

# Development of Web Portal for Risk Prediction and Image-based Disease Detection using AI Algorithms

Arsya Mohamed Ali<sup>1</sup>, Nithin Gangadharan Rangaraj<sup>2</sup>

<sup>1,2</sup> Student, Dept. of Biomedical Engineering, PSG College of Technology, Tamil Nadu, India

\*\*\*

**Abstract** – An accessible platform that integrates various provisions for abnormalities detection would help clinicians to solve different medical issues effectively. The platform developed is focused primarily on Risk Prediction and Image-based Disease Detection. The former employs two models that predict the risk of developing Cardiovascular diseases and Diabetes which helps clinicians for early prognosis while the latter incorporates two Image-based models that detect the presence of Hemorrhage and Pneumonia intending to prevent medical negligence. The models were trained using various Artificial Intelligence algorithms and the accurate models were deployed in the web portal after appropriate evaluation.

**Key Words:** Platform, Risk Prediction, Cardiovascular diseases, Diabetes, Image-based Disease Detection, Hemorrhage, Pneumonia.

## 1. INTRODUCTION

The human population has an upward trend all over the world and the rate of developing a disease or an abnormality is growing alongside. Therefore, it is essential to improvise the methods of diagnosis and prognosis to meet the rising demands. To improve the results of medical treatments, an integrated portal equipping the tasks of Image-based disease detection and Risk prediction is developed.

### 1.1 Image-based Disease Detection

The first section employs medical images such as X-rays, CT scans to detect the presence of Hemorrhage and Pneumonia. This provides fast and easy access for doctors and clinicians to diagnose several patients in a short period. In addition, it also helps in double verification of the results thereby preventing any medical negligence if caused.

Brain Hemorrhage is a condition that causes bleeding within the brain tissues or between the surrounding bone due to ruptured blood vessels. High blood pressure or even a sudden shock to the human skull may increase the person's risk. Symptoms are location-specific but numbness or weakness in the part of the face, sudden headache, dizziness, difficulties in speaking or walking are common. Around 15% of the strokes are caused by Intracerebral Hemorrhage (ICH), a common and disabling form of stroke [1].

Pneumonia is a microbial infection that primarily affects alveoli present in the lungs. The infection can be caused by either virus or bacteria and less commonly by other means

(Medication, Other Microorganisms, Autoimmune diseases). The symptoms exhibited by the patients generally include dry cough, chest pain, fever, and difficulty in breathing. A 2017 review conducted by Global Change Data Lab showed that around 2.56 million people died of pneumonia where one-third of the victims are children under the age of 5 [2].

The proposed method uses CT images of patients with cerebral hemorrhage and Chest X-ray images as the input data for Hemorrhage detection and Pneumonia detection, respectively. The review of the Image Classification Competition called ImageNet Large Scale Visual Recognition Challenge (ILSVRC) conducted in the year 2015 showed that ResNet 50 obtained the top-5 error rate as 6.71% while VGG16 obtained 9.33% [3]. Sirish Kaushik et al [4] developed four models employing Convolutional Neural Networks (CNNs) and their results proved that the number of layers in the model directly corresponds to the accuracy of prediction.

### 1.2 Risk Prediction

The second section predicts the susceptibility of the person towards cardiovascular diseases (Coronary heart diseases, stroke, congestive heart failure, hypertension, etc.) and Diabetes. On the one hand, the development of self-awareness is ensured from the patient's perspective. On the other hand, early predictions allow better treatment and prognosis by doctors.

Diseases that include heart and blood vessels are collectively termed Cardiovascular diseases (CVD). CVDs cause high mortality and morbidity rate than any other diseases worldwide. A statistical review showed that around 17.9 million people died in the year 2016 due to heart diseases which contributed to 31% of global deaths [5]. Certain lifestyle modifications and early treatment greatly reduce the chance of developing CVDs.

Diabetes is a disease that occurs when the blood sugar level is too high. There are various causes for this disease which include obesity, inactive lifestyle, immune system disorder, and many more. According to World Health Organization, around 1.5 million people died due to diabetes in the year 2019. Simple lifestyle changes greatly can prevent diabetes and hence, an early prediction could improve the treatment efficacy [6].

The parameters and symptoms for the diseases are used as the input data for classification purposes. Archana Singh et al [7] predicted heart disease using machine learning algorithms. They employed Support Vector Machine (SVM), k-Nearest Neighbor (KNN), and Decision tree algorithms for achieving their objective. Their results showed that KNN

outperformed with 87% testing accuracy. Muhammad Azeem Sarwar et al [8] predicted diabetes and inferred that a dataset with minimum or no missing attributes would provide better results. They concluded that a 99.99% accurate model demands a minimum of thousands of records with no missing attributes.

## 2. DATASET

The dataset used for Hemorrhage Detection consists of 200 Brain CT images where 50% of the images are Normal CT scans while the rest of the images are Hemorrhage CT scans. These images were collected from 200 different patients in .png format. It also contains a .csv file where all the images are labeled as either 'normal' or 'hemorrhage'. There is no distinction between the types of hemorrhage. The overall size of the dataset is 24.47 MB [9]. The sample images of normal and hemorrhage head CT scans are shown in figures 1 and 2.

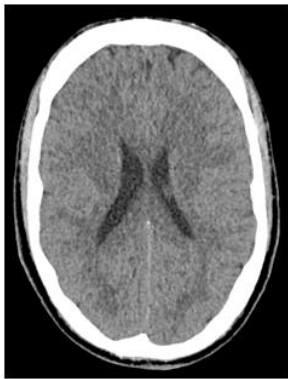


Fig -1: Normal CT

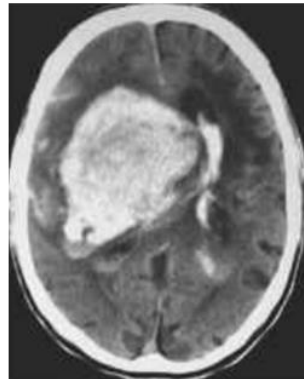


Fig -2: Hemorrhage CT

The dataset employed for Pneumonia Detection consists of 5863 Chest X-ray images in .jpeg format with an overall size of 1.15 GB. These images are anteroposterior chest X-ray images chosen from a retrospective cohort study of 1- to 5-year-old pediatric patients. The dataset contains three main folders for training, testing, and validating purposes. Each main folder has a subfolder for two image categories i.e., Pneumonia and Normal [10]. The sample normal and pneumonia chest X-ray images are shown in figures 3 and 4.



Fig -3: Normal X-ray



Fig -4: Pneumonia X-ray

The dataset for heart disease risk prediction contains information from a prospective study conducted by the U.S. Public Health Service to analyze the incidence and prevalence of cardiovascular diseases and their risk factors.

