a given observation over many classes.

[Fig. 8](#) illustrates the essential relationship of the Bayes theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Figure. 8 Bayes theorem

P(B) is constant across all classes, hence the only product that will maximize is P(B/A)*P(A).

iii. Support vector machine

SVM [\[6\]](#), a supervised machine learning algorithm can help with classification and regression problems. In an N-dimensional space, the SVM technique finds a hyperplane that classifies the input points. The size of the hyperplane is determined by the number of features. The hyperplane is just a line if there are only two input features. The hyperplane collapses to a two-dimensional plane when there are three input characteristics. The following are some benefits of support vector machines: effective in circumstances of enormous scope. When there are more dimensions than samples, the method is still useful.

iv. KNN

KNN [\[7\]](#) is a basic supervised machine learning (ML) approach that may be used to solve missing value imputation, classification, or regression problems. It is based on the idea that the observations in a data collection that is closest to a certain data point are the observations that are the most "similar," and that may thus classify unexpected points based on the values of the closest existing points. By choosing K, the user may specify the number of nearby observations to utilize in the method.

v. Logistic Regression

The Logistic Regression model is a popular statistical model in statistics that is used mostly for categorization. That is, given a collection of data, the Logistic Regression method assists us in categorizing these observations into two or more distinct groups. As a result, the target variable is discrete.

vi. Random forest

A popular classification and regression method that uses supervised learning is called Random Forest [\[11\]](#). One of the most crucial algorithms in machine learning is

this one. It is based on the idea of ensemble learning, which is the practise of integrating many classifiers to solve a challenging issue and enhance the model's performance.

Random Forest [11] is a classifier that uses a number of decision trees on different subsets of a given dataset and averages them to enhance the predicted accuracy of that dataset.

c) HYPERPARAMETER TUNING

RandomizedSearchCV [13] has "fit" and "score" methods. If the estimator supports them, it also implements "score samples", "predict", "predict_proba", "decision function", "transform", and "inverse transform". Cross-validated search across parameter settings is used to optimise the estimator parameters required to apply these approaches. In contrast to GridSearchCV, a defined number of parameter settings are sampled from the given distributions rather than all parameter values being tried out. n_iter specifies the number of parameter settings that are attempted. Sampling without replacement is conducted if all parameters are supplied as a list. Sampling with replacement is utilised if at least one parameter is presented as a distribution. Continuous distributions are strongly advised for continuous parameters.

2) DEEP LEARNING

Deep learning[12] is a subset of machine learning that is mostly used to train image models and video models that rely on neural networks with three or more layers. These models are capable of learning from the neurons built specifically for images.

a) Image dataset

i. Malaria

The dataset contains 2 folders

- Infected [Fig. 9](#)
- Uninfected [Fig. 10](#)
- And a total of 27,558 images.

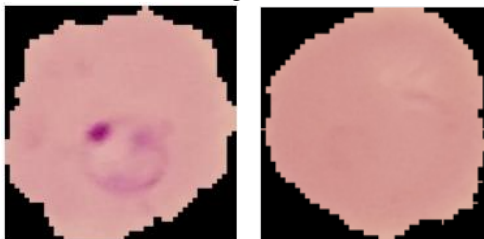


Figure 9. Infected Cell

Figure 10. Uninfected

ii. Pneumonia

The dataset is divided into three folders: train, test, and val, with subfolders for each picture category (Pneumonia/Normal). There are 5,863 JPG X-Ray pictures and two categories (Pneumonia/Normal)[Fig. 11](#) and [Fig. 12](#). The typical chest X-ray (left panel) shows clean lungs with no spots of aberrant opacification. Bacterial pneumonia (centre) is characterised by a focused lobar consolidation, in this example in the right upper lobe (white arrows), whereas viral

pneumonia (right) is characterised by a more widespread "interstitial" pattern in both lungs.



