iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52536

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Integrated Healthcare System Using Machine Learning

Prof. Archana Ugale[1], Abhijeet Gadakh[2], Roshan Sawant[3], Aniket Malunjkar[4], Vaibhav Dhakane[5]
[1, 2,3,4,5] *Department of Information Technology, Sir Visvesvaraya Institute of Technology, Nashik*

Abstract: ***In this digital world, data is an asset, and enormous data was generated in all fields. Data in the healthcare industry consists of all the information related to patients. Here a general architecture has been proposed for predicting disease in the healthcare industry.***
Many of the existing models are concentrating on one disease per analysis. There is no common system present that can analyse more than one disease at a time.
Thus, we are concentrating on providing immediate and accurate disease predictions to the users about the symptoms they enter along with the disease predicted. So, we are proposing an integrated system which used to predict multiple diseases by using Deep learning and machine learning algorithms. In this system, we are going to analyse Covid, Alzheimer's, and Pneumonia, Diabetes, Heart disease and Brain Tumor disease.
Later many more diseases can be included. To implement integrated healthcare systems, we are going to use machine learning and deep learning algorithms such as CNN and Random Forest. The importance of this system analysis is that while analysing the diseases all the parameters which cause the disease is included so it is possible to detect the disease efficiently and more accurately. In this prediction system, a user can analyse more than one disease on a single web application and if it is found that the result is positive then proper care guidance, diagnosis, and prescription will also give to the patient. Also, the printout of the generated result is given to the user.
In an integrated healthcare system, the user must select the name of the disease, enter its parameters and just click on predict. The corresponding machine learning model will be invoked and it would predict the output and display it on the screen. Also, all the information will be stored securely in the database for generating reports and information for future use in understanding previous disease patterns.
Keywords: Machine Learning, Deep Learning, CNN, Flask, Disease detection, XGBoost, Random Forest

I. INTRODUCTION

Machine learning and deep learning are subfields of Artificial Intelligence that are playing a huge role in today's time. From chatbots, object detection, recommendation system, and self-driving cars to medical fields, we are finding them everywhere. The medical field generates a huge amount of patient data which can be prepared in a lot of ways. So, with the help of the combination of machine learning, and deep learning, we have created an integrated disease prediction system that can detect six diseases at a time. Many of the existing systems can predict only one disease at a time and that too with lower accuracy and lower speed. Lower accuracy can seriously put a patient's health at risk. We have considered six diseases for now which are heart disease, brain Tumor, Diabetes, Covid, Pneumonia, and Alzheimer, and in the future, many more diseases can be added. The user has to enter input parameters of the disease or upload MRI, chest scan, CT scan, or X-ray image in the input field. The system would display the output whether he/she has the disease or not. This project can help a lot of people as one can monitor the patient's condition and take the necessary precautions thus increasing life expectancy. In this project, we are proposing a system used to predict multiple diseases by using Flask API and Python. We are trying to develop a system for patients and doctors so that we have to reduce the gap between them.

This system gives early detection and saves lots of lives by reducing the death rate of chronic diseases. Also, if a patient is found positive for that disease, we are trying to give them the information which is necessary for the diagnosis of that disease and a list of all other necessary factors about that disease. All the precautions related to that disease will give patients proper directions to cure that disease.

Also, all the data of that patients will be saved in the database so that in the future doctor have access to this information for better treatment.

©IJRASET: All Rights are Reserved

II. LITERATURE SURVEY

Paper Title	Authors	Publications and Details	Methodology	Advantages	Disadvantages
Multi Disease Prediction Using Data Mining Techniques	K. Gomathi, Dr. D. Shanmuga Priyaa	IJSSE Volume: 04, Issue: 2, Dec-2016	Data mining techniques can be used for predicting different types of diseases.	No. of tests can be reduced, Robust, Efficient to use	Limited to only three diseases, More time required
Prediction of Heart Disease using Machine Learning Algorithms	Santhana Krishnan J, Geeta S.	IEEE Xplore Apr-2019	Random Forest machine learning algorithm was used to predict heart disease	Fast and Robust, Higher accuracy	Specific to only one disease,
Multiple Disease Prediction System	Ankush S, Ashish Y, Saloni S, Prof. Renuka N.	IRJET Vol 09 No 03 Mar-2022	Three diseases were predicted by using 3 ML algorithms	User-friendly UI, Fast and efficient prediction	Internet connectivity is required

III. EXISTING SYSTEM

In the existing system, the dataset used for training and testing of the model is very small, for patients and diseases with specific disorders. These systems are mostly built for the prediction of only one disease at a time. These systems also give lower accuracy due to inadequate datasets. The pre-selected attributes may sometimes not satisfy the changes in the disease and its affecting factors which could lead to inaccuracy in results. Also, these existing systems only give focus on predictions of different diseases with various algorithms but they are not focusing treatment and any other tips for fast recovery from that disease. Also, the printing facility of generated disease results was not applied in earlier systems.

IV. PROPOSED SYSTEM

We are proposing a system that will be simple to handle and also provide a user-friendly user interface. It should be time efficient. To make it accurate and less time-consuming to do predictions we are providing adequate input entries to fulfil the values of all attributes. This system will work to reduce the gap between doctors and patients. For now, we are taking six diseases to predict in this system. Brain Tumor, Heart disease, Diabetes, Covid, Alzheimer, and Pneumonia. We are developing a deep learning model by using CNN and VGG16 to do better predictions of these four diseases Alzheimer's disease, Brain Tumor, Covid, and Pneumonia. Also, for the remaining disease, we are using Random Forest and XGBoost algorithms. The data-pre-processing and feature extractions are two very important steps for building an accurate model. In this system, if patients are found with a disease, then the system will automatically give health tips and all other necessary tips about that disease for better treatment and fast recovery from that disease. We are using the Flask framework for making a user-friendly user interface and integrating the front end and back end of this project.