It includes the information collected from the first 32 clinical exams. The study recruited 5,209 men and women sampled from Framingham, Massachusetts. Overall, the dataset includes 4240 rows with 15 attributes [11]. The attributes are listed as follows,

1. Gender,
2. Age,
3. Current Smoker or not,
4. Cigarettes per day,
5. BP medication,
6. Is he/she prevalent to stroke,
7. If he/she is prevalent to hypertension,
8. Does he/she have diabetes,
9. Total cholesterol level,
10. Systolic Blood pressure,
11. Diastolic Blood pressure,
12. Body Mass Index,
13. Heart rate,
14. Glucose,
15. If he/she have the chances of getting heart disease in 10 years.

Prediction of diabetes is carried out using a dataset that contains 520 instances with 17 attributes including the target. It is originally taken from patients belonging to an age group of 16-90 of Sylhet Diabetes Hospital collected in the form of a questionnaire [12]. The following are the attributes present in the dataset,

1. Age,
2. Gender,
3. Polyuria,
4. Polydipsia,
5. Sudden weight loss,
6. Weakness,
7. Polyphagia,
8. Genital Thrush,
9. Visual blurring,
10. Itching,
11. Irritability,
12. Delayed healing,
13. Partial paresis,
14. Muscle stiffness,
15. Alopecia,
16. Obesity,
17. Target (Positive/Negative).

## 3. METHODOLOGY

### 3.1 Image-based Disease Detection

#### 3.1.1 Data Preprocessing

Data preprocessing or data cleansing is performed to improve the image data by suppressing the undesired distortions and enhancing some image features relevant for further processing and analysis.

### 3.1.1.1 Resizing

The hemorrhage dataset contains images of different sizes as shown in figure 5; therefore, it is essential to resize. The images are resized to 128 x 128 pixels by analyzing the histogram. In addition, the summary of the histogram is shown in table 1.

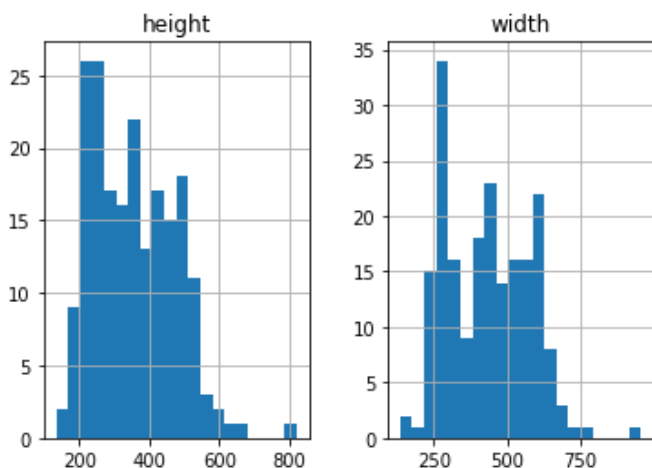


Fig -5: Histogram of Height and Width

	height	width
count	200.00	200.00
mean	355.50	433.72
std	116.78	142.05
min	134.00	135.00
25%	252.00	298.75
50%	345.00	435.50
75%	447.25	554.00
max	821.00	957.00

Table -1: Summary of the Histogram

The pneumonia dataset contains 5863 images of different heights and widths, before creating and training the model the images should be resized. The images are resized to 150 x 150 pixels by analyzing the dataset.

### 3.1.1.2 Normalization

In image processing, normalization is performed to change the range of pixel intensity values. Grayscale normalization means values of pixels are in the range from 0 to 1. It is performed on both the hemorrhage and pneumonia datasets to scale back the impact of illumination's differences. Moreover, CNN converges quicker on [0...1] data than on [0...255].

### 3.1.1.3 Augmentation

To improve accuracy and prevent overfitting, deep neural networks require a lot of training data. Image augmentation can increase the data by creating new examples from the existing ones, thereby improving the performance.

As both the classes in the hemorrhage dataset comprise 100 images, the dataset is balanced. The number of images in the dataset is less which might lead to overfitting. The efficiency of the dataset can be improved by adding augmented images as shown in figure 6. The performed augmentation operations are shown in table 2. To make sure that augmented images are not added to the test set, this process is executed after the dataset split.

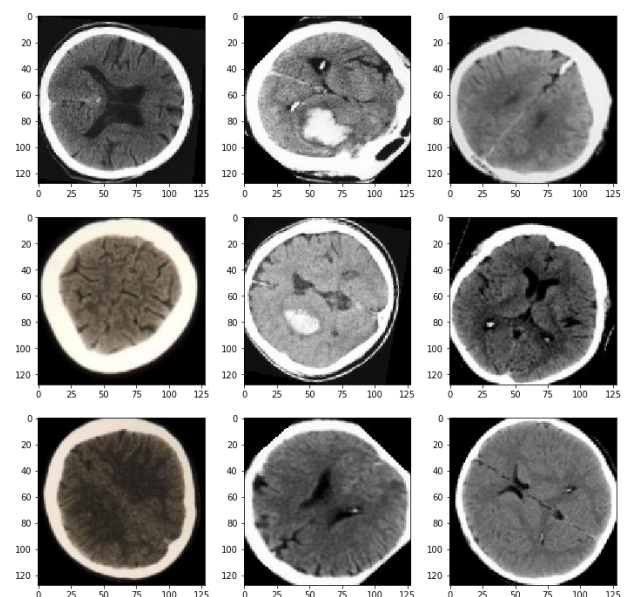


Fig -6: Augmented Sample CT Image

Operations	Value
Rescaling	255
Zoom	0.1
Horizontal Flip	True
Rotation Range	10 degree
Height, Width Shift	0.1

Table -2: Augmentation Parameters

The pneumonia dataset contains 3883 Pneumonia chest X-Ray images and 1349 normal chest X-Ray images; therefore, the dataset is imbalanced. The distribution of the target in the training set is shown in figure 7. To overcome the imbalance in the training set, data augmentation operations shown in table 3 are performed. The sample augmented images can be seen in figure 8.