Figure 11. Normal Lung

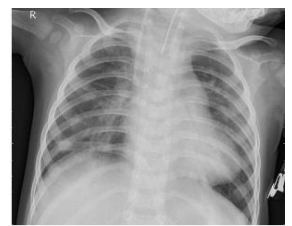


Figure 12. Infected Lung

b) Image Pre-processing

The term "image pre-processing" describes operations on images that are carried out at the most fundamental level of abstraction. These operations lessen rather than increase the information content of the pictures if entropy is a measure of information. Pre-processing aims to enhance image data by reducing undesirable distortions or strengthening specific visual characteristics necessary for later processing and analysis activities.

Image pre-processing techniques :

- Image filtering
- Image segmentation

c) Flowchart for Deep Learning

The flow of the deep learning procedure for the prediction of malaria and pneumonia disease is shown in the flowchart below [Fig. 13](#).

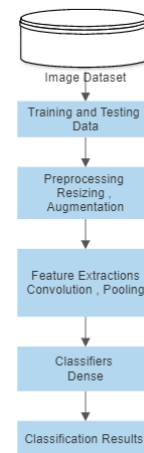


Figure. 13 Flowchart of Deep learning process

d) CNN

Convolutional neural networks (CNN) [15] , are a type of deep learning technique that may apply weights and biases to different objects and characteristics in an input picture. Comparatively speaking, a CNN requires far less pre-processing than other classification techniques. CNN can learn these filters and attributes, as opposed to manually-engineered filters in basic systems. A CNN's construction [Fig. 14](#) is influenced by the visual cortex's organization, which is a mirror of the neural connections present in the human brain.

Individual neurons only respond to inputs in this tiny area of the visual field known as the Receptive Field. These overlapping patterns are repeated until the entire visual field is covered.

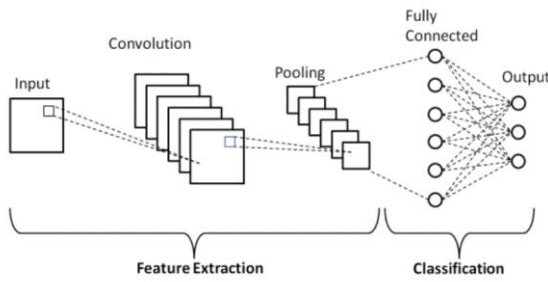


Figure. 14 CNN Layers [16]

A convolution neural network has multiple hidden layers [Fig. 14](#) that help in extracting information from an image. The four important layers in CNN are:

1. Convolution layer
2. ReLU layer
3. Pooling layer
4. Fully connected layer

Convolution Layer This is the first step in the process of extracting valuable features from an image. A convolution layer has several filters that perform the convolution operation. Every image is considered as a matrix of pixel values

ReLU layer ReLU stands for the rectified linear unit. Once the feature maps are extracted, the next step is to move them to a ReLU layer. ReLU performs an elementwise operation and sets all the negative pixels to 0. It introduces non-linearity to the network, and the generated output is a rectified feature map.

Pooling Layer Pooling is a down-sampling operation that reduces the dimensionality of the feature map. The rectified feature map now goes through a pooling layer to generate a pooled feature map.

3) EVALUATION METRICS

a) Confusion matrix

The confusion matrix was used to analyze classifier performance using the metrics listed below.

True Positive (TP): The proportion of classes that, when projected favorably, are likewise positive.

The True Negative (TN): statistic tracks the percentage of negatively predicted classes that are truly negative.

False Positive (FP): The number of classes that were incorrectly categorized while being favorably classified.

False Negative (FN): The amount of negatively classified courses that were mistakenly classified

b) Classification report

Equations 1-4 in [Fig. 15](#) state that the models are assessed based on four standard metrics: accuracy, precision, recall, F1-score, and specificity, where TP denotes true positive, TN

denotes true negative, FP denotes false positive, and FN denotes false negative.