V. SYSTEM ARCHITECTURE

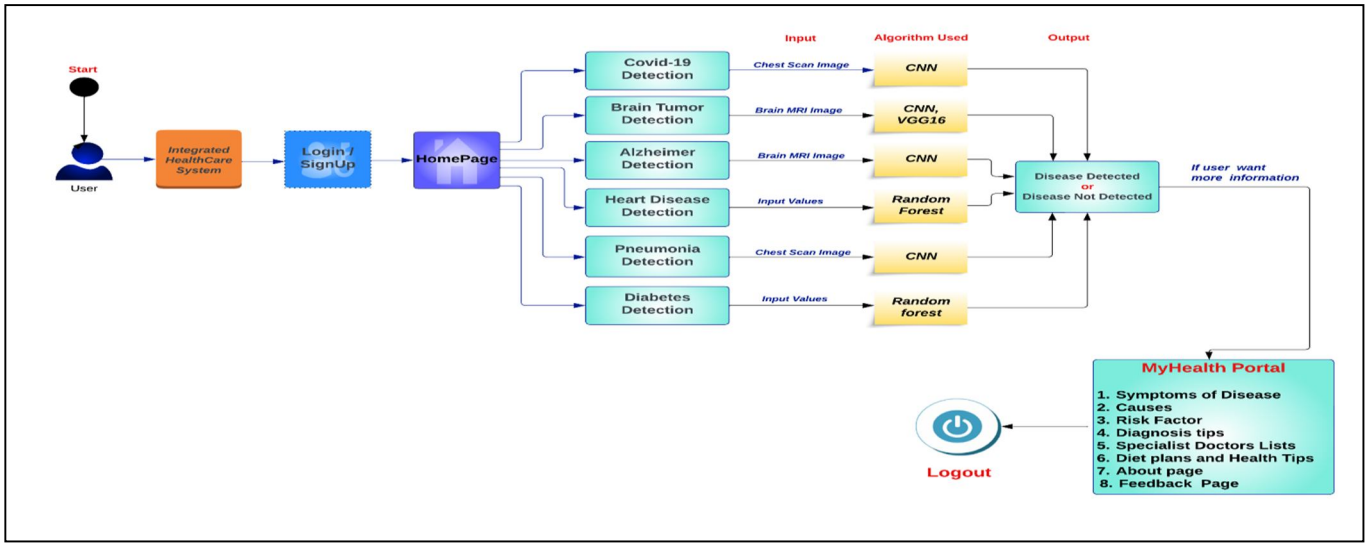


Fig. 1. System Architecture

In this system architecture, we are showing six diseases. For the prediction of each disease, the proposed model is different. The data-pre-processing is done on each dataset differently to remove all redundancies, missing values, and noise from the data. The input passed to each disease is different. After cleaning the data is fit for the training of different algorithmic models on it. After data is converted into training data (80%) and testing data (20%) the model is fitted to a particular algorithm. The model is built on training data and evaluated on testing data. When the model achieves good accuracy then the model is saved in a pickle file using the pickle module. Then using the Flask web application framework, we build a responsive user interface to handle user and system communication. Also, we have prepared another website where we are giving extra information about that disease that points are basic information, causes and symptoms, precaution and diet, risk factors, diagnosis and tests, re-infections, treatment, prevention, and frequently asked questions.

A. Some UML Diagrams

1) Use Case Diagram: Use case diagrams are usually referred to as behaviour diagrams used to describe a set of actions (use cases) that some system or systems (subject) should or can perform in collaboration with one or more external users of the system (actors). Each use case should provide some observable and valuable result to the actors or other stakeholders of the system.

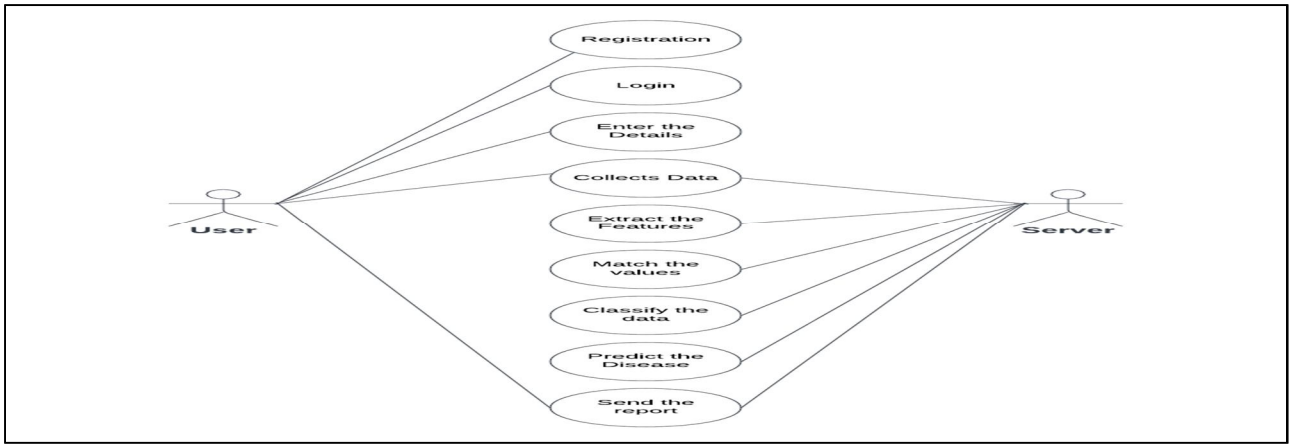


Fig. 2. Use Case Diagram

2) Sequence Diagram: A sequence diagram shows object interactions arranged in a time sequence. It depicts the objects and classes involved in the scenario and the sequence of messages exchanged between the objects needed to carry out the functionality of the scenario. Sequence diagrams are typically associated with use case realizations in the Logical View of the system under development. Sequence diagrams are sometimes called event diagrams or event scenarios.

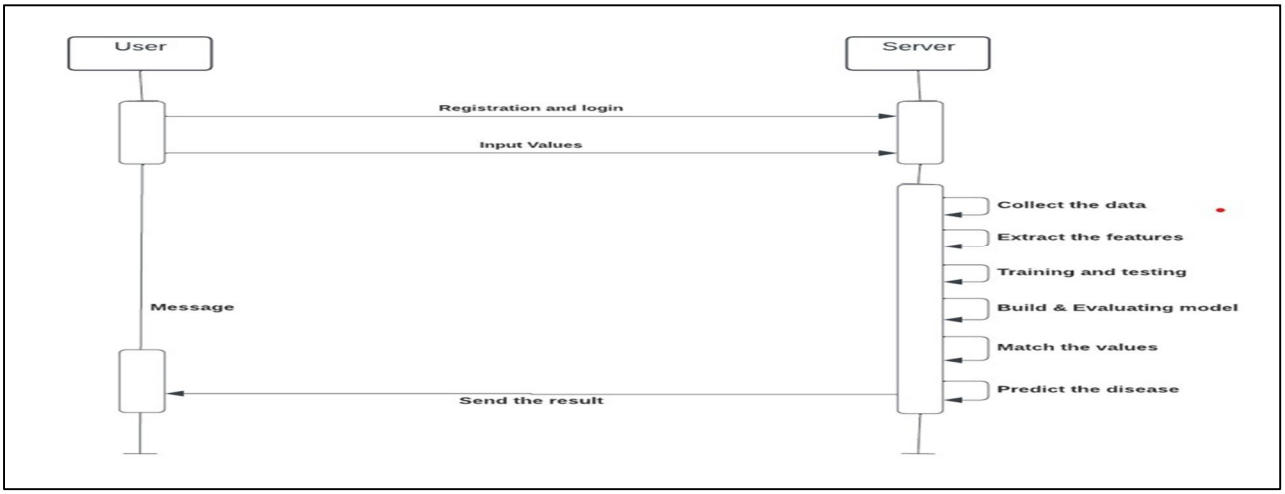


Fig. 3. Sequence Diagram

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

VI. SYSTEM REQUIREMENTS

A. Hardware Requirements

1) Works on: Laptops, Computers, and mobile
2) RAM: Minimum 2GB and above
3) CPU: Dual Core or Quad Core Processor

