

# Project Report

## HealthCure

Prepared By: Team Data Ninjas

Rajvi Shah

Yash Kamtekar

Sarjak Patel



## **Abstract**

Worldwide, infectious disease continues to emerge due to various factors such as urbanization, traveling of people from one place to another. The first virus detected was tobacco mosaic virus in 1892 and later in 1898 foot-and-mouth disease virus which and the first detected in humans was the yellow fever virus in 1901. New species of human viruses and diseases are still being detected every year such as the recent COVID-19 which may have originated in an animal and changed (mutated) so it could cause illness in humans. Lungs are the main organs affected by the COVID-19, however it can also affect other organs of the body. Other diseases such as Brain Tumor, Breast Cancer, Alzheimer, Diabetes, Pneumonia and Heart Disease are some of the diseases that, when detected at an early stage, can help cure this disease without going through a surgery.

Artificial intelligence can assist providers in a variety of patient care and intelligent health systems. Artificial Intelligence techniques ranging from machine learning to deep learning are prevalent in healthcare for disease diagnosis. Artificial intelligence techniques are most efficient in identifying the diagnosis of different types of diseases. The presence of computerized reasoning as a method for improved medical services offers unprecedented opportunities to recuperate patient and clinical group results, decrease costs, etc.

## **Introduction**

Healthcare is shaping up in front of our eyes with advances in digital healthcare technologies such as artificial Intelligence, 3D printing, robotics, nanotechnologies, etc. Artificial intelligent algorithms must be trained on population-representative information to accomplish presentation levels essential for adaptable “accomplishment”.

Enlargement in health care data struggles with the lack of well-organized mechanisms for integrating and reconciling these data ahead of their current silos. However, numerous frameworks and principles facilitate summation and accomplish adequate data quantity for AI.

## Related Work

**VGG16** is a convolution neural net (CNN ) architecture which was used to win ILSVR(Imagenet) competition in 2014. It is considered to be one of the excellent vision model architecture till date. Most unique thing about VGG16 is that instead of having a large number of hyper-parameter they focused on having convolution layers of 3x3 filter with a stride 1 and always used same padding and maxpool layer of 2x2 filter of stride 2. It follows this arrangement of convolution and max pool layers consistently throughout the whole architecture.

**Random forest** is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression. One of the most important features of the Random Forest Algorithm is that it can handle the data set containing continuous variables as in the case of regression and categorical variables as in the case of classification. It performs better results for classification problems.

**XGBoost**, which stands for Extreme Gradient Boosting, is a scalable, distributed gradient-boosted decision tree (GBDT) machine learning library. It provides parallel tree boosting and is the leading machine learning library for regression, classification, and ranking problems.

It's vital to an understanding of XGBoost to first grasp the machine learning concepts and algorithms that XGBoost builds upon: supervised machine learning, decision trees, ensemble learning, and gradient boosting.

A **Convolutional Neural Network (CNN)** is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data. CNNs are powerful image processing, artificial intelligence that use deep learning to perform both generative and descriptive tasks, often using machine vision that includes

image and video recognition, along with recommender systems and natural language processing.

The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.

The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex.