Type of Metric	Formula
Accuracy	$ACC = \frac{tp+tn}{tp+fp+tn+fn}$
Recall	$Recall = \frac{tp}{tp+fn}$
Precision	$Precision = \frac{tp}{tp+fp}$
F1-score	$F = 2 \cdot \frac{precision \cdot recall}{precision+recall}$

Figure 15. Evaluation Metrics

Precision

Precision is a classifier's ability to avoid labeling a negative occurrence as positive. For each class, it is defined as the ratio of true positives to the average of true positives and false positives.

Precision is a Positive prediction accuracy.

$Precision = \text{True Positive} / (\text{True Positive} + \text{False Positive})$ as in [Fig. 15](#)

Recall

A classifier's recall is its capacity to discover all positive occurrences. For each class, it is defined as the ratio of true positives to the average of true positives and false negatives.

The recall is a proportion of positives

$Recall = \text{True Positive} / (\text{True Positive} + \text{False Negative})$ as in [Fig. 15](#)

F1 Score

The F1 score is determined as a calculated average of accuracy and recall. with 1.0 being the highest and 0.0 being the poorest. F1 scores are lower than accuracy measurements because of factors in precision and recall. Use the average value of F1 rather than global accuracy to compare classifier models.

$F1 \text{ Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$ as in [Fig. 15](#)

c) Accuracy Comparison

For each disease different machine learning algorithm is applied to analysis the best algorithm

- i. Kidney Disease

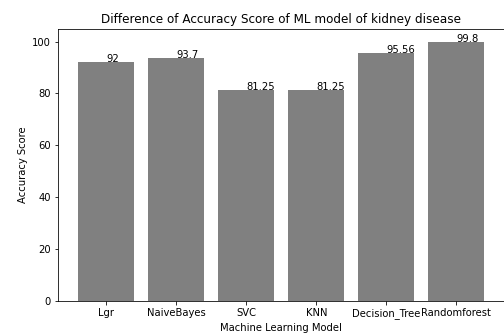


Figure 16. Represents the accuracy of different algorithm on Kidney datasets

[Fig. 16](#) Shows that Random forest has higher accuracy of 99.8 percentage .
Hence the pickle model is created for Random forest on kidney disease .

ii. Heart Disease

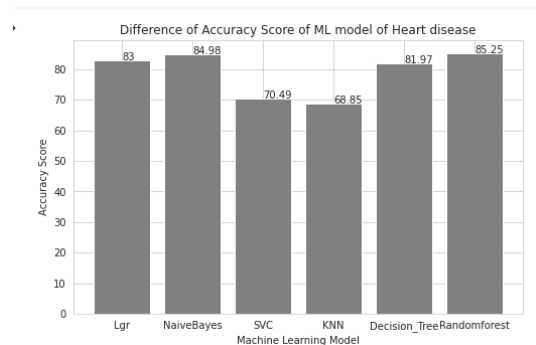


Figure 17. Represents the accuracy of different algorithm on Heart datasets

[Fig. 17](#) Shows that Random forest has higher accuracy of 85.25 percentage .
Hence the pickle model is created for Random forest on Heart disease .

iii. Liver Disease

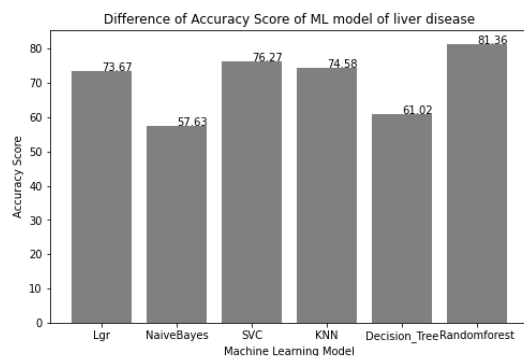


Figure 18. Represents the accuracy of different algorithm on Liver datasets

[Fig. 18](#) Shows that Random forest has higher accuracy of 81.36 percentage .
Hence the pickle model is created for Random forest on Liver disease .

