A Deep Learning Model for Multiple Disease Prediction using VGG-16

Nemani Anil
ISE department
Acharya Institute of Technology Bangalore
nemanin.19.beis@acharya.ac.in

Harsha Vardhana
ISE department
Acharya Institute of Technology Bangalore
harshal.19.beis@acharya.ac.in

Kiran Raju L
ISE department
Acharya Institute of Technology Bangalore
kirank.19.beis@acharya.ac.in

Preethi L
ISE department
Acharya Institute of Technology Bangalore
preethil.19.beis@acharya.ac.in

***Abstract*— All health-related problems must be precisely and professionally examined if it is to be prohibited and examined. In the case of a captious illness, the common technique of diagnosis might not be adequate. Early diagnosis of any health issue is crucial, as earlier disease discovery may result in more successful treatments or longer survival times.**

We may have serious illness, and if we do not address these symptoms right away, it may be too late to reverse the condition. The establishment of a Deep Learning (DL)-based medical analysis systems for disease forecast can lead to a prognosis that is more accurate than one made using historical techniques. Currently, the suggested approach can forecast Covid-19, Brain tumor, Breast Cancer, Alzheimer, Diabetes, Pneumonia. We can employ, a NN (Neural Network) known as CNN (Convolution Neural Network), which VGG 16 architectures for deep learning (visual geometry group) and random forest. For the early detection of a sickness, our diagnostic model can serve as a doctor, enabling timely treatment and the possibility of saving lives.

Keywords- CNN (Convolutional neural networks), VGG-16, Deep Learning, Machine Learning and Random Forest.

I. Introduction

Artificial intelligence has enhanced machine brain power and has given the capacity to think. Machine- learning and Several studies have been conducted on the many branches of AI known as deep learning. According to data, India has a doctor-to-population ratio of 1:1456 compared to the WHO's suggested ratio of 1 doctor to 1000 patients, demonstrating a doctor shortage[5]. This will lessen the congestion at hospital OPDs and reduce the workload for the medical staff. Additionally, this approach will lessen the need for expensive treatment and fear in the final phases, allowing for the timely administration of appropriate medication and a reduction in the fatality rate.Machines have always been considered as more skillful to humans because they can complete jobs more quickly and continually with a high level of accuracy without the prospect of human error. A disorder predictor, known as a virtu doctor, is able to accurately predict any patient's ailment without the involvement of a human. A disease predictor can also be a gift in cases like COVID-19 and EBOLA because it can detect a person's illness without any physical contact[7]. In order to diagnose a patient and provide treatments, the Real Time health Prediction steps at this point. Therefore, putting in place an applications that eventually shapes a proper health predictive model can provide patients with immediate guidance and help stop diseases from progressing to later stages[2]. Countless lives can be saved by timely data analysis and improving disease prediction performance.

As technology developed, deep learning gained popularity because of increasingly powerful computers and the accessibility of datasets from open-source sources. Therefore, developing a deep learning model, applying it to the data set, and adding specific patient data can help with prediction. The outcome of the analysis will be dependent on the data provided, making it specific to that person[4]. Doctors utilise a variety of scientific techniques and procedures to identify and diagnose not only common illnesses but also a number of fatal conditions.

A successful therapy is almost always attributed to an accurate and correct diagnosis. Disease prediction systems that use machine learning algorithms help to achieve reliable findings in such situations because doctors occasionally make mistakes when diagnosing a patient's illness.

Regretfully, people often overlook their health because they are too preoccupied with their everyday tasks. Children and elderly people are both capable of ignoring or failing to recognize the crucial indicators that can later lead to more serious problems. It is advisable to get treatment before the illness worsens and progresses. Such individuals can benefit from preventative care and early detection of their health conditions with the use of a prediction system. This facilitates access to primary health services in isolated regions.

Several experts argue that learning is necessary for the development of insight. These approaches can be unsupervised, semi-supervised, supervised, reinforcement learning, evolutionary learning, and deep leaning, to name a few.[11]

Here we are using algorithms like random forest and CNN. A A Deep-Learning technique called a convolutional network may take a picture as input, assign various elements and objects in the image importances, and be able to discern between them.

Random Forest is a classifier that employs many decision trees on various subsets of the provided dataset and computes an overall average to increase the dataset's predictive accuracy. The more trees there are in the forest, the higher the accuracy and the less overfitting there is.

II .METHODOLOGY

3.a Existing system

Many of the analyses that have already been done looked at specific diseases. A user must apply one analysis when they want to analyze diabetes, and another analysis when they want to analyze heart disease[9]. This process takes a while. Additionally, if a user has multiple diseases but the system in place can only anticipate one of them, there is a potential that the death rate may rise as a result of the inability to foresee the other diseases.