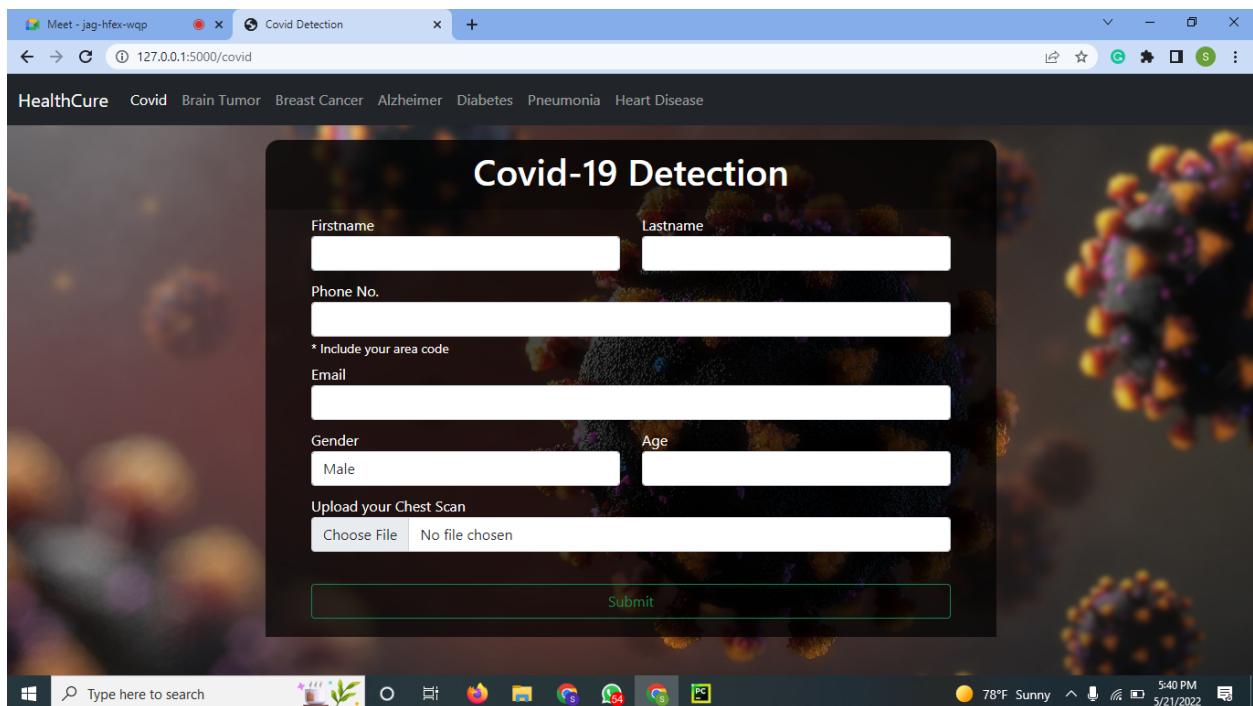
A ConvNet is able to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters.

The role of the ConvNet is to reduce the images into a form that is easier to process, without losing features that are critical for getting a good prediction.