iv. Breast Cancer

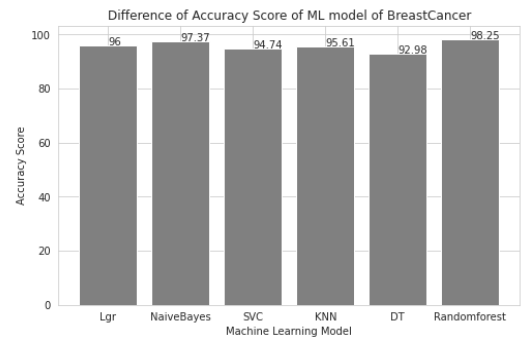


Figure 19. Represents the accuracy of different algorithm on Breast datasets

[Fig. 19](#) Shows that Random forest has higher accuracy of 98.25 percentage .
Hence the pickle model is created for Random forest on Breast Cancer .

v. Diabetes Disease

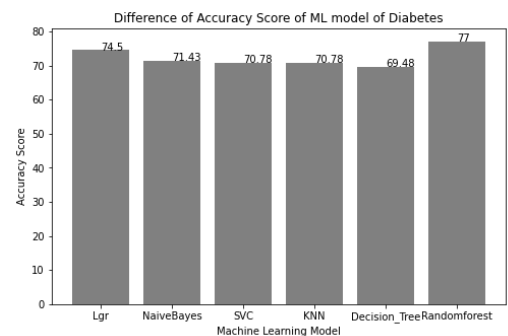


Figure 20. Represents the accuracy of different algorithm on Diabetes datasets

[Fig. 20](#) Shows that Random forest has higher accuracy of 77 percentage .
Hence the pickle model is created for Random forest on Diabetes disease .

V) RESULTS

a) Machine Learning Model Classifier

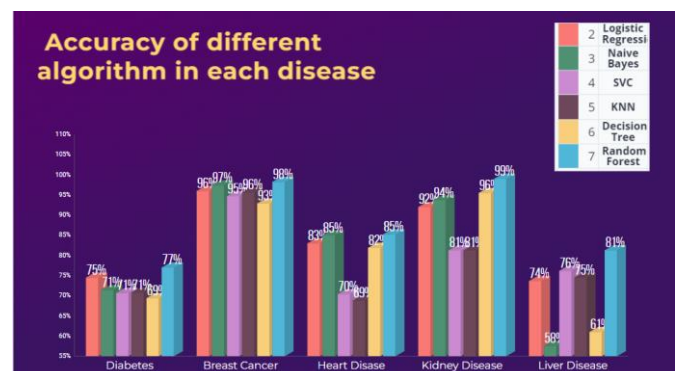


Figure 21. Accuracy score of different algorithms on all disease

[Fig. 21](#) Represents the accuracy score for different algorithms on each disease datasets on machine learning model .

Based on the analysis of [Fig. 21](#) , Random forest has high accuracy for all disease , Hence the Random forest is created a model for disease prediction .

Accuracy Score	
Disease	Accuracy score
Diabetes	77 %
Breast Cancer	98.25%
Heart Disease	85.75%
Kidney Disease	99.8%
Liver Disease	81.36%

Table 1. Accuracy score of different disease of high accuracy

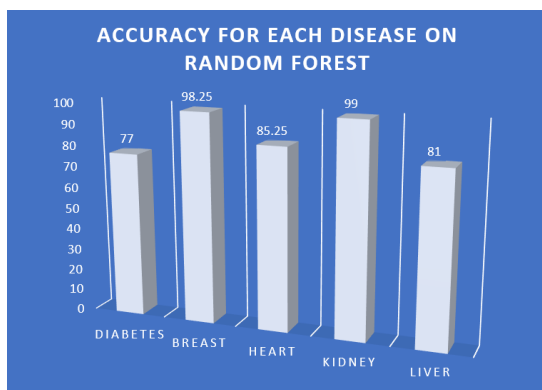


Figure 22. Represents the accuracy for all disease on Random forest algorithm

[Fig. 22](#) and [Table. 1](#) demonstrate the display of individual base classifiers for each disease .

b) Deep Learning Model

Deep learning model is created with CNN layers and tested with validation data which has high of [Fig. 23](#) and [Table. 2](#)