3.b Proposed system

Multi-disease detection models can detect more than one disease simultaneously. Therefore, users do not have to look at any models for disease detection. This shortens time and help increasing mortality due to simultaneous forecasting by multiple diseases. In this model we combined structured and unstructured data from the healthcare sectors to assess illness risk. The creation of a latent component model to missing information in online-sourced medical records. We employ the convolution neural network method to automatically choose features for unstructured files and random forest for structured data. CNN and Random Forest algorithm are used to train the models. The Trained CNN and Random Forest Models in the form of pickle files are consumed by the flask file housing at the local system.

3.c Algorithm

Machine learning's most well-known algorithm Random Forest belongs to the ensemble learning family. This is an improved decision tree that combines multiple decision trees to make predictions. As the name describes, "A random forest is a classifier that averages a set of decision trees from various subsets of a given dataset to increase the accuracy of the dataset's prediction of correctness." Instead of relying on trees, it takes predictions from each tree and Based on the majority vote of the predictions, random forest predicts the outcome.

A. RANDOM FOREST

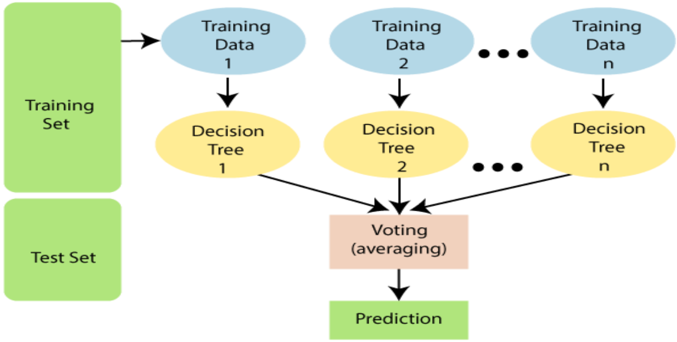


Fig1- RANDOM FOREST ARCHITECTURE

B. CNN

Transfer--learning is an experience and understanding method that shortens the amount of training data needed, as well as the time and money needed to develop learning models. You can transfer training from a new model using a pre-trained model with the aid of deep learning. Uses for transfer learning include the classification of tumors, the prediction of software flaws, the identification of activities, and the classification of moods [7]. In this instance, we contrasted the effectiveness of the Deep CNN model put forth by with that of the well-liked transfer learning technique VGG16.

C. Visual Geometry Group 16

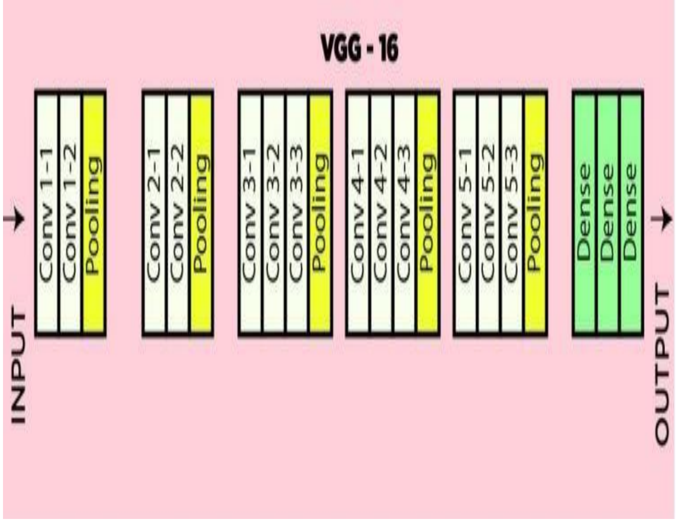


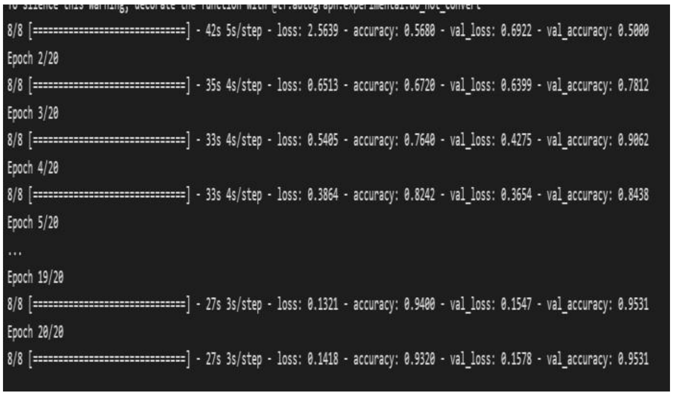
Fig 2 VGG-16 ARCHITECTURE

III RESULT ANALYSIS

DISEASES	PERFORMANCE			
	KNN	DT	ANN	ACCURACY
COVID 19	85.68%	88%	91.1%	95%
BRAIN TUMOR	89%	90.69%	92%	94%
ALZHEIMER	78.56%	79.35%	80%	81%
BREAST CANCER	82%	85.58%	90.32%	93%
DIABETES	77%	77.89%	78%	79%
Pneumonia	84%	86.95%	88%	90%