The following are the disease detections that we have done in our project

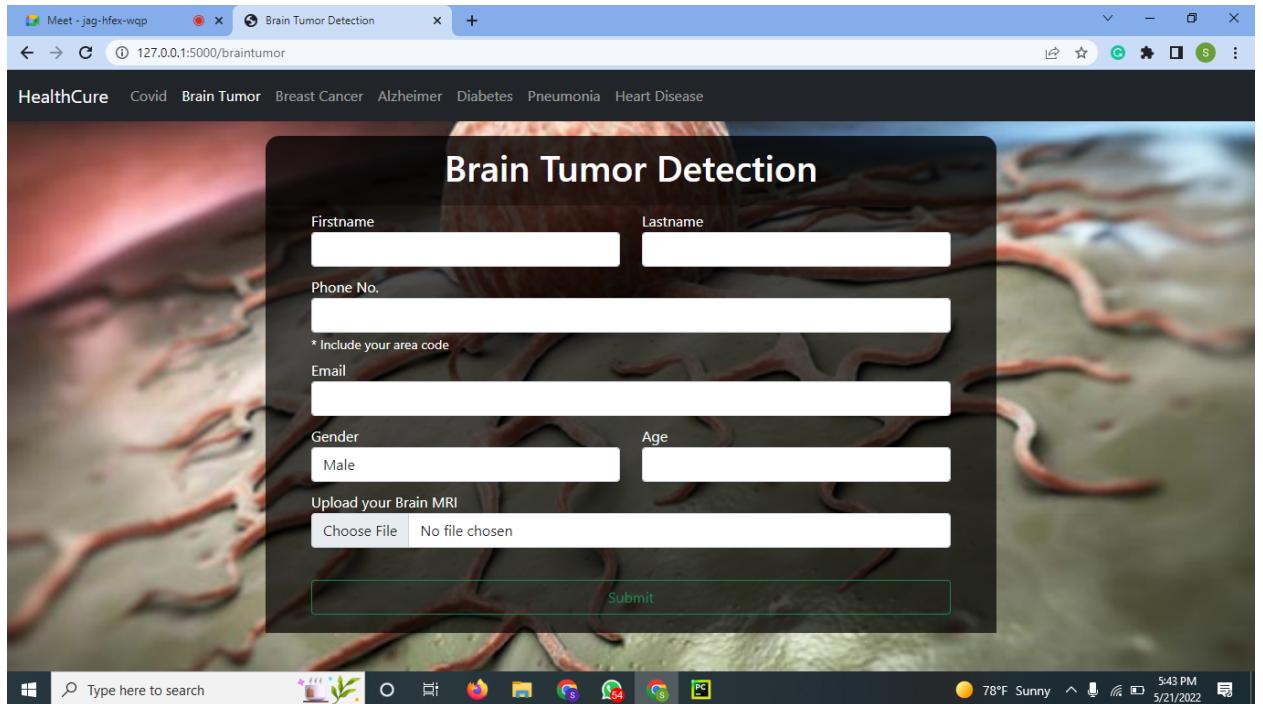
## 1. Covid-19 Detection

We have used a custom based CNN architecture for this decision and we achieved an accuracy of 93%.



## 2. Brain Tumor Detection

We have used VGG-16 for feature extractions and used a custom-made CNN and achieved an accuracy of 100% as we tested this only for 10 images.



### 3. Breast Cancer Detection

We have used Random Forest for this and achieved an accuracy of 91.81%.

The screenshot shows a web browser window titled "Breast Cancer Detection" with the URL "127.0.0.1:5000/breastcancer". The page has a dark header with the "HealthCure" logo and navigation links for Covid, Brain Tumor, Breast Cancer, Alzheimer, Diabetes, Pneumonia, and Heart Disease. The main content area is titled "Breast Cancer Detection" and contains a form with the following fields:

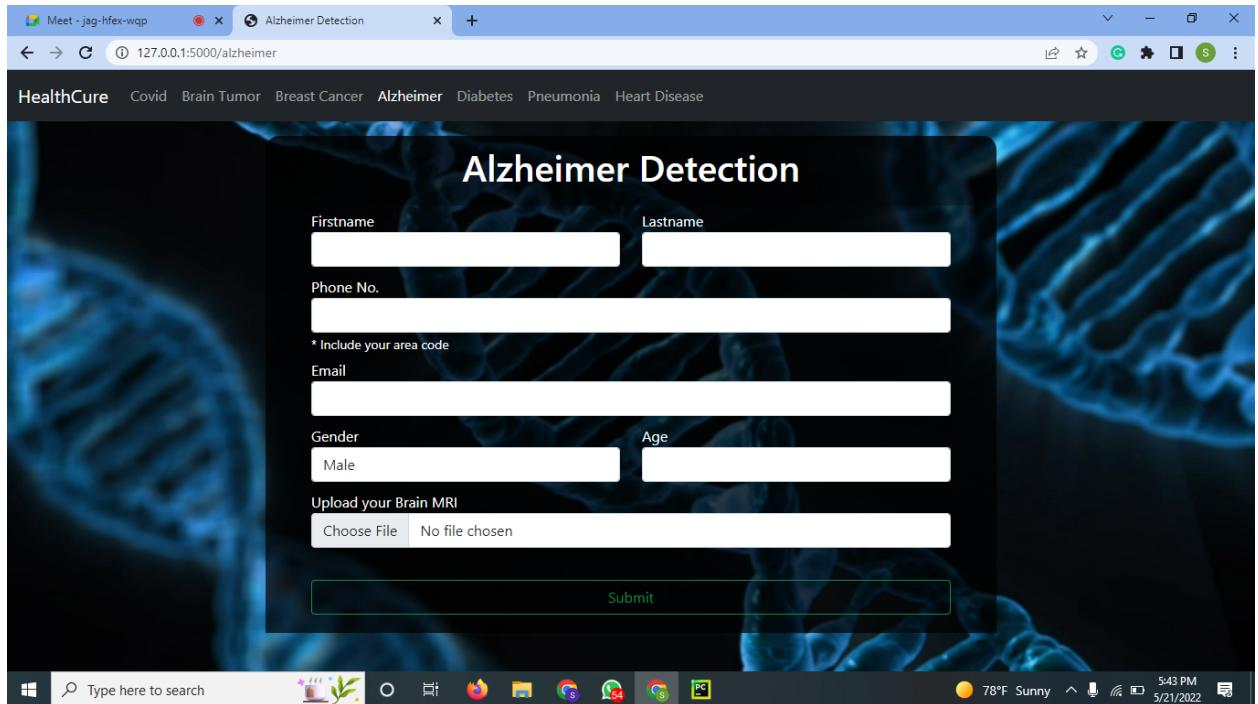
- Firstname
- Lastname
- Phone No. (with a note: \* Include your area code)
- Email
- Gender (Male)
- Age
- Concave Points Mean
- Area Mean
- Radius Mean
- Perimeter Mean
- Concavity Mean

At the bottom of the form is a "Submit" button.