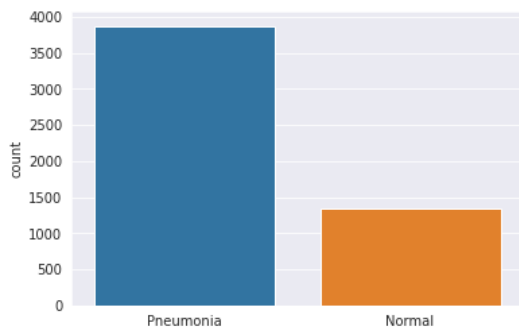


Fig -7: Target Distribution in the Training set

Operations	Value
Rotation Range	30 degree
Zoom	0.2
Horizontal Flip	True
Height Shift	0.1
Width Shift	0.1

Table -3: Augmentation Parameters

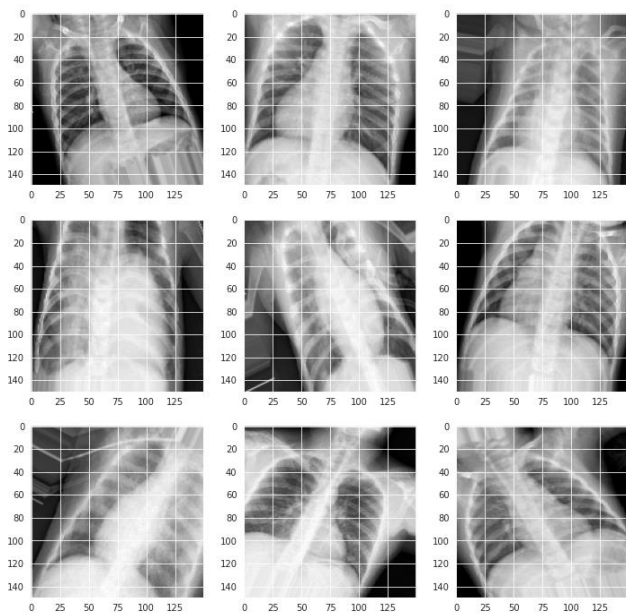


Fig -8: Augmented Sample X-Ray Image

### 3.1.2 Model Training

For each image-based detection model, three deep learning networks (CNN, ResNet50 & VGG16) are trained and evaluated.

The architecture of the CNN model is shown in figure 9. It comprises three convolutional blocks each consisting of a convolutional layer, max pooling layer & fully connected layer. Out of several types of non-linear activations, sigmoid and ReLU activation functions are used. The sigmoid activation takes a real-valued number as input and outputs 0 or 1. It is used in the final layer to obtain the class. ReLU

operation is used in the convolution layers which outputs zero for all negative inputs. A dropout layer is incorporated to regularize the network.

Model: "sequential\_4"

Layer (type)	Output Shape	Param #
conv2d_12 (Conv2D)	(None, 126, 126, 32)	320
activation_20 (Activation)	(None, 126, 126, 32)	0
max_pooling2d_12 (MaxPooling)	(None, 63, 63, 32)	0
conv2d_13 (Conv2D)	(None, 61, 61, 32)	9248
activation_21 (Activation)	(None, 61, 61, 32)	0
max_pooling2d_13 (MaxPooling)	(None, 30, 30, 32)	0
conv2d_14 (Conv2D)	(None, 28, 28, 64)	18496
activation_22 (Activation)	(None, 28, 28, 64)	0
max_pooling2d_14 (MaxPooling)	(None, 14, 14, 64)	0
flatten_4 (Flatten)	(None, 12544)	0
dense_8 (Dense)	(None, 64)	802880
activation_23 (Activation)	(None, 64)	0
dropout_4 (Dropout)	(None, 64)	0
dense_9 (Dense)	(None, 1)	65
activation_24 (Activation)	(None, 1)	0
Total params: 831,009		
Trainable params: 831,009		
Non-trainable params: 0		

Fig -9: CNN Architecture

The ResNet-50 model has an input layer, a functional block, and an output layer which is shown in figure 10. The functional block used is the ResNet-50 pre-trained model from Keras which is a deep residual network comprising 50 layers. The pre-trained ResNet-50 model has five stages each with a convolutional and identity block. Each convolutional and identity block has 3 convolutional layers.

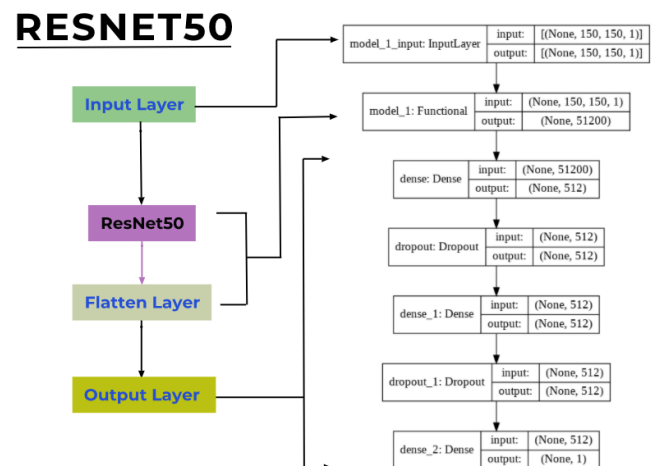


Fig -10: ResNet-50 Architecture

The VGG16 model has an input layer, a functional block, and an output layer which is shown in figure 11. The functional block used is the VGG16 pre-trained model from Keras which is a convolutional neural network containing 16 layers. The pre-trained model has 13 convolutional layers and 3 fully connected layers.

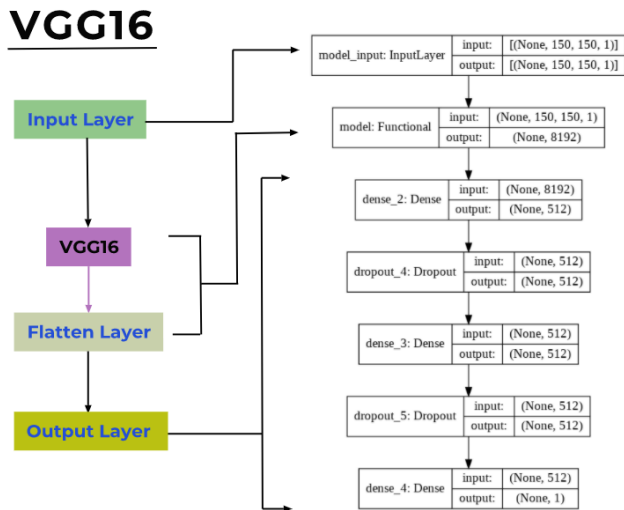


Fig -11: VGG16 Architecture

For the hemorrhage detection, the three models discussed above are trained with 160 CT images, tested with 20 CT images, and validated with 20 CT images. Therefore, the train-test-val split is 80:10:10.

For pneumonia detection, the three models (CNN, ResNet-50 & VGG16) are trained using 5216 X-Ray images, tested using 624 X-Ray images, and validated using 16 X-Ray images. Therefore, the train-test-val split is 89:10:1. Table 4 shows the number of epochs each model trained.

Model	No. of Epochs	
	Hemorrhage (Batch Size = 10)	Pneumonia (Batch Size = 32)
CNN	100	50
ResNet-50	400	12
VGG16	100	50

Table -4: Number of Epochs Trained

### 3.1.3 Model Evaluation

The validity of the proposed method is related to the accuracy, loss, and F1 score. The evaluation results for each detection model are shown in tables 5 & 6.