TABLE 1 Accuracy of Different Algorithms

C . COVID -19

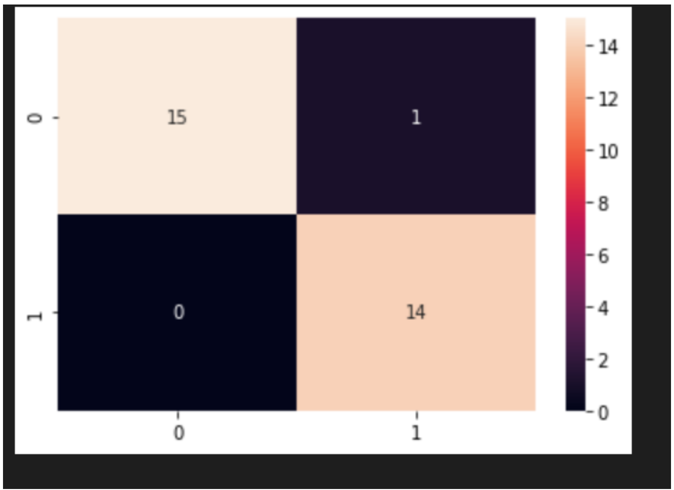


```
To silence this warning, decorate the function with @tf.autograph.experimental.do_not_convert
8/8 [==========================] - 42s 5s/step - loss: 2.5639 - accuracy: 0.5680 - val_loss: 0.6922 - val_accuracy: 0.5000
Epoch 2/20
8/8 [==========================] - 35s 4s/step - loss: 0.6513 - accuracy: 0.6720 - val_loss: 0.6399 - val_accuracy: 0.7812
Epoch 3/20
8/8 [==========================] - 33s 4s/step - loss: 0.5405 - accuracy: 0.7640 - val_loss: 0.4275 - val_accuracy: 0.9062
Epoch 4/20
8/8 [==========================] - 33s 4s/step - loss: 0.3864 - accuracy: 0.8242 - val_loss: 0.3654 - val_accuracy: 0.8438
Epoch 5/20
...
Epoch 19/20
8/8 [==========================] - 27s 3s/step - loss: 0.1321 - accuracy: 0.9400 - val_loss: 0.1547 - val_accuracy: 0.9531
Epoch 20/20
8/8 [==========================] - 27s 3s/step - loss: 0.1418 - accuracy: 0.9320 - val_loss: 0.1578 - val_accuracy: 0.9531
```

Fig 3 Covid-19 disease accuracy here we can see that training racy and validation accuracy is after 20 epochs

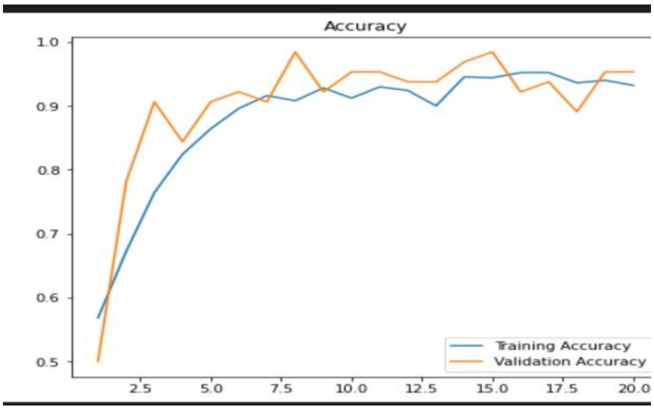
D. Training and Accuracy of covid 19

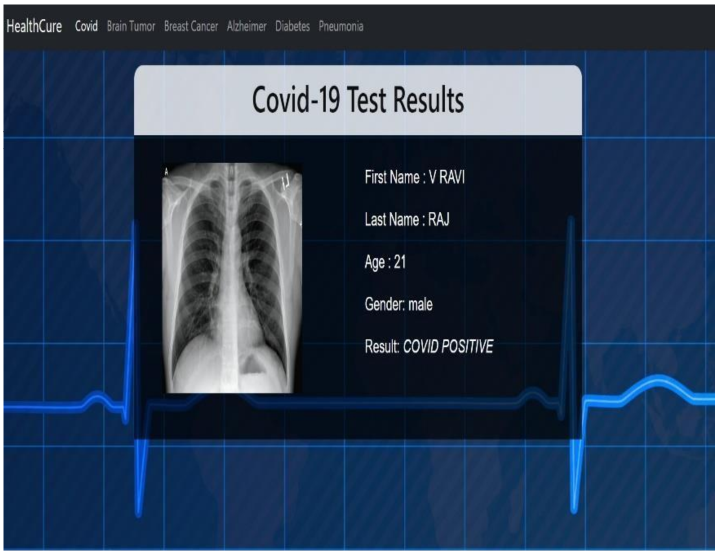
Fig 4: - This figure represents the covid 19 Training andaccuracy graph