The browser's taskbar at the bottom shows various open applications and the system status bar indicates it's 5:43 PM, 78°F Sunny, and the date is 5/21/2022.

#### 4. Alzheimer Detection

Like Breast Cancer Detection we have used Random Forest algorithm and managed to achieve an accuracy of 73%



## 5. Diabetes Detection

We have used Random Forest again and achieved an accuracy of 66%.

Meet - jag-hfex-wqp

Diabetes Detection

127.0.0.1:5000/diabetes

HealthCure Covid Brain Tumor Breast Cancer Alzheimer Diabetes Pneumonia Heart Disease

Diabetes Detection

Firstname

Lastname

Phone No.

\* Include your area code

Email

Gender  Male

No. of pregnancies

Glucose conc.

Blood Pressure

Skin Thickness

Insulin

BMI

Diabetes Pedigree

Age

Submit

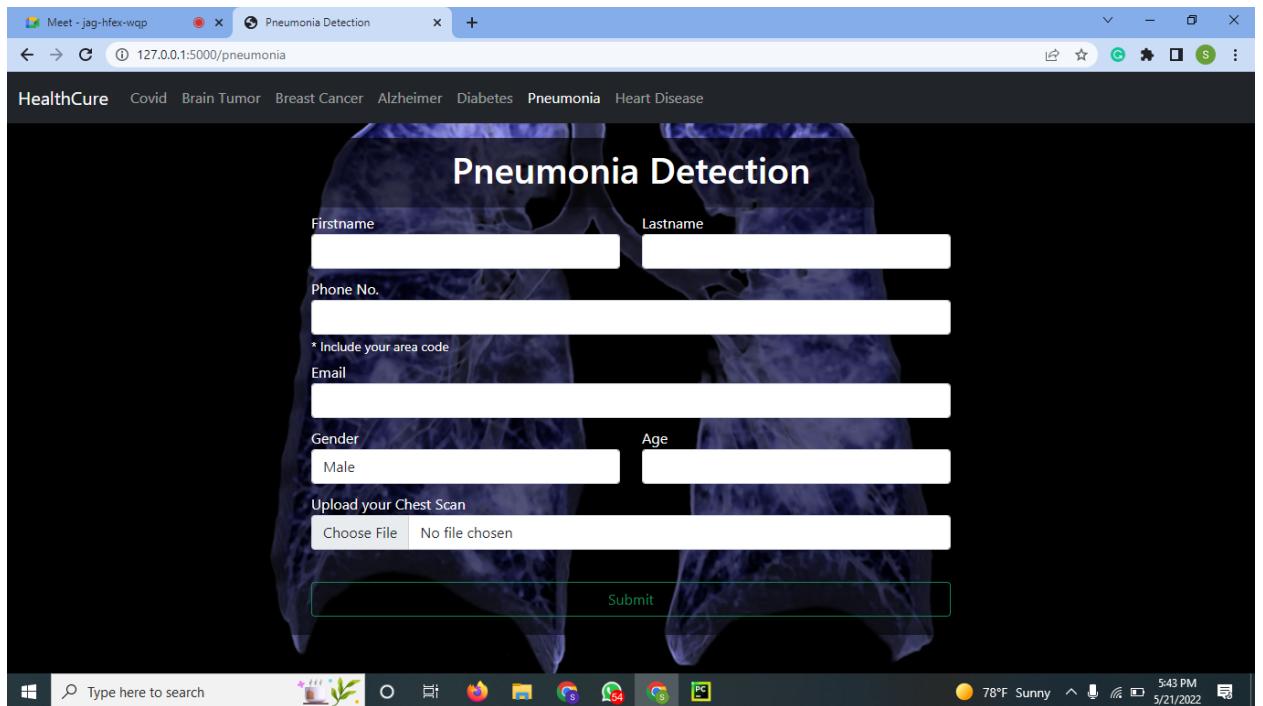
Type here to search

Windows Start button

78°F Sunny 5:43 PM 5/21/2022

## 6. Pneumonia Detection

We have used CNN architecture for this and achieved an accuracy of 83%.



## 7. Heart Disease Detection

For detecting heart disease we have used XGBoost and achieved an accuracy of almost 87%.