B. Software Requirements

1) Browsers supported: Google Chrome, Brave, Microsoft Edge
2) Programming languages used: HTML, CSS, JavaScript, Python, Bootstrap
3) The framework used: Flask
4) The library used: Pandas, Numpy, Maptplotlib, Sklearn, Tensorflow, Keras,
5) Platform: VS Code, Jupyter Notebook, Google Colab

VII. PROJECT IMPLEMENTATION

A. Algorithm Used

1) Convolutional Neural Network

So, when there is an image dataset present and we have to do the image classification or image feature extraction at that moment we apply the CNN algorithm to that data to get the best insights and results from that data. In this project, we applied CNN Algorithm for Covid Detection and Brain Tumor Detection. Alzheimer's Detection and Pneumonia Detection.

2) What is Convolutional Neural Network?

CNN stands for Convolutional Neural Network, which is a deep learning algorithm that is primarily used for image and video processing. CNNs consist of multiple layers, including convolutional layers, pooling layers, and fully connected layers, that work together to extract features from input data. Convolutional layers use filters to scan through the input data and extract features by convolving the filters with the input data. Pooling layers reduce the dimensionality of the feature maps by down-sampling them to preserve the most important information while discarding redundant information. Fully connected layers are used to classify the features extracted by the convolutional and pooling layers into different categories. CNN (Convolutional Neural Networks) is used in image classification because they are designed to automatically learn and extract relevant features from images, capturing spatial hierarchies. This enables CNNs to effectively handle the complex patterns and structures present in images, leading to high accuracy in image classification tasks.

3) Steps for implementing CNN Algorithm:

a) Collect and pre-process the data:
b) Build the CNN architecture:

- Define the convolutional layer, pooling layer and fully connected layers
- Define the required activation function, number of filters, kernel size

c) Compile the model:

- Compile the model by specifying the optimizer, loss function and metrics to be used.

d) Train the model:

- Set numbers of epochs,
- Monitor the performance by plotting the graphs

e) Evaluate the model:

- Evaluate the performance by accuracy, precision, recall and F1 score.

f) Optimize and deploy the model:

4) Random Forest Algorithm:

Random forest is a machine learning algorithm that builds a collection of decision trees and combines their outputs to make predictions. Each decision tree is trained on a random subset of the data and a random subset of the features, to reduce overfitting and improve performance.

©IJRASET: All Rights are Reserved

The final prediction is made by aggregating the predictions of all the decision trees, using a majority vote (in classification problems) or average (in regression problems). Random forest is an example of an ensemble learning algorithm, which combines multiple models to improve prediction accuracy. The random forest algorithm is easy to use and can handle a variety of input data types, including categorical and continuous variables. It is also robust to outliers and missing data, making it a popular choice for real-world applications. The random forest can be used for both classification and regression problems, making it a versatile algorithm.

5) XGBoost Algorithm:

XGBoost is a machine-learning algorithm that combines the predictions of multiple decision trees to make accurate predictions. It starts with a simple decision tree and iteratively improves by building additional trees that correct the mistakes of the previous ones. Each tree focuses on the examples that were predicted incorrectly and assign them higher weights to correct the errors. The algorithm continues this process, gradually reducing the overall error and creating a strong predictive model. It uses boosting techniques to emphasize the examples that are more challenging to predict. Finally, it combines the predictions of all the trees to make the final prediction. XGBoost incorporates regularization techniques to prevent overfitting, which occurs when a model becomes too complex and performs well on the training data but fails to generalize to new data. Regularization helps maintain simplicity and generalization by adding penalties to the objective function during model training.

XGBoost supports parallel processing, which means it can utilize multiple CPU cores during training, making it faster and more efficient. This parallelization allows for faster model building, especially when dealing with large datasets.