Model	CNN	ResNet-50	VGG16
Training Accuracy	97.5%	96.24%	98.12%
Testing Accuracy	94.9%	89.9%	94.99%
Validation Accuracy	89.99%	94.99%	85.0%
F1 Score	95%	83%	95%

Table -5: Performance Metrics – Hemorrhage Model

Model	CNN	ResNet-50	VGG16
Training Accuracy	92.71%	90.68%	94.71%
Testing Accuracy	91.5%	81.73%	91.03%
Validation Accuracy	81.25%	68.75%	87.5%
F1 Score	90.5%	79%	90%

Table -6: Performance Metrics – Pneumonia Model

From the performance metrics comparison of hemorrhage and pneumonia models, it can be inferred that CNN and VGG16 have better accuracy than the ResNet50 model.

In the world of medicine, false classifications play a significant role. Having considered this point of view the trained models are evaluated using the F1 score as it is sensitive to false negative and false positive classifications. The F1 score comparison of the three models for each detection model is shown in figures 12 & 13.



Fig -12: F1 Score Comparison – Hemorrhage Model



Fig -13: F1 Score Comparison – Pneumonia Model

On comparing the three models, the ResNet-50 has a lesser F1 score of 83% and 79% for the hemorrhage and pneumonia detection respectively. A lesser F1 score depicts more chance for misclassification to occur.

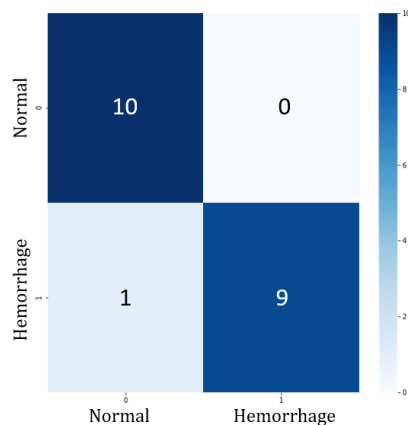


Fig -14: Confusion Matrix (CNN) - Hemorrhage

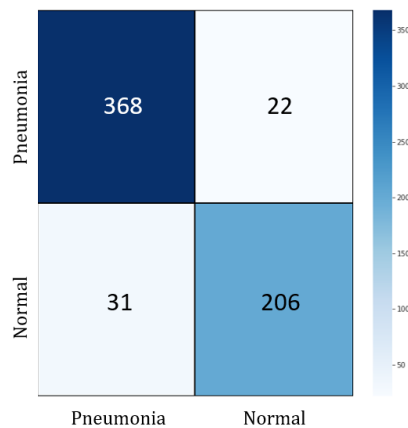


Fig -15: Confusion Matrix for CNN - Pneumonia

From the F1 score comparison, it is seen that for both hemorrhage and pneumonia detection, CNN has the highest score of 95% and 90.5% respectively. The confusion matrices of CNN for both hemorrhage and pneumonia are shown in figures 14 & 15. There is 1 false-positive image from the hemorrhage test set and 51 misclassified images (22 false negatives and 31 false positives) from the pneumonia test set.

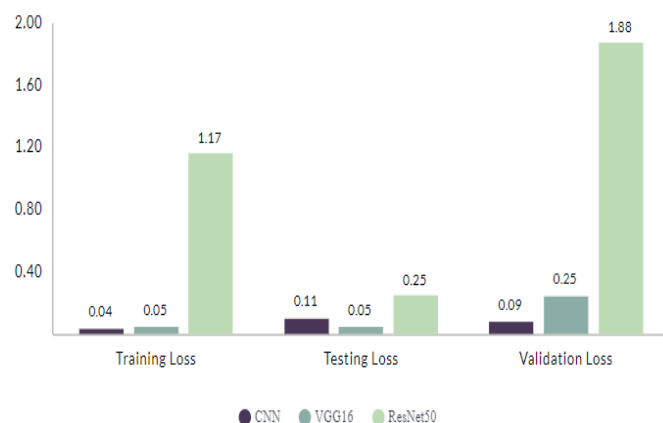


Fig -16: Loss Comparison – Hemorrhage Model

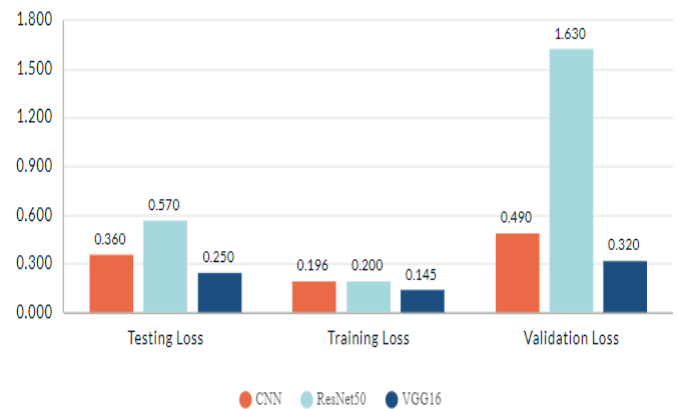


Fig -17: Loss Comparison – Pneumonia Model

From the bar plot shown in figure 16, it can be concluded that ResNet50 has high training and validation loss of 1.17 and 1.88 respectively. In figure 17, the ResNet50 has a validation loss of 1.63 which is a greater loss compared to the other two models.

Based on the accuracy, loss, and F1 score comparison, CNN & VGG16 have similar results and also performed better than the ResNet-50 model. For model deployment, the size of the model needs to be considered. The size of the CNN is 9.61 MB which is smaller than VGG16 and ResNet-50 which are of sizes 146 MB and 310 MB respectively. Also, CNN has comparatively less loss value than the VGG16.

Thus, for both hemorrhage and pneumonia detection, the CNN model is selected for deployment on the web portal.

### 3.2 Risk Prediction

#### 3.2.1 Data Preprocessing

##### 3.2.1.1 CVD Prediction

As mentioned in section 2, the heart study dataset consists of 4240 instances with 15 attributes including the target. Analysis shows that the data consists of 12.74% of null values. The distribution of null values is shown in table 7.

Attributes	Total	Percentage
Glucose	388	9.15
BPMeds	53	1.25
totCHol	50	1.18
CigsPerDay	29	0.68
BMI	19	0.45
heartRate	1	0.02

Table -7: Null Values Distribution

The presence of null values greatly affects the performance of the model and hence the rows containing

null values are removed entirely from the dataset which resulted in 3751 instances. The ratio of positive to negative cases of the 3751 patients is 15:85. To avoid the biasing issue due to class imbalance, Synthetic Minority Oversampling TEchnique (SMOTE) is employed to balance the targets.

The working of SMOTE is that it oversamples the minority class with newly created synthetic samples along the line segment joining some or all the k minority class nearest neighbors [13]. With the sampling strategy set as 0.9, the minority class is oversampled to have 90% of the number of samples in the majority class. On the other hand, by using Random under sampler, the number of samples in the majority class is reduced by 1.11 times (1/0.9) the minority class. The target distribution before and after the application of SMOTE is shown in figure 18.