HealthCure Covid Brain Tumor Breast Cancer Alzheimer Diabetes Pneumonia Heart Disease

### Heart Disease Detection

Firstname

Lastname

Phone No.

\* Include your area code

Email  Gender  Male

Old Peak  Max. Heart Rate achieved  Exercise induces angina  No

No. of major vessels  Type of Chest Pain  Age  Thal   
0 typical angina 41 1

Submit

Type here to search 78°F Sunny 5:43 PM 5/21/2022

# Output

Meet - jag-hfex-wqp Covid Detection 127.0.0.1:5000/covid

HealthCure Covid Brain Tumor Breast Cancer Alzheimer Diabetes Pneumonia Heart Disease

### Covid-19 Detection

Firstname: Sarjak      Lastname: Patel

Phone No.: 1234567890

\* Include your area code

Email: sarjak.patel@sjsu.edu

Gender: Male      Age: 23

Upload your Chest Scan  
Choose File: 142.jpg

Submit

Windows Taskbar: Type here to search, File Explorer, Edge, Firefox, Google Sheets, Google Slides, Google Sheets, Google Slides, 78°F Sunny, 5/21/2022, 5:42 PM

Meet - jag-hfex-wqp Covid Detection 127.0.0.1:5000/result

HealthCure Covid Brain Tumor Breast Cancer Alzheimer Diabetes Pneumonia Heart Disease

### Covid-19 Test Results

First Name : Sarjak  
Last Name : Patel  
Age : 23  
Gender: male  
Result: COVID POSITIVE



Windows Taskbar: Type here to search, File Explorer, Edge, Firefox, Google Sheets, Google Slides, Google Sheets, Google Slides, 78°F Sunny, 5/21/2022, 5:42 PM

Meet - jag-hfex-wqp Brain Tumor Detection

127.0.0.1:5000/braintumor

HealthCure Covid Brain Tumor Breast Cancer Alzheimer Diabetes Pneumonia Heart Disease