DL Model	Accuracy score
Malaria	96%
Pneumonia	95%

Table 2. Accuracy of CNN model

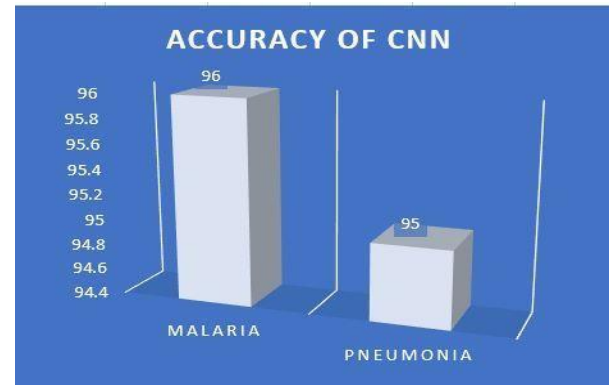


Figure 23. Represents the accuracy of malaria and pneumonia in CNN

[Fig. 23](#) and [Table. 2](#) demonstrate the accuracy of image datasets under deep learning model .

This model is saved by save_weight from keras and used as backend for web application.

c) Front End web Design

This web application is useful for doctors to predict disease based on the features .

Web application consists of 10 Pages.

1. Home Page
2. Disease Page
3. Diabetes Prediction Page
4. Heart Disease Prediction Page
5. Liver Disease prediction page
6. Breast cancer Prediction page
7. Kidney Disease Prediction page
8. Malaria Prediction Page
9. Pneumonia Prediction page
10. Analysis Page

Here the some pages from web applications

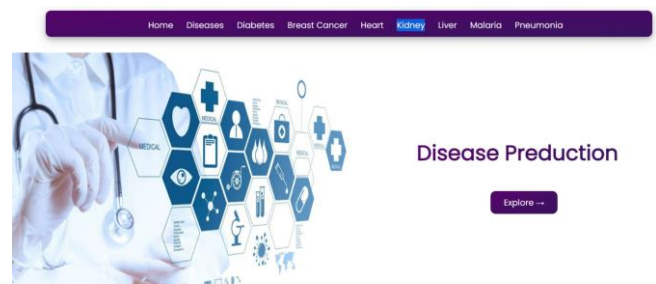


Figure 24. Home page of web application

The page [Fig. 24](#) consists of navigation to all other pages and also contact information for email and phone numbers are included .

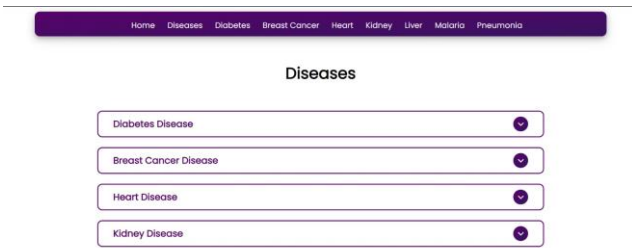


Figure 25. Shows the symptoms ,details, and awareness about each diseases

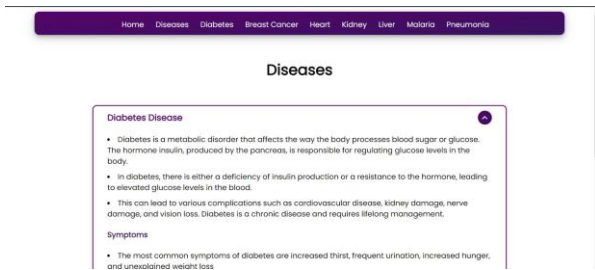


Figure 26. i.e for Diabetes disease syptoms and information

[Fig. 25](#) and [Fig. 26](#) represents the Disease Page , which consists of all details and symptoms of each disease . Same as [Fig. 25](#) and [Fig. 26](#), Symptoms and details are included for all other disease .