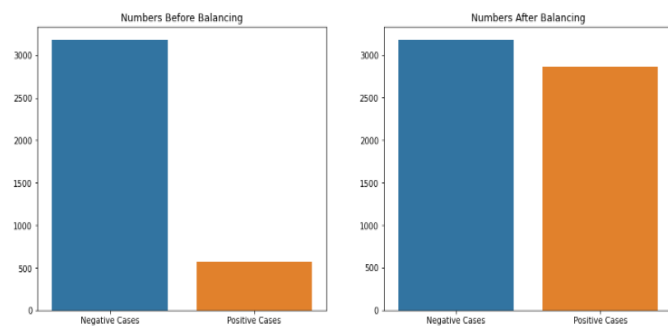


Fig -18: Target Distribution before and after balancing

The correlation study shows that there is no single feature that has a correlation greater than 0.5 and hence it is concluded that one or two features will not be sufficient to achieve the objective. Furthermore, attributes conveying similar information such as CigsPerDay & Current Smoker, diabetes & glucose (multicollinear features) are present in the dataset which might affect the performance of the model. Therefore, it is essential to perform feature selection to extract the desired features for training purposes. The correlation heat map can be seen in figure 19.

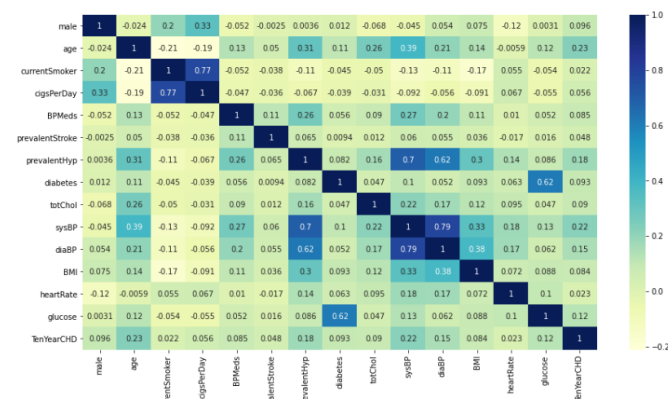


Fig -19: Correlation heat map – CVD

Boruta is the feature selection algorithm implemented to reduce unnecessary noise by removing repetitive features. It is an algorithm that uses Random Forest Classifier to determine the relevant features. Firstly, the algorithm creates shadow attributes for all existing attributes and shuffles them. Secondly, Random Forest Classifier is trained using the extended attributes and Z-Score is calculated. Z-Score is computed by dividing average loss by standard deviation. The maximum Z-Score (MZSA) among the shadow attributes are found. Thirdly, all the attributes with Z-Score significantly higher than MZSA are marked as ‘important’ while those with significantly lower Z-Scores are marked ‘unimportant’ and are removed. This process is repeated until the importance is assigned for all the attributes [14]. The features are ranked in order according to Boruta as seen in table 8.

Attributes	Rank
Age	1
SysBP	1
BMI	2
totChol	3
diaBP	4
Glucose	5
heartRate	6
cigsPerDay	7
prevalentHyp	8
Male	9
Current Smoker	10
Diabetes	11
BPMeds	12
prevalentStroke	13

Table -8: Ranking of Attributes - CVD

Initially, features ranking from 1-6 are considered for training. The final rank is fixed during model evaluation. The features of this dataset are in varied ranges which might cause issues while training. Therefore, it is essential to normalize the data. The Standard Scaler method from the Scikit-learn library is implemented for normalization. This method subtracts the mean from the value (Centering) and divides the value by standard deviation (Scaling). A value is standardized as follows,  
 Standardized value = (Value – Mean) / Standard Deviation

### 3.2.1.2 Diabetes Prediction

The Diabetes study dataset consists of 520 instances & 17 attributes with the last attribute “class” being the target variable. The “Gender” attribute is either Male/Female, the



“class” attribute is either Positive/Negative and all other attributes other than “Age” are either Yes/No. Hence, all the attributes except “Age” are converted to 1’s and 0’s by using the Label Encoder function provided by the Scikit-learn package.

Further analysis confirms the absence of null values in the dataset. The ratio of positive to negative cases is 60:40 and hence, there is no class imbalance issue to be accounted for. The target distribution is shown in figure 20.



Fig -20: Target Distribution

The Correlation study shows that Polyuria and Polydipsia have the highest positive correlation with diabetes with values 0.67 and 0.65 respectively. On the other hand, Gender has the highest negative correlation of 0.45 with diabetes. In addition, only a few attributes such as itching, delayed healing, and obesity have the least correlation while the others have a significant correlation with diabetes. Hence, most of the features must be considered into account which can be achieved by proper feature selection. The correlation heat map of the diabetes study dataset can be seen in figure 21.

Feature selection is carried out using the Boruta algorithm explained in section 3.2.1.1. Table 9 shows the order of the attributes ranked after the application of the Boruta algorithm. There are five features Age, Gender, Polyuria, Polydipsia, and Sudden Weight loss which are of the highest importance. Initially, features ranking from 1-9 are considered for training, the final rank fixation is executed during the evaluation stage.

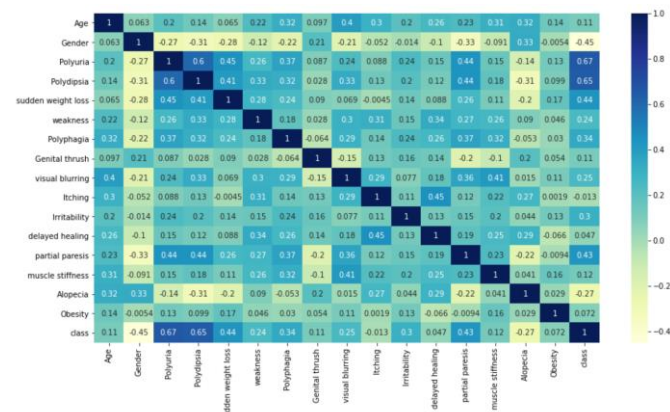


Fig -21: Correlation heat map – Diabetes

Attributes	Rank
Age	1
Gender	1
Polyuria	1
Polydipsia	1
Sudden Weight Loss	1
Partial paresis	2
Irritability	3
Alopecia	4
Polyphagia	5
Delayed healing	6
Itching	7
Visual blurring	8
Weakness	9
Genital thrush	9
Muscle stiffness	11
Obesity	12

Table -9: Ranking of Attributes - Diabetes

### 3.2.2 Model Training

Three classifiers Support Vector Machine (SVM), k-Nearest Neighbor (KNN), and Decision tree are trained for both the prediction tasks. The best values for the hyperparameters for all the three classifiers are obtained by implementing the Grid Search method. The hyperparameters for each classifier are listed in table 10.