E. Confusion matrix

Fig 5 Covid-19 output is been represented and the result isShowed to the user by loading the model





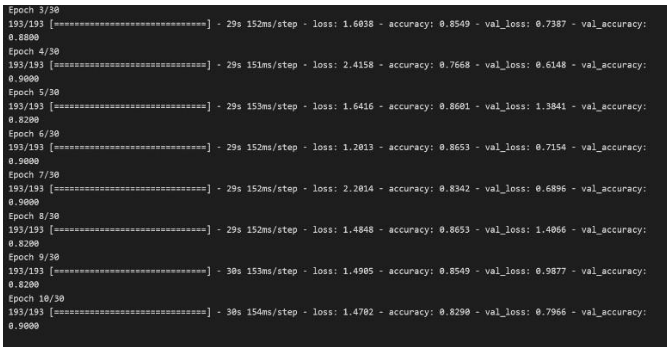
First Name : V RAVI

Last Name : RAJ

Age : 21

Gender: male

Result: *COVID POSITIVE*



```
Epoch 3/30
193/193 [==============================] - 29s 152ms/step - loss: 1.6038 - accuracy: 0.8549 - val_loss: 0.7387 - val_accuracy:
0.8800
Epoch 4/30
193/193 [==============================] - 29s 151ms/step - loss: 2.4158 - accuracy: 0.7668 - val_loss: 0.6148 - val_accuracy:
0.9000
Epoch 5/30
193/193 [==============================] - 29s 153ms/step - loss: 1.6416 - accuracy: 0.8601 - val_loss: 1.3841 - val_accuracy:
0.8200
Epoch 6/30
193/193 [==============================] - 29s 152ms/step - loss: 1.2013 - accuracy: 0.8653 - val_loss: 0.7154 - val_accuracy:
0.9000
Epoch 7/30
193/193 [==============================] - 29s 152ms/step - loss: 2.2014 - accuracy: 0.8342 - val_loss: 0.6896 - val_accuracy:
0.9000
Epoch 8/30
193/193 [==============================] - 29s 152ms/step - loss: 1.4848 - accuracy: 0.8653 - val_loss: 1.4066 - val_accuracy:
0.8200
Epoch 9/30
193/193 [==============================] - 30s 153ms/step - loss: 1.4905 - accuracy: 0.8549 - val_loss: 0.9877 - val_accuracy:
0.8200
Epoch 10/30
193/193 [==============================] - 30s 154ms/step - loss: 1.4702 - accuracy: 0.8290 - val_loss: 0.7966 - val_accuracy:
0.9000
```