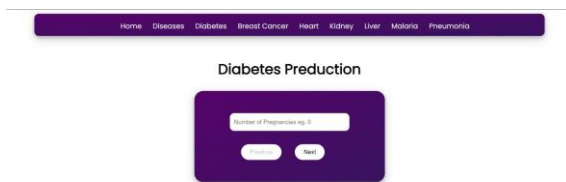


Figure 27. i.e for Diabetes Prediction Page

[Fig. 27](#) is the example of front end form for all disease like Liver , Breast cancer, Kidney disease ,Heart Disease , Diabetes Disease .

[Fig. 27](#) Form has to be filled based on the example input and submit the form .

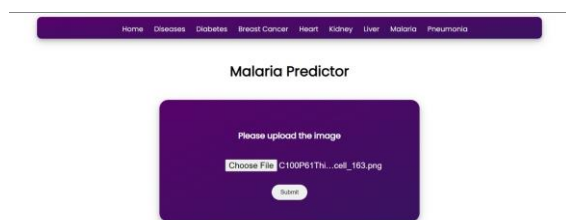


Figure 28. Shows the form for uploading the image for malaria prediction



Figure 29. Shows the Output after predictions

[Fig. 28](#) and [Fig. 29](#) represents the design form for the malaria prediction page where the input is uploaded as image.

[Fig. 28](#) and [Fig. 29](#) has design same as for Pnuemonia prediction .

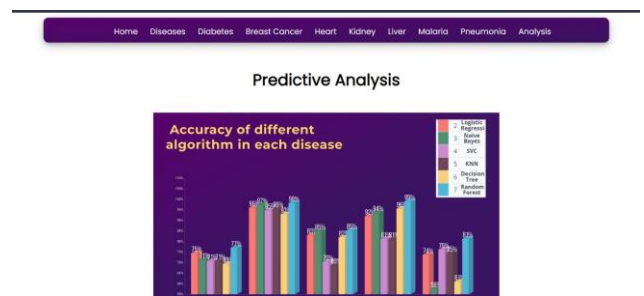


Figure 30. shows the predictive analysis

[Fig. 30](#) Represents the predictive analysis of model created in backend process

The web application is developed with HTML , CSS , Javascript for frontend and Flask for Backend .

V. CONCLUSION AND FUTURE WORKS

The primary goal of this study was to develop a system that could accurately forecast several diseases. The doctors doesn't have to navigate via several websites thanks to this project, which also saves time. The project is set up so that the gadget uses the user's symptoms as input and produces illness prognosis as the output.

Early diagnosis of diseases can both lengthen your life and spare you from financial hardship. To attain the highest level of accuracy, we have employed a variety of machine learning techniques, including Random Forest.

Our ideology is practicable and, if possible, might be developed as an application. Making image detection easier by segmenting and creating images from all conceivable angles. Other than binary classification, various other features may be used to discover further classifications such as low, mild, medium, and high (chances). Paper committed such a concept for detection using both Machine Learning and Deep Learning, as well as Edge detection technology, which is almost always used in various Domains and is primarily

known for its detection accuracy. Concluding here that our system can be adapted to any kind of Environment and it supports well with any Ideology to make the change.

In the future, it can add more strategies like the XGB Algorithm which is an Extreme Gradient Boosting algorithm that can make even high accuracy of predictions. It and also be developed as a Web or Android Application. The Identification and tracking system can be implemented if possible. Making React JavaScript or Other Attractive web apps for the model can make the system more familiar among the others, both the positive and negative results should be conveyed in a good understandable manner for the user/client who uses the system.

A high level of Visualization can be implemented before developing the model to make the process understandable for the user too, since visualization is a visual technique there is no need for one to have Coding knowledge to understand it.

The incorporation of several innovative ideas into CNN's architectural design has impacted the trajectory of research, notably in the fields of image processing and CV. The successful use of CNN to grid-like topological data illustrates its utility as an image representational model.

A promising field of study, CNN architectural design is expected to become one of the most widely utilised AI techniques in the future.