VIII. GRAPHICAL USER INTERFACE

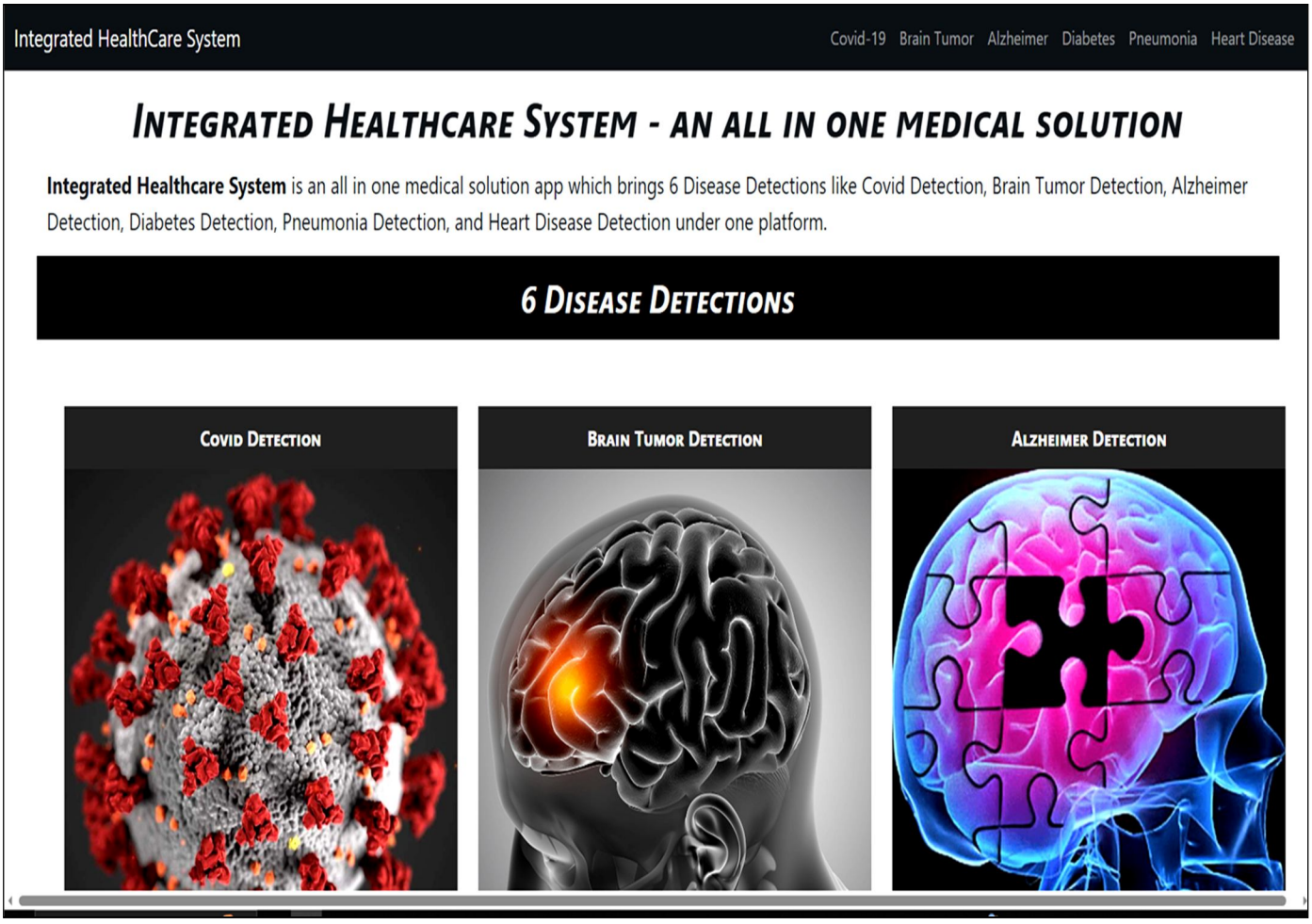


Fig. 4. Graphical User Interface

VIII. PROJECT IMPLEMENTATION SCREENSHOT

A. *Covid Detection*

Input Output

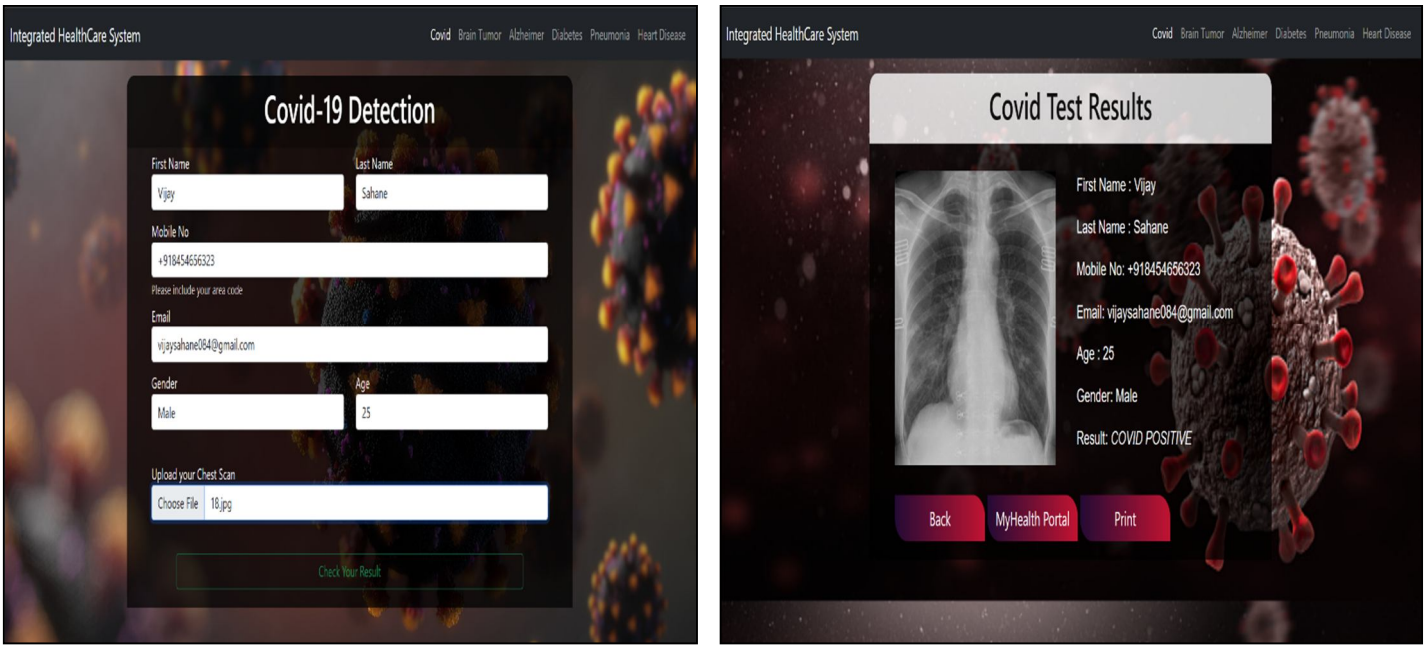


Fig. 5. Covid Detection

B. *Brain Tumor Detection*