G.Training and Accuracy of Brain Tumor

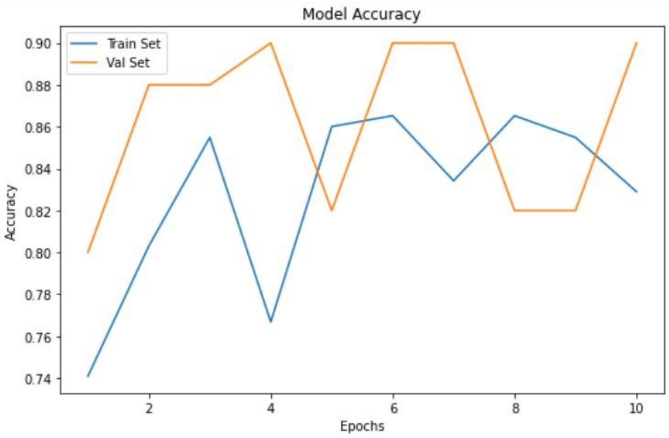


Fig 8: - this figure represents the covid 19 Training and accuracy graph

H. Confusion Matrix

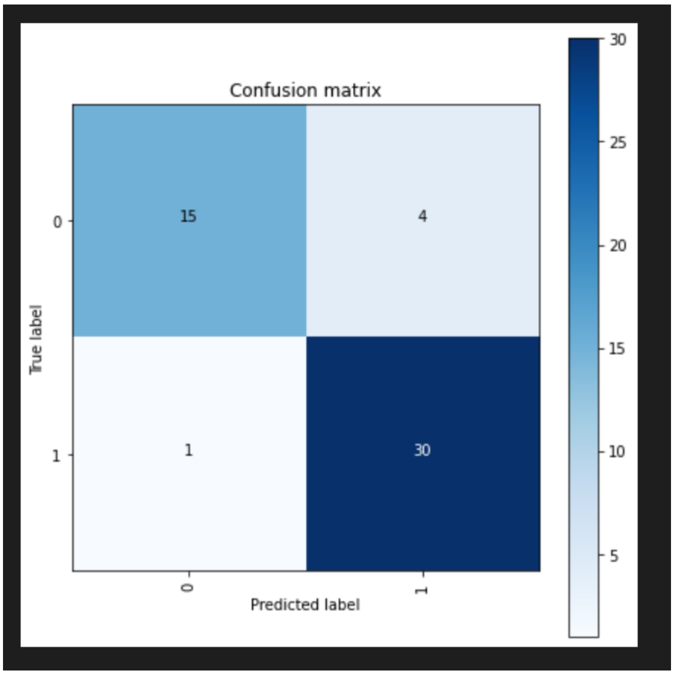
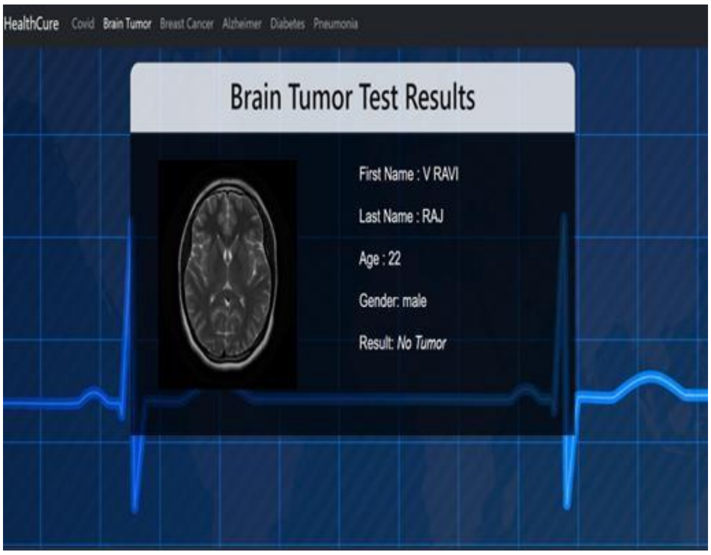


Fig 9 Confusion matrix of brain tumor.

I.Brain Tumor



Brain Tumor Test Results

First Name : V RAVI

Last Name : RAJ

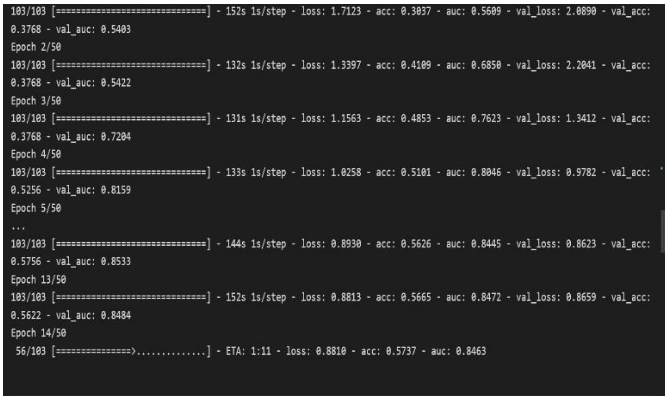
Age : 22

Gender: male

Result: *No Tumor*

Fig 10 The Brain Tumor result is shown and the user can see the result by loading out the corresponding model

J. ALZHEIMER



```
103/103 [==============================] - 152s 1s/step - loss: 1.7123 - acc: 0.3037 - auc: 0.5609 - val_loss: 2.0890 - val_acc:
0.3768 - val_auc: 0.5403
Epoch 2/50
103/103 [==============================] - 132s 1s/step - loss: 1.3397 - acc: 0.4109 - auc: 0.6850 - val_loss: 2.2041 - val_acc:
0.3768 - val_auc: 0.5422
Epoch 3/50
103/103 [==============================] - 131s 1s/step - loss: 1.1563 - acc: 0.4853 - auc: 0.7623 - val_loss: 1.3412 - val_acc:
0.3768 - val_auc: 0.7204
Epoch 4/50
103/103 [==============================] - 133s 1s/step - loss: 1.0258 - acc: 0.5101 - auc: 0.8046 - val_loss: 0.9782 - val_acc:
0.5256 - val_auc: 0.8159
Epoch 5/50
...
103/103 [==============================] - 144s 1s/step - loss: 0.8930 - acc: 0.5626 - auc: 0.8445 - val_loss: 0.8623 - val_acc:
0.5756 - val_auc: 0.8533
Epoch 13/50
103/103 [==============================] - 152s 1s/step - loss: 0.8813 - acc: 0.5665 - auc: 0.8472 - val_loss: 0.8659 - val_acc:
0.5622 - val_auc: 0.8484
Epoch 14/50
 56/103 [==============>...............] - ETA: 1:11 - loss: 0.8810 - acc: 0.5737 - auc: 0.8463
```

Fig 11 ALZHEIMER disease accuracy here we can see that training racy and validation accuracy is after 15 epochs.