Ensemble learning is one of the possible areas of research in CNNs. The combination of multiple and diverse architectures can benefit the model by boosting generality and robustness on many categories of pictures by extracting several layers of semantic representations.

The same is true for concepts such as batch normalisation, dropout, and new activation functions, all of which should be highlighted.

VI. DATA AVAILABILITY

a) Heart Disease

[Heart Disease Dataset | Kaggle](#)

b) Kidney Disase

[Chronic Kidney Disease dataset | Kaggle](#)

c) Liver Disease

[Indian Liver Patient Records | Kaggle](#)

d) Breast Cancer

[Breast Cancer Wisconsin \(Diagnostic\) Data Set | Kaggle](#)

e) Diabetes

[Pima Indians Diabetes Database | Kaggle](#)

f) Malaria

[Malaria Cell Images Dataset | Kaggle](#)

g) Pneumonia

[Chest X-Ray Images \(Pneumonia\) | Kaggle](#)

REFERENCES

- [1] Nagasankar, Tamilselvi & Ankaryarkanni, B.. (2016). Performance Analysis of Edge Detection Algorithms on Various Image Types. *Indian Journal of Science and Technology*. 9. 10.17485/ijst/2016/v9i21/95207.
- [2] Wenzhou Gao, Xiaoguang Zhang, Lei Yang and Huizhong Liu, "An improved Sobel edge detection," *2010 3rd International Conference on Computer Science and Information Technology*, 2010, pp. 67-71, DOI: 10.1109/ICCSIT.2010.5563693.
- [3] Anil K. Bharodiya, Atul M. Gonsai, An improved edge detection algorithm for X-Ray images based on the statistical range, *Heliyon*, Volume 5, Issue 10, 2019, e02743, ISSN 2405-8440.
- [4] Y. Menga, Z. Zhang, H. Yin, et al. Automatic detection of particle size distribution by image analysis based on local adaptive canny edge detection and modified circular Hough transform *Micron*, 106 (2018), pp. 34-41
- [5] K. Zhang, Y. Zhang, P. Wang, et al. An improved Sobel edge algorithm and FPGA implementation *Procedia Comput. Sci.*, 131 (2018), pp. 243-248
- [6] Polat H, Danaei Mehr H, Cetin A. Diagnosis of Chronic Kidney Disease Based on Support Vector Machine by Feature Selection Methods. *J Med Syst*. 2017 Apr;41(4):55. doi: 10.1007/s10916-017-0703-x. Epub 2017 Feb 27. PMID: 28243816.
- [7] N. Saranya, M. Sakthi Samyuktha, S. Isaac and B. Subhanki, "Diagnosing Chronic Kidney Disease using KNN Algorithm," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 2038-2041, DOI: 10.1109/ICACCS51430.2021.9442039.
- [8] Ilyas H, Ali S, Ponum M, Hasan O, Mahmood MT, Iftikhar M, Malik MH. Chronic kidney disease diagnosis using decision tree algorithms. *BMC Nephrol*. 2021 Aug 9;22(1):273. doi: 10.1186/s12882-021-02474-z. PMID: 34372817; PMCID: PMC8351137.
- [9] Subasi, A., Alickovic, E., Kevric, J. (2017). Diagnosis of Chronic Kidney Disease by Using Random Forest. In: Badnjevic, A. (eds) *CMBEBIH 2017. IFMBE Proceedings*, vol 62. Springer, Singapore.
- [10] U. N. Dulhare and M. Ayesha, "Extraction of action rules for chronic kidney disease using Naïve Bayes classifier," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2016, pp. 1-5, DOI: 10.1109/ICCIC.2016.7919649.
- [11] Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
- [12] A. Shrestha and A. Mahmood, "Review of Deep Learning Algorithms and Architectures," in *IEEE Access*, vol. 7, pp. 53040-53065, 2019, doi: 10.1109/ACCESS.2019.2912200.
- [13] Yang, Li & Shami, Abdallah. (2020). On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice.
- [14] Jia Wu, Xiu-Yun Chen, Hao Zhang, Li-Dong Xiong, Hang Lei, Si-Hao Deng, Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimization, *Journal of Electronic Science and Technology*, Volume 17, Issue 1, 2019, Pages 26-40, ISSN 1674-862X, <https://doi.org/10.11989/JEST.1674-862X.80904120>.
- [15] Alzubaidi, L., Zhana, J., Humaidi, A.J. et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **8**, 53 (2021). <https://doi.org/10.1186/s40537-021-00444-8>
- [16] <https://www.upgrad.com/blog/basic-cnn-architecture/#>
- [17] Naveen Kishore G,V .Rajesh ,A.Vamsi Akki Reddy, K.Sumedh,T.rajesh Sai Reddy, "Prediction Of Diabetes Using Machine Learning Classification Algorithms".
- [18] Deepti Sisodia, Dilip Singh Sisodia," Prediction of Diabetes using classification Techniques.
- [19] Archana Singh ,Rakesh Kumar, "Heart Disease Prediction Using Machine Learning Algorithms", 2020 IEEE, International Conference on Electrical and Electronics Engineering (ICE3)
- [20] NisharBanu, MA; Gomathy, B.; (2013). Disease Predicting System Using Data Mining Techniques.