To improve the results, stacked generalization is implemented to obtain precision in classification. This method is efficient as it combines the predictions from the individual classifiers present in the stack. It is an ensemble machine learning algorithm. In stacked generalization, several models are combined to generate biases with the help of the training dataset and then filter out the biases. It consists of several base models and one meta model. The specialty of this method is that the meta model learns from the predictions of the base models [15].

The three classifiers SVM, KNN, and Decision Tree are the base models of the stack while Logistic Regression is the meta model. The best hyperparameters obtained using the Grid Search method for the three classifiers are employed while defining the base models in the stack. Overall, four models that include SVM, KNN, Decision Tree, and Stack are trained for both the prediction tasks. The datasets of both the prediction tasks are split at the ratio of 80:20, where 80% of the data are used for training while the remaining 20% for testing.



Models	Hyperparameters	Best Values	
		CVD	Diabetes
SVM	Kernel Co-efficient (C)	10	10
	Regularization Parameter (gamma)	1	0.1
KNN	Number of neighbors (n_neighbors)	2	1
Decision Tree	Max_features	Auto	Auto
	Min_samples_leaf	1	1
	Min_samples_split	2	2

Table -10: Best Hyperparameters after Grid Search

### 3.2.3 Model Evaluation

Evaluation of the four models is accomplished using various performance metrics such as training accuracy, testing accuracy, F1-Score, and Area Under Curve.

For CVD Prediction, the values obtained for all the classifiers are listed in table 11. The values for the stack mentioned are those obtained when features ranking from 1-6 are considered. As F1-Score is the best metric on which the models can be evaluated for medical applications, a separate comparison bar plot is shown in figure 22.

Performance Metrics	SVM	KNN	Decision Tree	Stack (Rank 1-6)
Training Accuracy	92.38%	97.9%	100%	100%
Testing Accuracy	83.61%	84.93%	75.83%	89.65%
F1-Score	82.35%	83.89%	74.43%	88.44%
AUC	91.1%	88.2%	76.1%	96.4%

Table -11: Performance Metrics - CVD

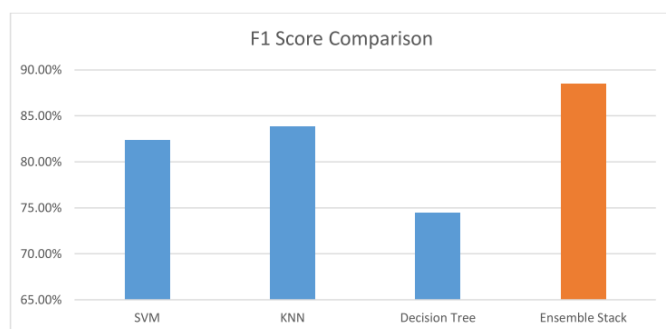


Fig -22: F1-Score Comparison – CVD

From the bar plot, it can be inferred that the stack model has outperformed the three individual classifiers with a score of 88.44%. Having the best model on hand, the performance of the stack with different ranks is analyzed. The line plot for rank vs F1 score is shown in figure 23.

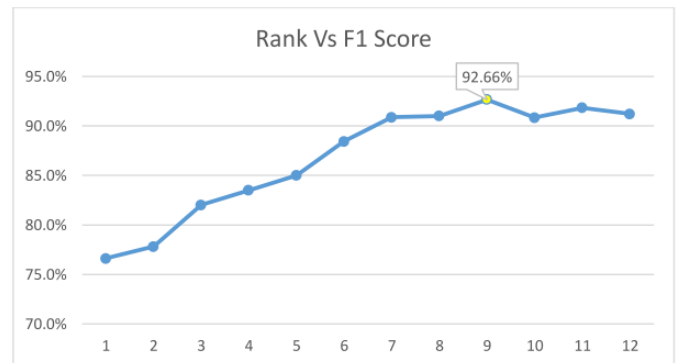


Fig -23: Line plot for Rank Vs F1 Score -CVD

From the line plot, the Stack trained with features ranked from 1-9 has produced the maximum F1-score with the value of 92.66%. Thus, the Stack model with Rank 9 is finalized. The training and testing accuracies for this model are 100% and 92.22% respectively. The ROC-AUC Curve and Confusion matrix for this model are shown in figures 24 & 25. The confusion matrix is obtained using the test data.

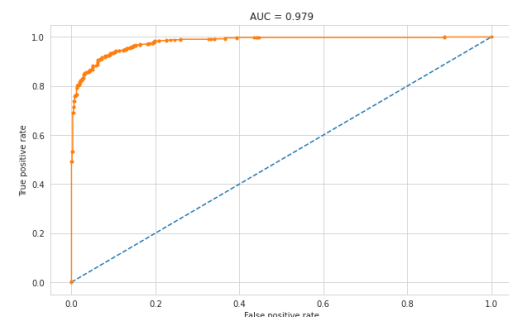


Fig -24: ROC-AUC Curve (Stack) - CVD

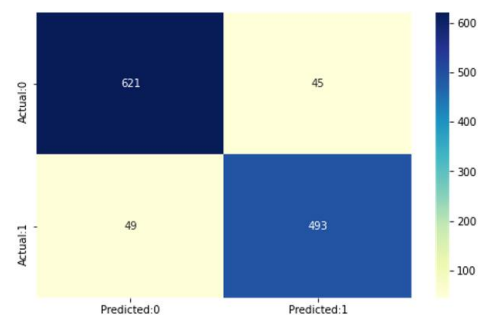


Fig -25: Confusion Matrix (Stack) – CVD

The values of the performance metrics for diabetes prediction are listed in table 12. A separate comparison bar plot on F1-Score is shown in figure 26.

Performance Metrics	SVM	KNN	Decision Tree	Stack (Rank 1-9)
Training Accuracy	100%	100%	100%	100%
Testing Accuracy	96.15%	96.15%	97.12%	99.04%
F1-Score	97.06%	96.97%	97.71%	99.26%
AUC	98.2%	96.4%	97.8%	98.5%

Table -12: Performance Metrics - Diabetes

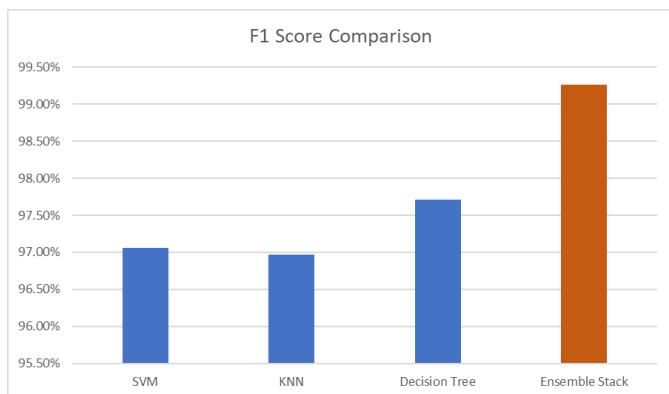


Fig -26: F1-Score Comparison – Diabetes

From the above bar plot, it can be observed that the stack has outperformed the other three individual classifiers. Having the best model on hand, the performance of the stack with different ranks is analyzed. The line plot for rank vs F1 score is shown in figure 27.