K. Training and Accuracy of Alzheimer

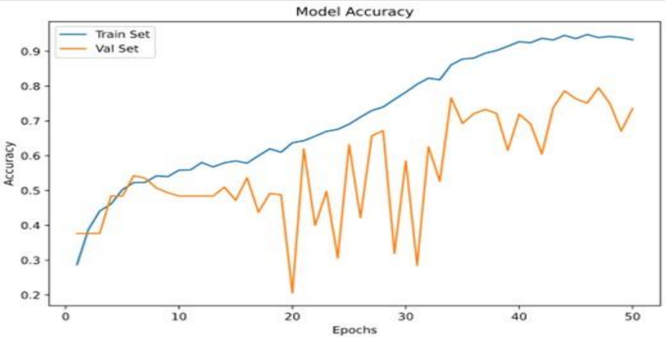
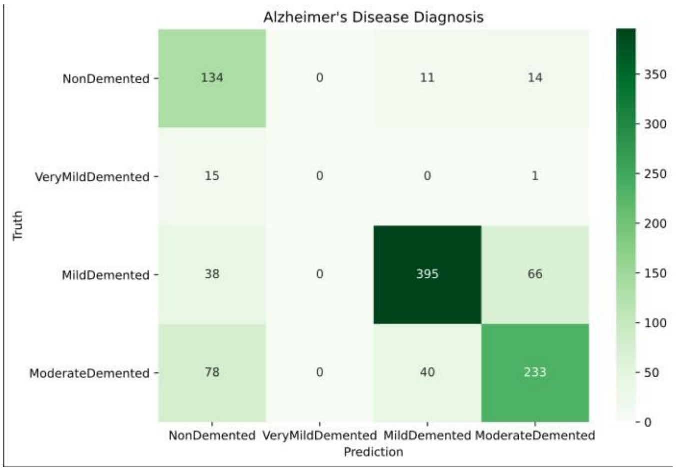


Fig 12: - this figure represents the Alzheimer Training and accuracy graph

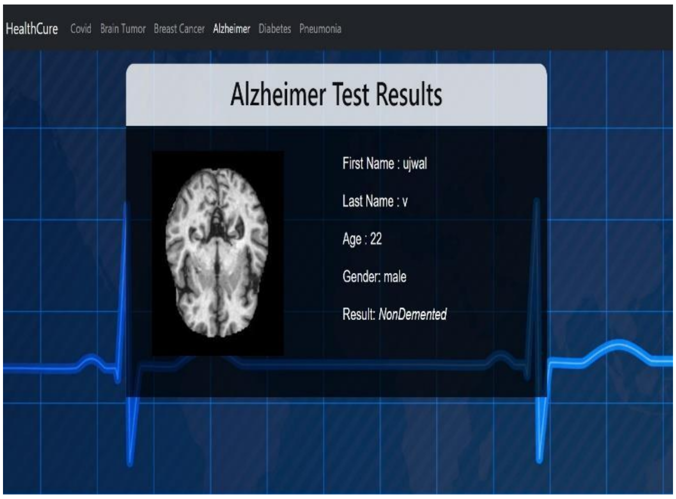
L. Confusion Matrix



	NonDemented	VeryMildDemented	MildDemented	ModerateDemented
NonDemented	134	0	11	14
VeryMildDemented	15	0	0	1
MildDemented	38	0	395	66
ModerateDemented	78	0	40	233

Fig 13 Confusion matrix of Alzheimer.

M. Alzheimer



Alzheimer Test Results

First Name : ujwal
Last Name : v
Age : 22
Gender: male
Result: *NonDemented*

Fig 14 The output of the Alzheimer disease is shown by loading the corresponding model and showing it to the end user.

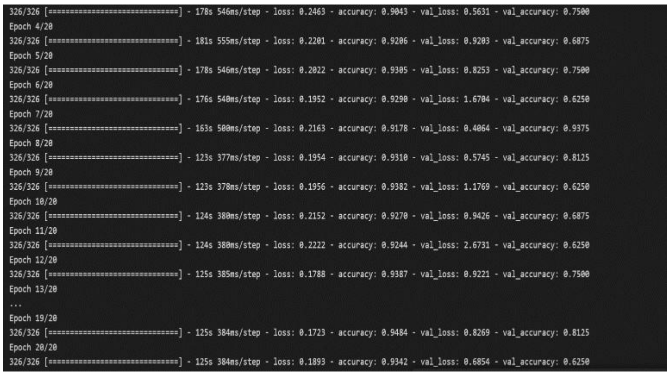
N. DIABETES

Diabetes Test Results

First Name : UJWAL
Last Name : V
Age : 22
Gender: male
Result: *NEGATIVE*

Fig 15 Diabetes output is shown and the model is loaded Showing the result to the end user