## Brain Tumor Detection

Firstname: Rajvi      Lastname: Shah

Phone No.: 1234567890

\* Include your area code

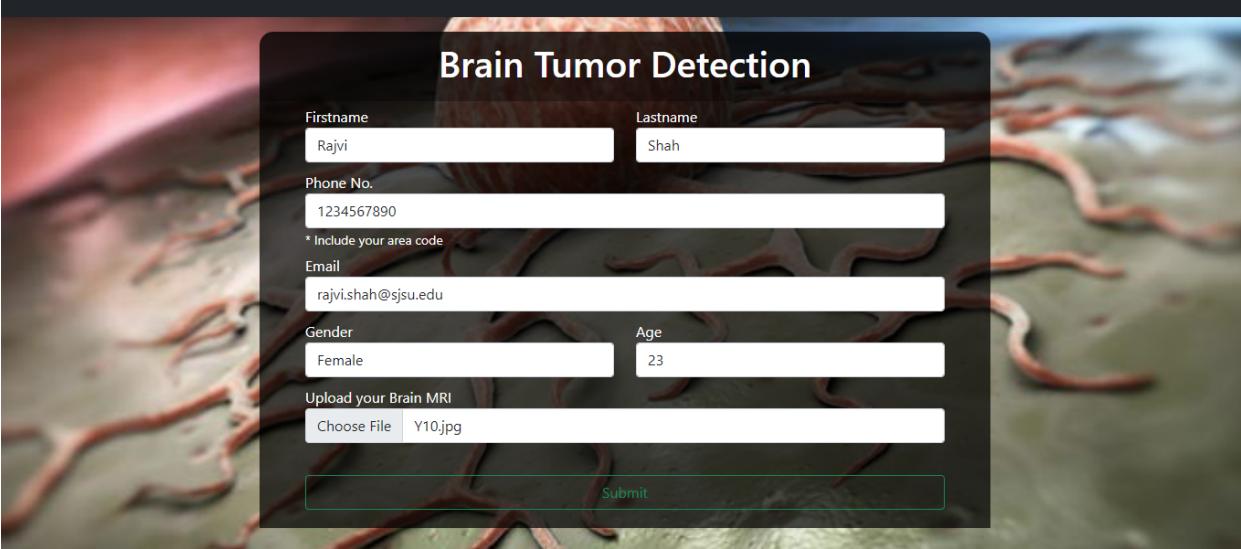
Email: rajvi.shah@sjsu.edu

Gender: Female      Age: 23

Upload your Brain MRI

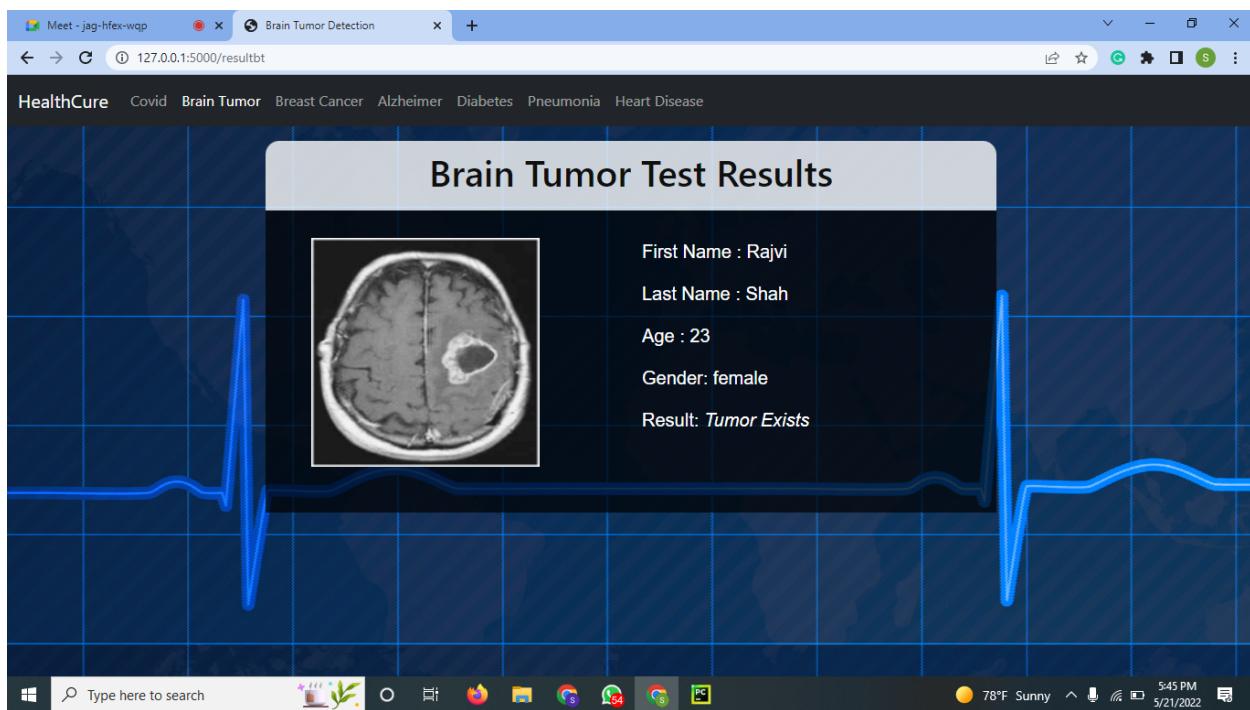
Choose File: Y10.jpg

Submit



Type here to search

5:49 PM 5/21/2022



Meet - jag-hfex-wqp Pneumonia Detection

127.0.0.1:5000/pneumonia

HealthCure Covid Brain Tumor Breast Cancer Alzheimer Diabetes **Pneumonia** Heart Disease