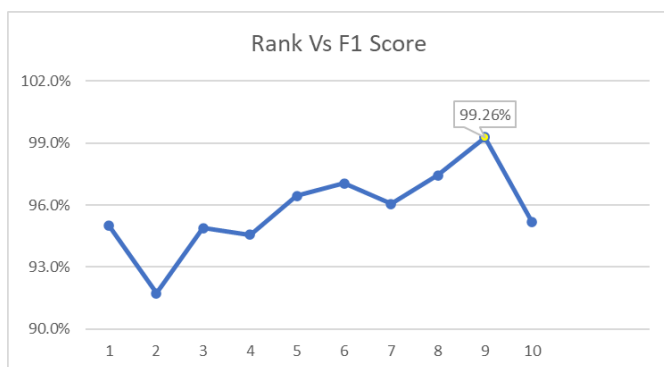


Fig -27: Line plot for Rank Vs F1 Score – Diabetes

From the obtained line plot, it can be inferred that when features ranking from 1-9 is considered, the stack has performed its best with a 99.26% F1 score. Thus, the stack with rank 9 is finalized for diabetes prediction. The training and testing accuracies for this model are 100% and 99.04% respectively. The ROC-AUC curve and Confusion matrix for this model are shown in figures 28 & 29.

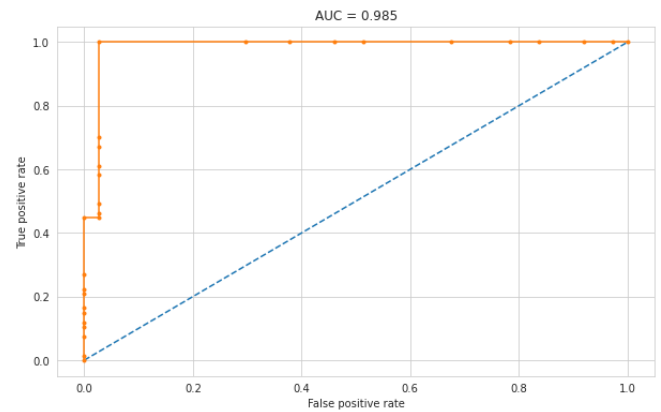


Fig -28: ROC-AUC Curve - Diabetes

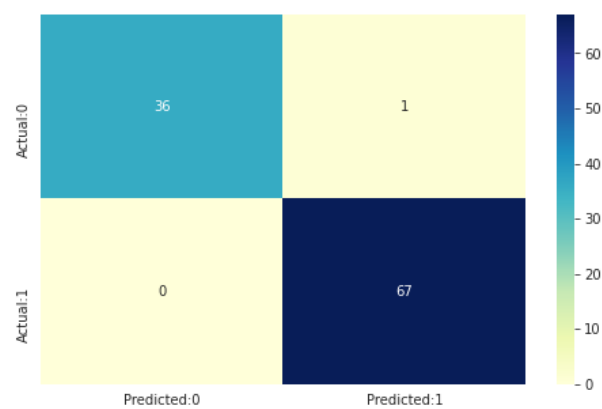


Fig -29: Confusion Matrix – Diabetes

### 3.3 Model Deployment

Streamlit is a python framework used to create a web application for data science and machine learning in a short span of time. Four web applications are separately designed for image-based disease detection (hemorrhage & pneumonia) and risk prediction (CVD & diabetes) using Streamlit. Individual GitHub repositories are created to house all the files (requirements, main python file & model) that are required for web deployment. These repositories are then inherited by Streamlit share for public accessibility.

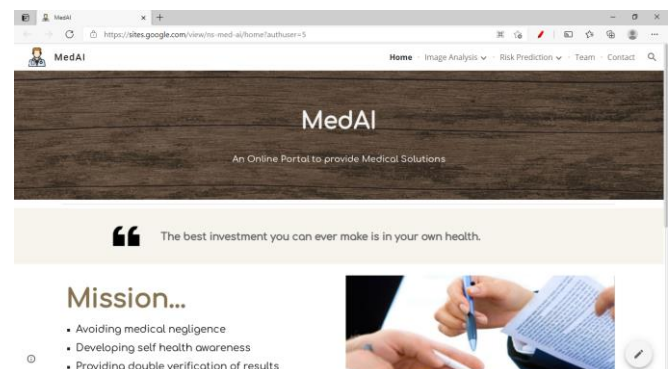


Fig -30: Home Page

To enhance the user interface, all the four Streamlit share sites are embedded using google sites. MedAI, a web portal developed using google sites hosts a home page, an image analysis page, a risk prediction page & a contact page. The screenshot of the home page is shown in figure 30.

#### 4. RESULTS

The screenshots for all the prediction tasks are shown from figures 31 to 34.

Website link: <https://sites.google.com/view/ns-med-ai>

#### Hemorrhage Detection

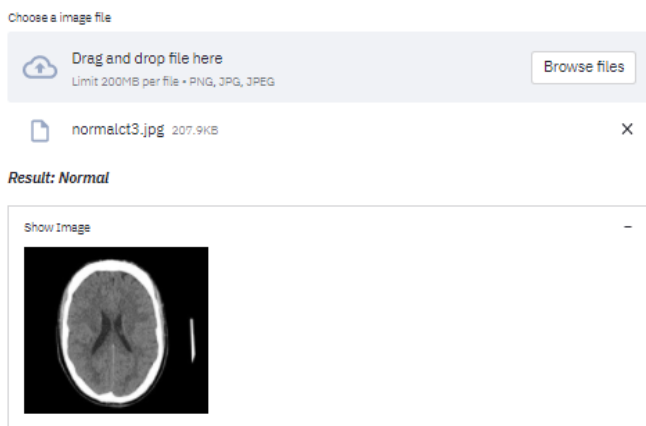


Fig -31: Detection of "Normal" Brain CT

#### Pneumonia Detection

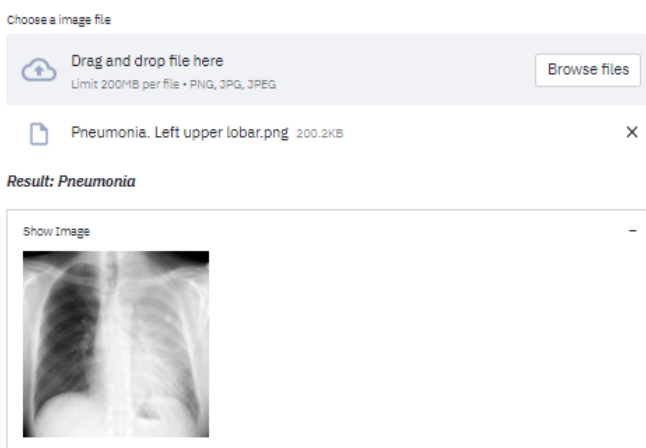


Fig -32: Detection of "Pneumonia" Chest X-Ray

For image-based detection, the input image should be in .png, .jpg or .jpeg format. In the case of risk prediction, if the probability is greater than 50% it is predicted that the person is at risk. In addition, the information regarding the symptoms and the units of the parameters the user has to enter is provided in the portal.