Input Output

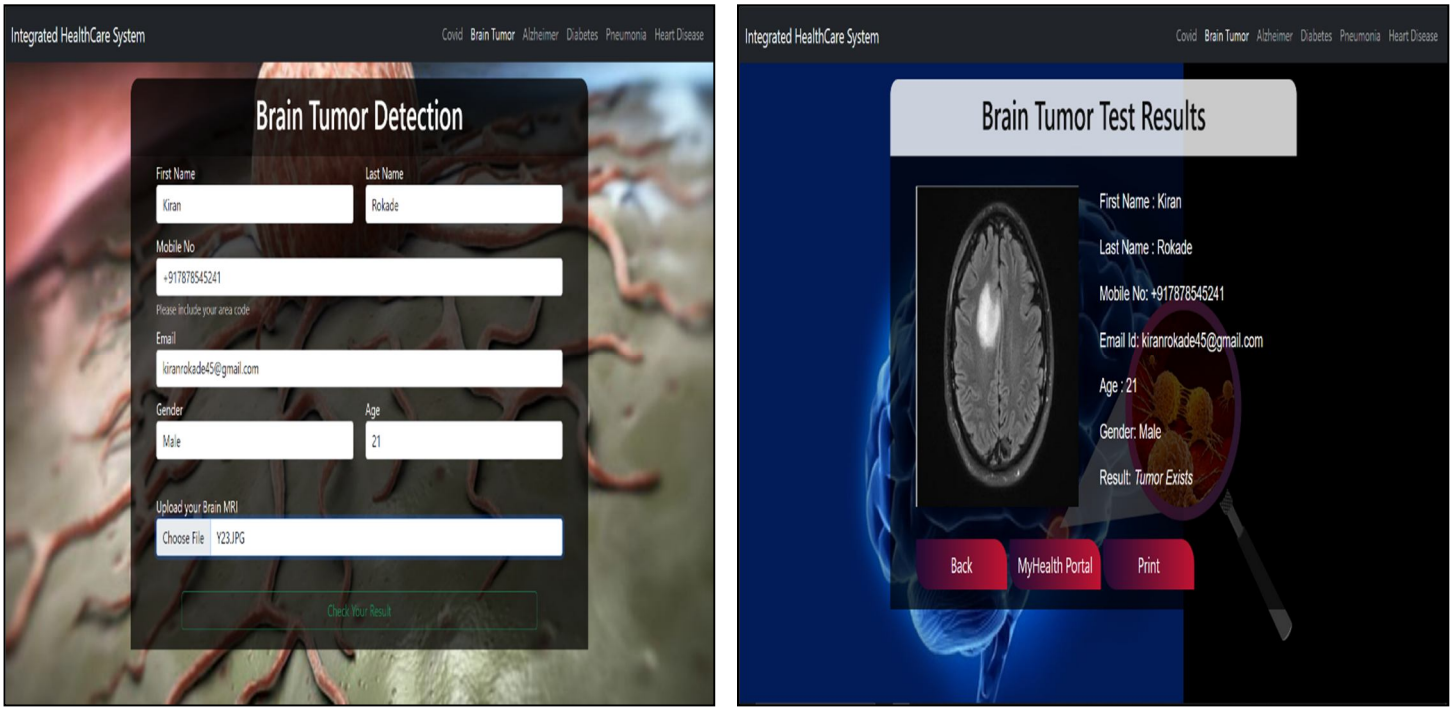
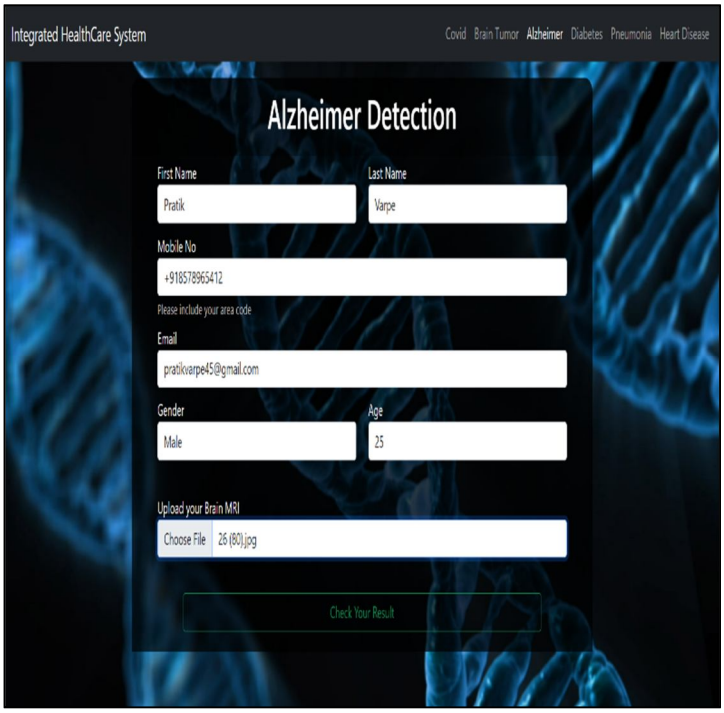


Fig. 6. Brain Tumor Detection

C. Alzheimer's Disease Detection

Input



Output

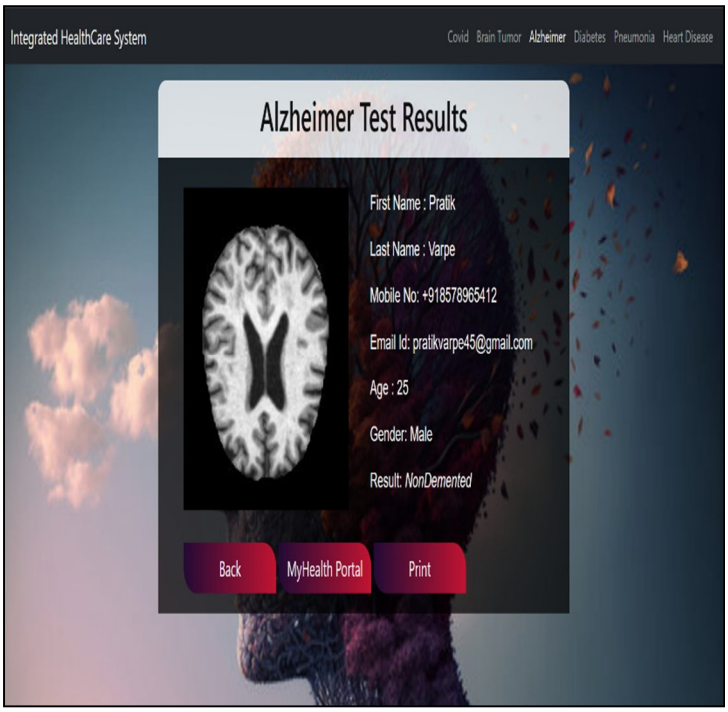
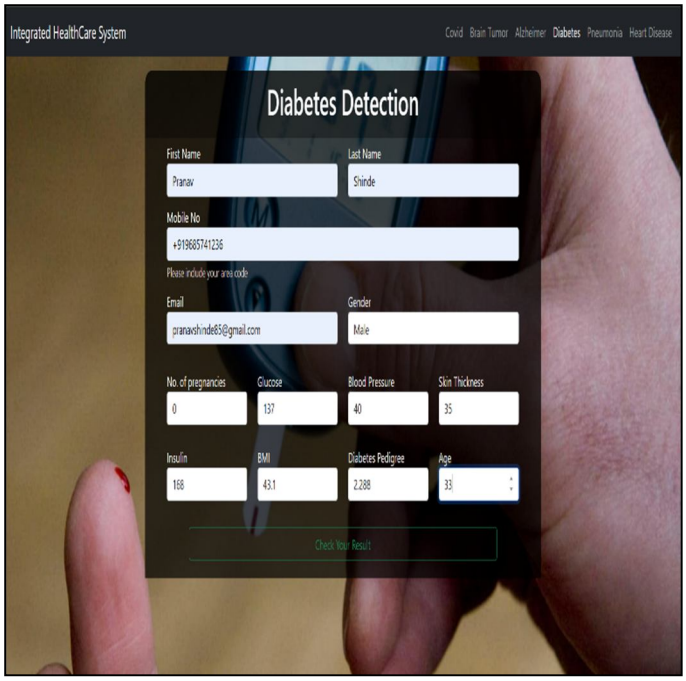