O. Pneumonia



```
326/326 [==============================] - 178s 546ms/step - loss: 0.2463 - accuracy: 0.9043 - val_loss: 0.5631 - val_accuracy: 0.7500
Epoch 4/20
326/326 [==============================] - 181s 555ms/step - loss: 0.2201 - accuracy: 0.9206 - val_loss: 0.9203 - val_accuracy: 0.6875
Epoch 5/20
326/326 [==============================] - 178s 546ms/step - loss: 0.2022 - accuracy: 0.9305 - val_loss: 0.8253 - val_accuracy: 0.7500
Epoch 6/20
326/326 [==============================] - 176s 540ms/step - loss: 0.1952 - accuracy: 0.9290 - val_loss: 1.6704 - val_accuracy: 0.6250
Epoch 7/20
326/326 [==============================] - 163s 500ms/step - loss: 0.2163 - accuracy: 0.9178 - val_loss: 0.4064 - val_accuracy: 0.9375
Epoch 8/20
326/326 [==============================] - 123s 377ms/step - loss: 0.1954 - accuracy: 0.9310 - val_loss: 0.5745 - val_accuracy: 0.8125
Epoch 9/20
326/326 [==============================] - 123s 378ms/step - loss: 0.1956 - accuracy: 0.9382 - val_loss: 1.1769 - val_accuracy: 0.6250
Epoch 10/20
326/326 [==============================] - 124s 380ms/step - loss: 0.2152 - accuracy: 0.9270 - val_loss: 0.9426 - val_accuracy: 0.6875
Epoch 11/20
326/326 [==============================] - 124s 380ms/step - loss: 0.2222 - accuracy: 0.9244 - val_loss: 2.6731 - val_accuracy: 0.6250
Epoch 12/20
326/326 [==============================] - 125s 385ms/step - loss: 0.1788 - accuracy: 0.9387 - val_loss: 0.9221 - val_accuracy: 0.7500
Epoch 13/20
...
Epoch 19/20
326/326 [==============================] - 125s 384ms/step - loss: 0.1723 - accuracy: 0.9484 - val_loss: 0.8269 - val_accuracy: 0.8125
Epoch 20/20
326/326 [==============================] - 125s 384ms/step - loss: 0.1893 - accuracy: 0.9342 - val_loss: 0.6854 - val_accuracy: 0.6250
```

Fig 16 Pneumonia disease accuracy here we can see that training accuracy and validation accuracy is after 20 epochs.

P. Training and Accuracy of Pneumonia

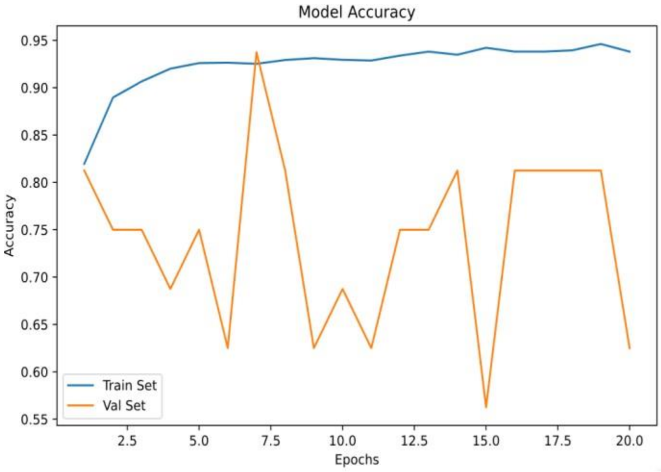
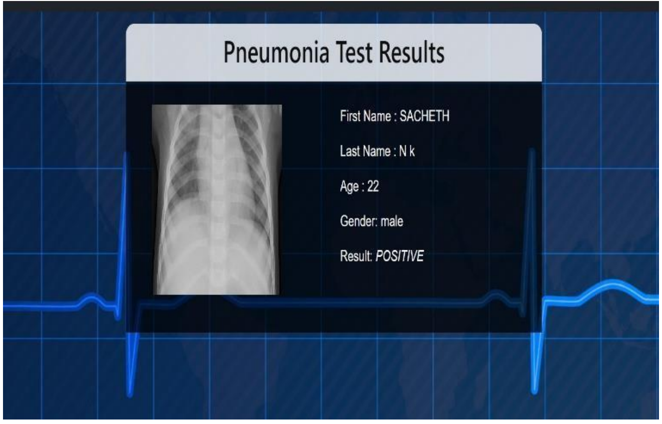


Fig 17: - this figure represents the Pneumonia Training and accuracy graph.