- [21] S. Grampurohit, C. Sagarnal, Disease prediction using machine learning algorithms, 2020 Int. Conf. Emerg. Technol. (INCET) (2020) 1–7, <https://doi.org/10.1109/INCET49848.2020.9154130>.
- [22] R.J.P. Princy, S. Parthasarathy, P.S. Hency Jose, A. Raj Lakshminarayanan, S. Jeganathan, Prediction of Cardiac Disease using Supervised Machine Learning Algorithms, in: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 570–575, <https://doi.org/10.1109/ICICCS48265.2020.9121169>.
- [23] Shrestha,Ranjit.,& Chatterjee,Jyotir. Moy. (2019).Heart Disease Prediction System Using Machine Learning . LBEF Research Journal of Science Technology and Management , 1(2), 115-132.
- [24] Battineni ,Gopi. , Sagaro,Getu.Gamo. ,Chinatalapudi, Nalini. ,&Amenta,Francesco. (2020). Application Of Machine Learning Predictive Models in the Chronic Disease .International of PersonalisedMedicine , 10(21), 1- 11.
- [25] Godse, Rudra A.,Gunjal,Smita S., JagtapKaran A .,Mahamuni ,Neha S., &Wankhade, Prof. Suchita. (2019). Multiple Disease Prediction Using Different Machine Learning Algorithms Comparatively.International Journal of Advance Research in Computer and Communication Engineering, 8(12), 50-52
- [26] Chauhan Raj H., NaikDaksh N. ,Halpati,Rinal A., Patel,Sagarkumar J. , &PrajapatiMr. A.D. (2020). Disease Prediction using Machine Learning.International Research Journal of Engineering and Technology, 7(5) , 2000-2002.
- [27] P. Prabhu, S. Selvaabharathi. Deep Belief Neural Network Model for Prediction of Diabetes Mellitus. In 2019 3rd International Conference on Imaging, Signal Processing and Communication, ICISPC 2019 (pp. 138–142) Institute of Electrical and Electronics Engineers Inc. ISBN:9781728136639. 2019.
- [28] Marimuthu , M. , Abinaya, M. ,Hariesh,K.S., Madhan,K.,& Pavithra, Kumar. V.(2018).A Review of Heart Disease Prediction Using Machine Learning and Data Analytics Approach .International Journal of Computer Application , 181(18), 20-25.
- [29] Sharma A, Nagalli S. Chronic Liver Disease. [Updated 2022 Jul 4]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK554597/>
- [30] R.J.P. Princy, S. Parthasarathy, P.S. Hency Jose, A. Raj Lakshminarayanan, S. Jeganathan, Prediction of Cardiac Disease using Supervised Machine Learning Algorithms, in: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), 2020, pp. 570–575, <https://doi.org/10.1109/ICICCS48265.2020.9121169>.