Fig. 7. Alzheimer's Detection

D. Diabetes Detection

Input



Output

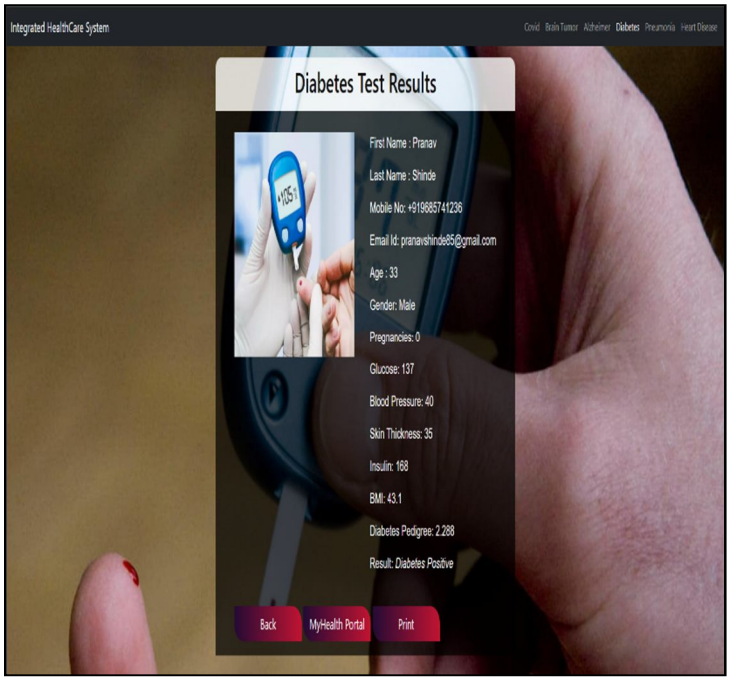
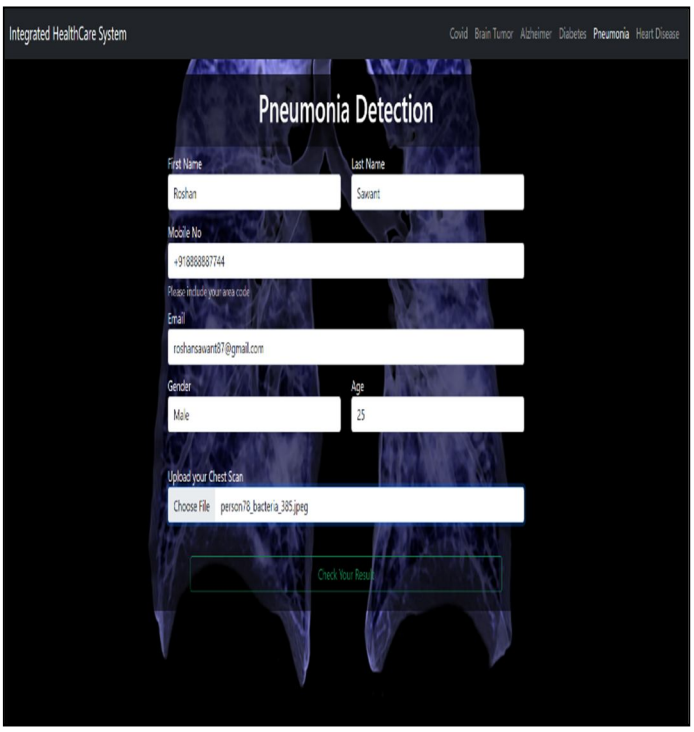


Fig. 8. Diabetes Detection

E. Pneumonia Detection

Input



Output

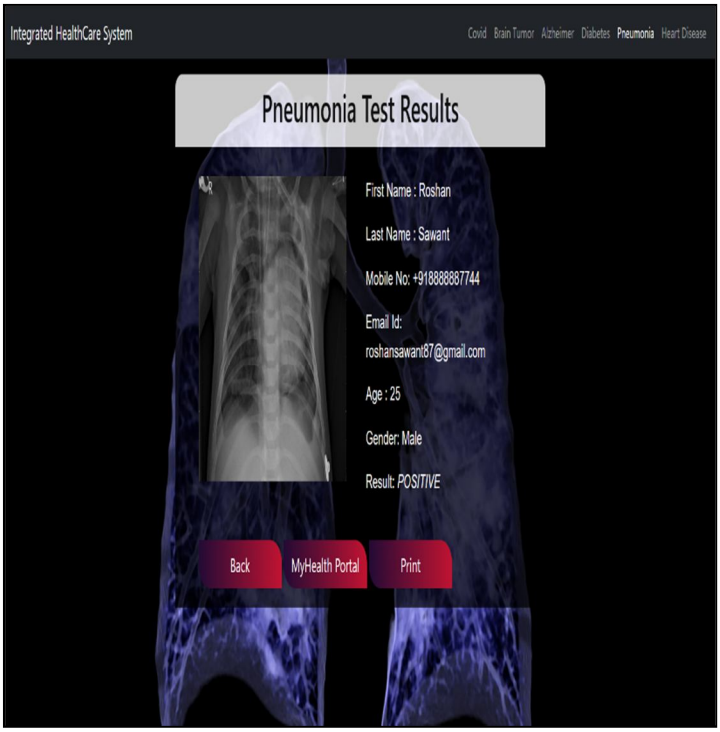
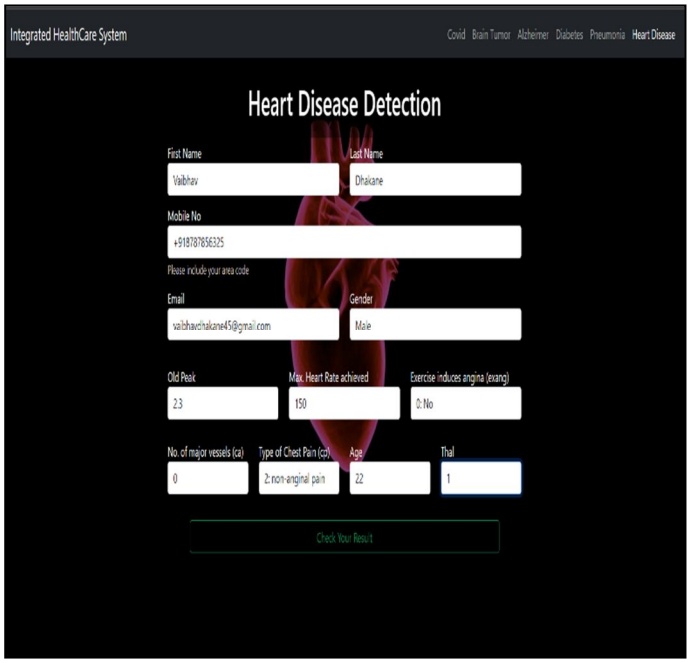