Q. Pneumonia



Pneumonia Test Results

First Name : SACHETH
Last Name : N k
Age : 22
Gender: male
Result: *POSITIVE*

Fig 18 The pneumonia disease output is shown By loading the model and is visible to the end user.

1V CONCLUSION

Multiple disease detection using VGG16 is a technique that utilizes a pre-trained convolutional neural network (CNN) architecture known as VGG16 for the classification of diseases. The VGG16 architecture was trained on a large dataset of images, allowing it to recognize patterns in medical images that are indicative of specific diseases. The conclusion of multiple disease detection using VGG16 is that it has the potential to boost the correctness and speed of disease diagnosis, especially in the case of complex diseases that are difficult to diagnose through traditional methods.

However, it is important to note that while VGG16 has shown promising results in several studies, it is not a substitute for traditional diagnostic methods and should be used in conjunction with other diagnostic tools and clinical expertise. Additionally, the performance of VGG16 in disease detection may vary depending on the quality of the training dataset and the specific application. Further research and development are required to validate and improve the performance of VGG16 in multiple disease detection.
The conclusion of multiple disease detection is important because it can help healthcare providers diagnose and treat illnesses more effectively, leading to improved patient outcomes. However, it is important to note that accurate diagnosis requires a combination of multiple diagnostic methods and a thorough understanding of the patient's medical history. Additionally, the interpretation of test results must be done by a qualified healthcare provider to ensure the correct diagnosis and treatment plan are established.

V . ACKNOWLEDGEMENT

We express our regard to Mr. Lokanna Kadakolmath Assistant Professor, our project guide, and Mr. M K Dhananjaya Assistant Professor, for continuously monitoring the project's progress and establishing exact timelines. Their insightful counsel served as the driving force behind finishing the task progress and establishing exact timelines of the task.

V11. REFERENCES

[1] Aayushi Agrawal, Shubham Dalal, Dhvanan Rangrej and Nitin Choubey,“Multiple Disease Prediction System” International Journal of Recent Advancesin Multidisciplinary TopicsVolume 2, Issue 10, October2021

[2] Dhiraj Dahiwade,“ Designing Disease Prediction Model Using Machine Learning Approach” IEEE: Proceedings of the Third International Conference on Computing Methodologies and Communication

[3] Md. Ehtisham Farooqui, “A Detailed Review on Disease Prediction Models that uses Machine Learning” International Journal of Innovative Research in Computer Science & Technology (IJIRCST) ISSN: 2347-5552, Volume, 8, Issue, 4, July, 2020

[4] Muhammad Nabeel , Shumaila Majeed , Mazhar Javed Awan, Hooria Muslih-ud-Din , Mashal Wasique and Rabia Nasir, “Review on Effective Disease Prediction through Data Mining Techniques” International Journal on Electrical Engineering and Informatics - Volume 13, Number 3, September 2021

[5] Akkem Yaganteeswarudu, “Multi Disease Prediction Model by using Machine Learning and Flask API” IEEE: Proceedings of the Fifth International Conference on Communication and Electronics Systems

[6] A Detailed Review on Disease Prediction Models that uses Machine Learning Md. Ehtisham Farooqui, Dr. Jameel Ahmad

[7] Rudra A. Godse, Smita S. Gunjal, Karan A. Jagtap, Neha R. Mahamuni, “Multiple Disease Prediction Using Different Machine Learning Algorithms Comparatively” International Journal of Advanced Research in Computer and Communication Engineering

[8] Abhishek Singh, “Prediction and Analysis of Multiple Diseases Using Machine Learning Techniques” INTERNATIONAL RESEARCH JOURNAL OF ENGINEERING AND TECHNOLOGY

[9] Marouane Ferjani, Dr. Jameel Ahmad, “Disease Prediction System using Support Vector Machine and Multilinear Regression”

[10] K. Arumugam a , Mohd Naved b , Priyanka P. Shinde, Orlando Leiva-Chauca d , Antonio Huaman-Osorio e and Tatiana Gonzales-Yanac, “Multiple disease prediction using Machine learning algorithms”

[11] Aditya Sharad Ahirrao, “Multi Disease Detection and Predictions Based On Machine Learning” Electonic journal February 2020

[12] Priyanka Shinde, “Multiple disease prediction using Machine learning algorithms”

[13] Divya Mandem and B. Prajna, “Multi Disease Prediction System”, November 2021| IJIRT | Volume 8 Issue 6 | ISSN: 2349-6002

[14] Mehfuza Holia and Ankita Parmar, “A survey on Brain Tumor detection using deep learning”,International Journal of Advance Engineering and Research

[15] Akash C. Jamgade, “Disease Prediction Using Machine Learning”, International Research Journal of Engineering and Technology