## Pneumonia Detection

Firstname: Yash      Lastname: Kamtekar

Phone No.: 1234567890

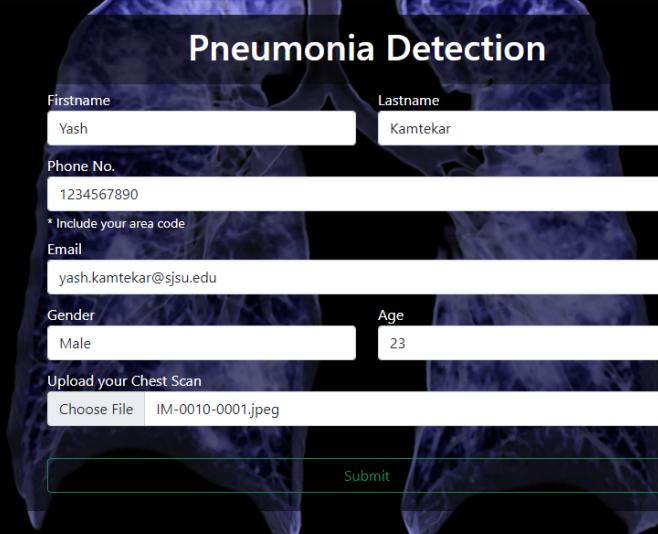
\* Include your area code

Email: yash.kamtekar@sjtu.edu

Gender: Male      Age: 23

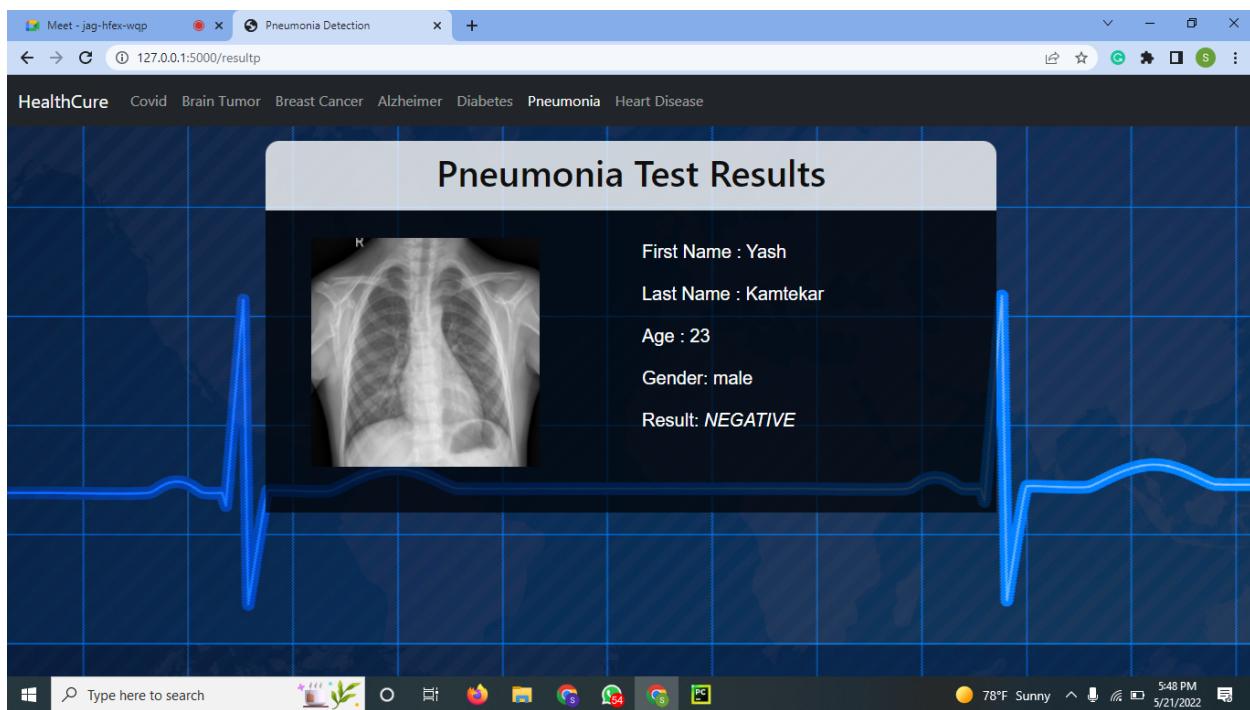
Upload your Chest Scan  
Choose File: IM-0010-0001.jpeg

Submit



Type here to search

78°F Sunny 5:49 PM 5/21/2022



## **Future Scope**

We plan to add more data to our training dataset to get more accurate results.

We also plan to detect disease using XRay images or just by inputting numbers.

We also plan to add features such as giving precautions to the user for the disease that the user has been detected to.

Another feature can be giving some remedies to the user of the disease that the user has been detected.

## References

<https://machinelearningprojects.net/healthcare-medical-project/>

<https://www.techtarget.com/searchenterpriseai/definition/convolutional-neural-network>

<https://towardsdatascience.com/step-by-step-vgg16-implementation-in-keras-for-beginners-a833c686ae6c>

<https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>

<https://www.nvidia.com/en-us/glossary/data-science/xgboost/>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8754556/#:~:text=The%20AI%20techniques%20are%20also,results%2C%20decrease%20costs%2C%20etc.>