FIG. 9. Pneumonia Detection

F. Heart Disease Detection

Input



Output

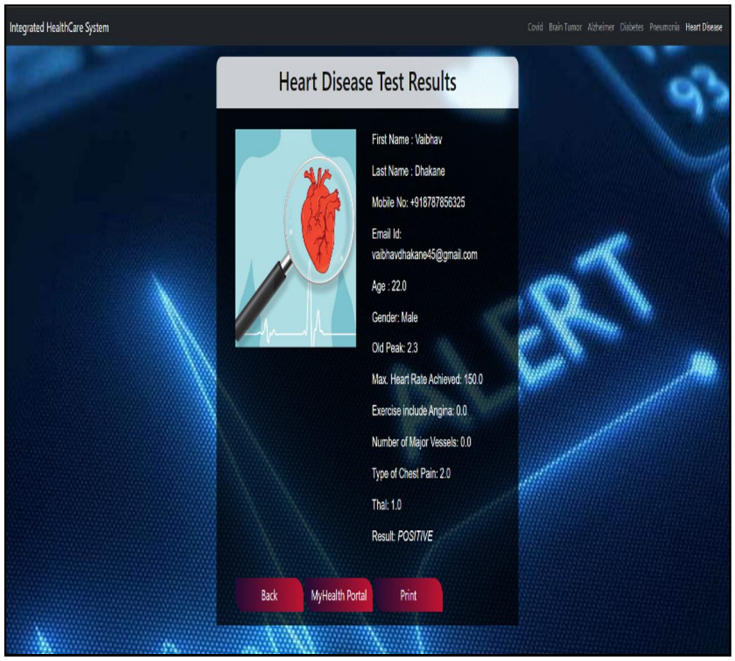


FIG. 10. Heart Disease Detection

G. Healthcare Portal Screenshot

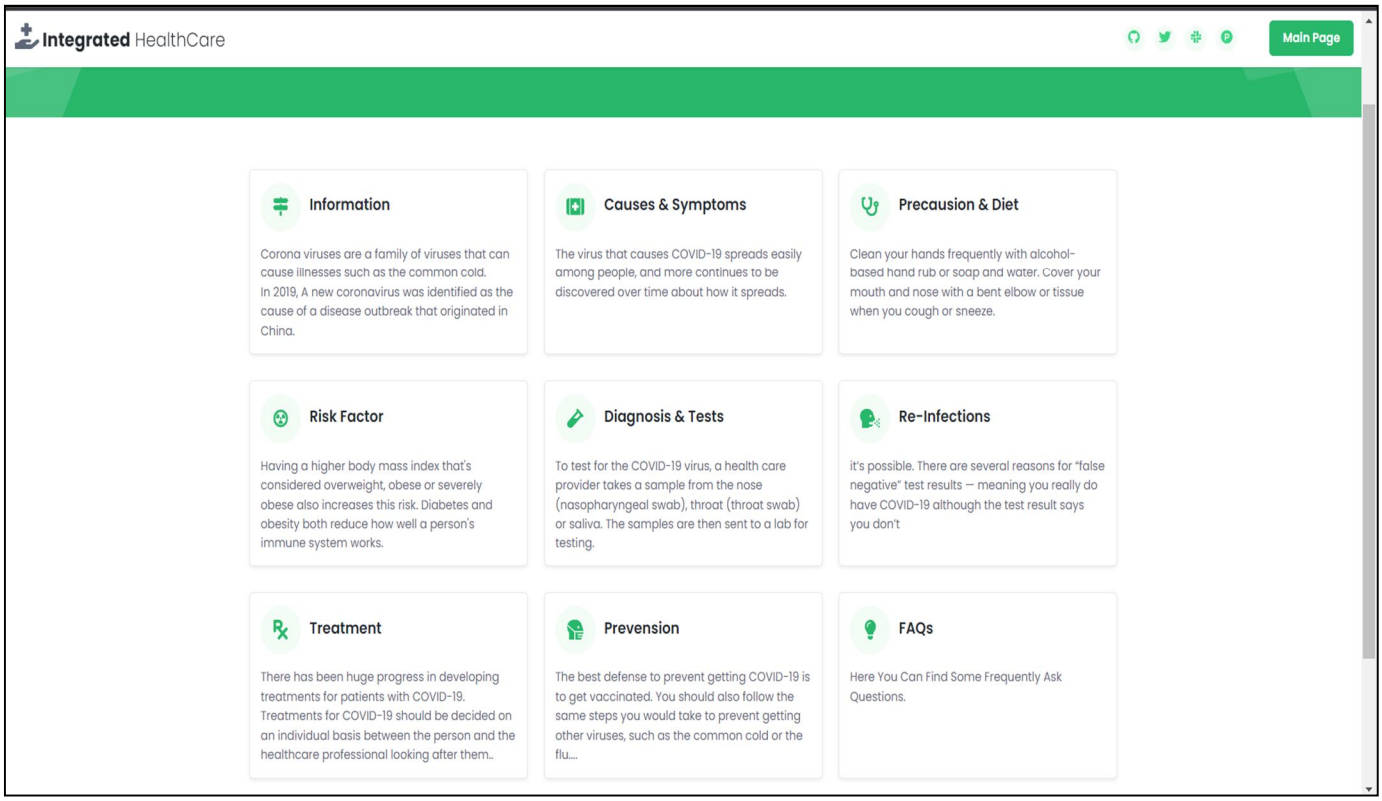


Fig. 11. Healthcare Portal

IX. RESULT AND ANALYSIS

A. Result

Sr. No	Disease Name	Algorithm Used	Input Data type and Size	Accuracy of Model
1	Covid Detection	CNN	Input: Image Dataset Size: (384,2)	97.20 %
2	Brain Tumor Detection	CNN VGG16	Input: Image Dataset Size: (253,2)	83.42 %
3	Alzheimer's Detection	CNN	Input: Image Dataset Size: (5122,4)	97.24 %
4	Diabetes Detection	Random Forest	Input: Values Dataset Size: (768,9)	77.27 %
5	Pneumonia Detection	CNN	Input: Image Dataset Size: (5856,2)	94.33 %
6	Heart Disease Detection	Random Forest	Input: Values Dataset Size: (303,14)	84.78 %

B. Analysis

1) The classification models showed high accuracy in identifying Covid, Brain Tumor, Alzheimer's, Pneumonia, Diabetes, and heart disease.

2) Random Forest algorithm performed well in predicting Diabetes and heart disease. Convolutional Neural Networks performed well in predicting all other diseases.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

3) The GUI developed using HTML, CSS, and Bootstrap, integrated with Flask, makes it easy for users to interact with the disease prediction models.

4) The CNN models used for Covid, Brain Tumor, Alzheimer's, and Pneumonia, were trained on image datasets, and have the potential to provide accurate diagnoses using medical imaging.

5) The accuracy of the classification models can be further improved by increasing the size and diversity of the training datasets, optimizing hyperparameters, and using other advanced machine learning algorithms.

6) Overall, the developed disease classification models and the user-friendly GUI can be useful tools for healthcare professionals to diagnose and treat patients with the selected diseases with high accuracy and efficiency.