#### Diabetes Risk Prediction

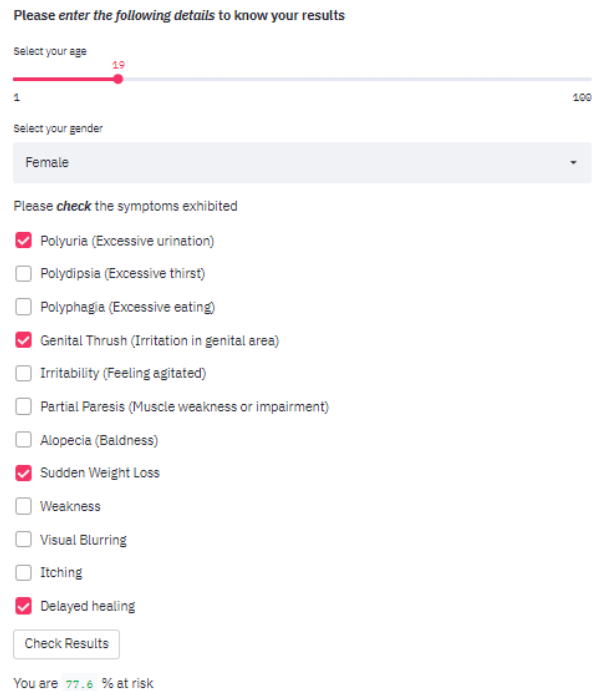


Fig -33: Diabetes Risk Prediction

#### Heart Risk Prediction

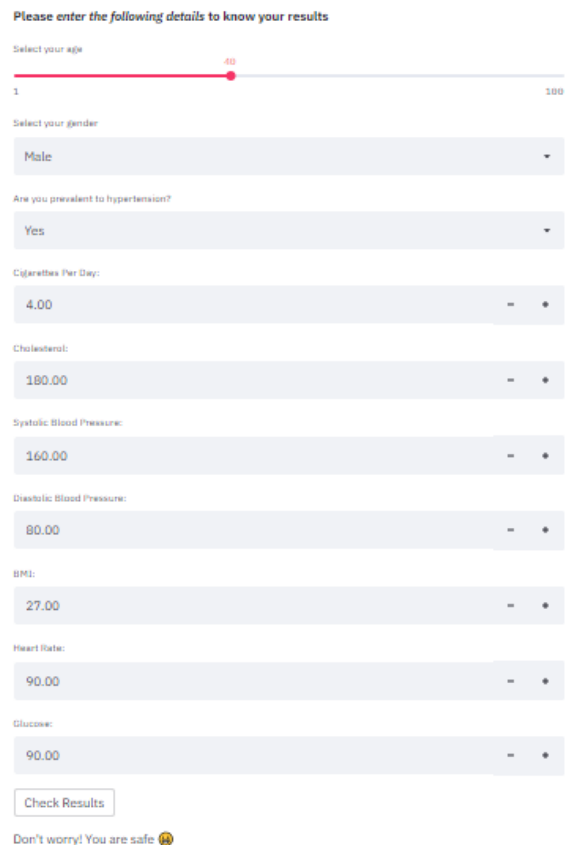


Fig -34: Heart Risk Prediction

## 5. CONCLUSIONS

Overall, a web portal is developed by incorporating all four models. To improve the performance of the models, datasets with more samples are essential. Similarly, other medical diagnosis models can be added to the portal, thereby creating a standalone web access to assist in various medical issues.

## REFERENCES

- [1] Ramandeep Sahni, Jesse Weinberger, "Management of intracerebral hemorrhage", *Vascular Health and Risk Management (Vasc Health Risk Manag)*, Oct 2007
- [2] <https://ourworldindata.org/pneumonia>, Accessed on 6th May 2021
- [3] <https://towardsdatascience.com/review-resnet-winner-of-ilsvrc-2015-image-classification-localization-detection-e39402bfa5d8>, Accessed on 6th May 2021
- [4] V. Sirish Kaushik, Anand Nayyar, Gaurav Kataria and Rachna Jain, "Pneumonia Detection Using Convolutional Neural Networks (CNNs)", *Proceedings of First International Conference on Computing, Communications, and Cyber-Security (IC4S 2019)*, Apr 2020, doi:10.1007/978-981-15-3369-3\_36
- [5] <https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-cvds>, Accessed on 7th May 2021
- [6] <https://www.who.int/news-room/fact-sheets/detail/diabetes>, Accessed on 7th May 2021
- [7] Archana Singh, Rakesh Kumar, "Heart Disease Prediction Using Machine Learning Algorithms", *2020 International Conference on Electrical and Electronics Engineering (ICE3)*, Feb 2020, doi: 10.1109/ICE348803.2020.9122958
- [8] Muhammad Azeem Sarwar, Nasir Kamal, Wajeeha Hamid, Munam Ali Shah, "Prediction of diabetes Using Machine Learning Algorithms", *24th International Conference on Automation and Computing (ICAC)*, Sep 2018, doi:10.23919/ICAC.2018.8748992
- [9] Kaggle, "Head CT- hemorrhage", 2018. Available: <https://www.kaggle.com/felipekitamura/head-ct-hemorrhage>, Accessed on 26th Jan 2021.
- [10] Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification", *Mendeley Data*, V2, doi: 10.17632/rsbjbr9sj.2
- [11] Kaggle, "Framingham Heart study dataset", 2017. Available: <https://www.kaggle.com/amanajmera1/framingham-heart-study-dataset>, Accessed on 26th Jan 2021
- [12] UCI ML Repository, Center for Machine Learning and Intelligent Systems, Available: <https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset>, Accessed on 26th Jan 2021
- [13] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, W. Philip Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique", *Journal of Artificial Intelligence Research* 16(1):321-357, Jun 2002, doi:10.1613/jair.953
- [14] Maya Gopal P.S, Bhargavi R, "Feature Selection for Yield Prediction Using Boruta Algorithm", *International*

*Journal of Pure and Applied Mathematics*, Volume 118 No. 22 2018, 139-144

- [15] Ma Z, Wang P, Gao Z, Wang R, Khalighi K, "Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose", *PLoS ONE* 13(10): e0205872, 2018, <https://doi.org/10.1371/journal.pone.0205872>