X. CONCLUSION

This paper gives the research of multiple experiments done in the healthcare field. This healthcare system supports many diseases prediction in a single web app using different machine learning and deep learning algorithms. The main purpose of this project is to build a system that would predict five to six diseases at one time with high accuracy. Also, the system must be highly secure and have an easy and responsive user interface. Because of this integrated system, the patient doesn't need to traverse different websites which saves time. The project successfully achieved its objectives by utilizing machine learning and deep learning techniques to develop a highly efficient medical system with a user-friendly interface. The second phase of this project further improved the system by incorporating real-time data and integrating front-end and back-end using Flask. The models accurately predicted disease and provided effective diagnosis tips. Overall, the project can be considered a success in providing a reliable and user-friendly medical system. Hence, we understand the drawbacks of existing systems that we have reduced in this system and give highly efficient solutions for the medical system with a user-friendly user interface and proper treatment and effective diagnosis tips after a disease is predicted successfully.

XI. ACKNOWLEDGMENT

We would like to express our sincere gratitude to our project guide Prof. Archana Ugale for providing help during the research, which would have seemed difficult without their motivation, constant support, and valuable suggestions. We also thank our college and our respected teachers for giving us the platform to prepare a project on the topic "Integrated Healthcare System using Machine Learning". The research carried out for the project helped us learn a lot of things and also gave us practical thinking with the technologies that were used in this project.

REFERENCES

[1] Akkem Yaganteeswarudu "Multi Disease Prediction Model by Using Machine Learning and Flask API". In 2020 5th International Conference on Communication and Electronics Systems (ICCES), pages 1242-1246. IEEE, 2020.

[2] P Hamsagayathri and S Vigneshwaram. Symptoms-based disease prediction using machine learning techniques. In 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), pages 747-752. IEEE, 2021.

[3] Priyanka Sonar, Prof. K. Jaya Malini, "Diabetes Prediction Using Different Machine Learning Approaches", 2019 IEEE,3rd International Conference on Computing Methodologies and Communication (ICCMC)

[4] Archana Singh, Rakesh Kumar, "Heart Disease Prediction Using Machine Learning Algorithms", 2020 IEEE, International Conference on Electrical and Electronics Engineering (ICE3)

[5] Varun Kumar Ojha, Vandana Bhattacherjee, and Sanjeev Sharma, "A Comparative Study of Machine Learning Algorithms for Disease Prediction," International Journal of Advanced Computer Science and Applications, vol. 11, no. 6, 2020.

[6] Anjali Sharma, Kamal Kumar, and Neha Sharma, "Comparison of Different Machine Learning Algorithms for Heart Disease Prediction," International Journal of Computer Sciences and Engineering, vol. 7, no. 10, 2019.

[7] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different Supervised machine learning algorithms for disease prediction," BMC Medical Informatics and Decision Making, vol. 19, no. 1, pp. 1–16, 2019.

[8] www.kaggle.com

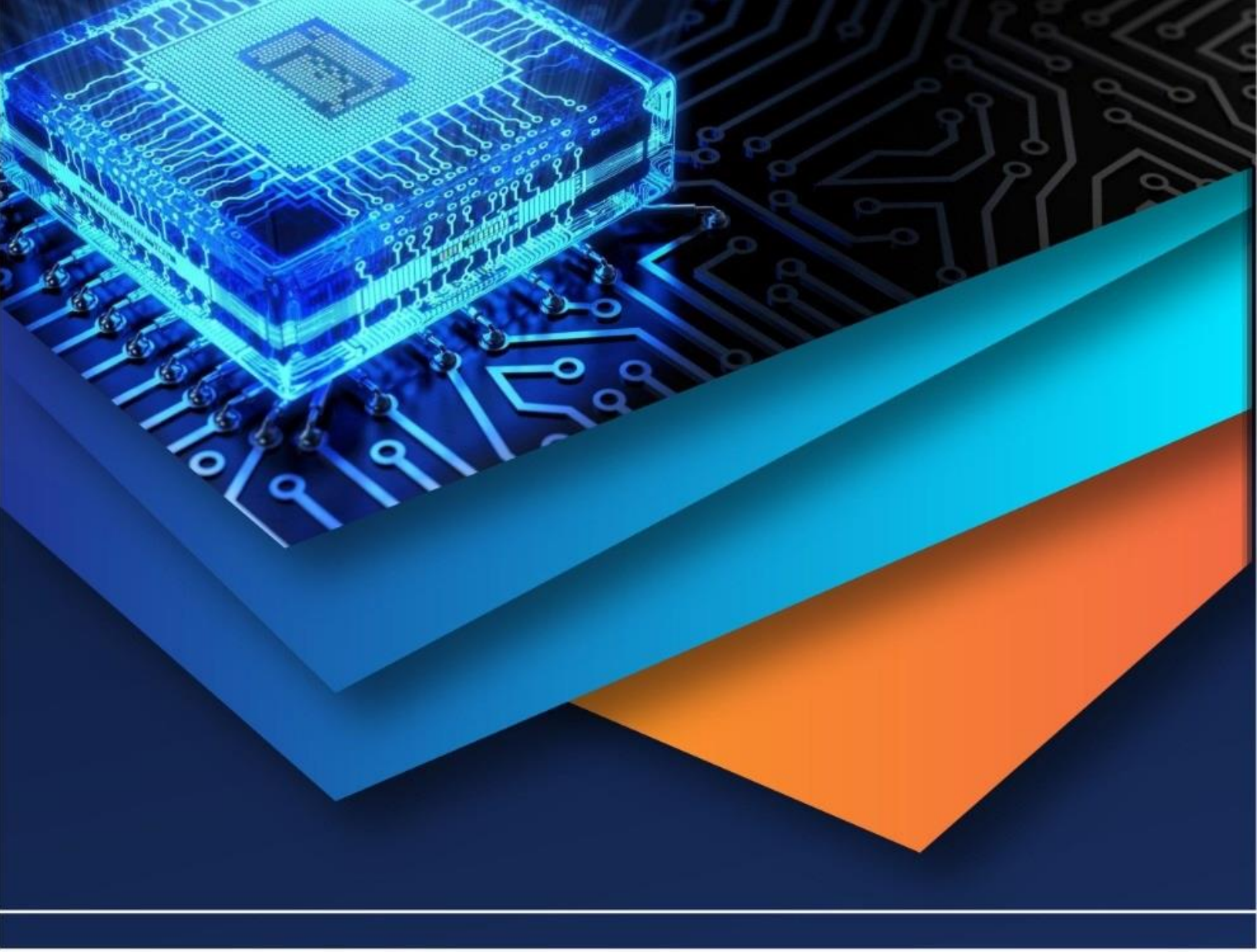
[9] https://archive.ics.uci.edu/ml/index.php

[10] https://healthcare.ai

[11] www.youtube.com

[12] Stanford Machine Learning Group (stanfordmlgroup.github.io)

[13] https://machinelearningmastery.com/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)