iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** II **Month of publication:** February 2023

DOI: https://doi.org/10.22214/ijraset.2023.49166

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 II Feb 2023- Available at www.ijraset.com

Streamlit Interface for Multiple Disease Diagnosis

MS P Keerthi[1], G Srikar Reddy[2], V Sai Raghava[3], K Buchi Reddy[4]
[1, 2, 3, 4]*Department of CSE, Vardhaman College of Engineering,* Hyderabad, India

Abstract: ***Nowadays most people are suffering with the many chronic diseases and getting effected by those without having the knowledge. In this paper we are trying to establish a single streamlet interface for three different machine learning disease prediction models. Our interface includes three diseases which are heart disease, diabetes, and pneumonia. Our heart disease model is developed using the logistic regression which explores whether the person have healthy heart or not by taking the needed inputs which are explored in further sections. Our diabetes model predicts whether the person has diabetes or not and it is developed through SVM classifier. The pneumonia model is developed through convolution neural network (CNN) which enhances whether the person has pneumonia or not by taking chest x-ray image as the input. All these models are embedded in the streamlit interface which is the python framework for making website interface. User can interact with these models through this interface by giving needed inputs and predicting their disease and enacting to them early before they get worst.***
Keywords: Pneumonia, diabetes, SVM, CNN.

I. INTRODUCTION

Machine learning has become the most developing technology in the current scenario where it is giving the good outcomes and accuracy. Building disease prediction models through these techniques will be very useful for many aspects. In this paper we build models for diabetes, heart disease and pneumonia. Diabetes mellitus (D.M.) is one of the most extensive diseases the world has confronted, particularly in developed or developing countries. The D.M. is a metabolic disease that causes high blood sugar. The hormone insulin transfers sugar from the blood into cells for energy utilization. Diagnosing Diabetes as early as possible is the need of an hour because it can lead to various complications. When it comes to heart disease, as per different health organization reports most of the people nowadays are suffering with unhealthy heart and in many cases it resulting to attacks and death. So, predicting our heart is healthy or not is the most important thing in these days. The pneumonia cases have been increased due the COVID effect. It is the highly ranked cause of death and it also many varieties. So predicting it early and getting treatment may decrease the chance of the huge complications and we may save our life. So we have proposed a method where we used a python framework called streamlit and made interface for these three diseases . After developing those machine learning models and making them together at this interface and making user easy to use these models for predicting their disease at single place.
In this paper, we discuss related work, methodology, implementation and results of our proposed method.

II. RELATED WORK

This section briefly reviews past work on different disease detection models we worked in this paper.
Firstly, when it comes diabetes disease prediction model, Wilson et al. [1] developed a D.M. prediction model that predicts middle-aged American adults' risk. Logistic Regression is being used for performing this task. The factors considered to predict the risk were parental history of D.M., obesity, high blood pressure, low levels of high-density lipoprotein cholesterol, elevated triglyceride levels, and impaired fasting glucose. About 3140 Samples were taken, and the results showed that the developed model gave an accuracy of 85%. The evaluation of the proposed algorithm was performed by [2] on the Canadian population in which about 4403 samples were taken, and after performing experiments, the algorithm showed 78.6% accuracy. [4] used various machine learning algorithms to predict prediabetes. Some of the algorithms used in [4] work are Decision Tree, Logistic Regression and Artificial Neural Network. In their work, there were about 735 patients who were diagnosed as Diabetic and 752 as non-diabetic. The results showed that the decision tree outperformed than other two algorithms by showing an accuracy of 77.3%. In this work, we are using a Support Vector Machine for performing binary classification.
When it comes to heart disease model, some of the disease models are developed and the some of the enhancements are taken from the references [13][14][15].
when it comes to Pneumonia detection, [6] deployed VGG-16, VGG-19, Resnet50, and InceptionV3 network architectures and transfer learning for the detection of viral pneumonia, obtaining 71–88% accuracy on 624 test images. Dey et al. [7] combined the VGG-19 architecture with various classifiers like SVM or random forest and based on a set of 1650 X-ray images they achieved accuracy, precision, and F1 score values ranging up to 98%, 95%, and 96%, respectively.

©IJRASET: All Rights are Reserved

Brunese et al. [8] established a three-class pneumonia detection framework based on VGG-16 architecture and the so-called GradCAM algorithm for visual debugging and achieved a 96.2% accuracy on a set of 6523 CXR images. Panwar et al. [9] combined the VGG-19 architecture with the GradCAM algorithm and achieved 95.6% accuracy in a three-class pneumonia classification framework. Similarly, Ibrahim et al. Jin et al. [16] proposed a three-stage hybrid model consisting of feature extractor, feature selector, and SVM classifier stage, which achieved a remarkable 98.62% accuracy on a private set of 1743 X-ray images. Karthik et al. [17] employed the so-called Channel-Shuffled Dual-Branched (CSDB) CNN architecture to distinguish various types of pneumonia from several publicly available datasets, achieving F1 scores of 94–98%. Quan et al. [18] combined the Dense Net and the CapsNet architecture and achieved 96% recall on a small sized COVID-19 dataset, but only 90.7% accuracy on larger set of pneumonia X-ray images.
Streamlit interface has been made with the reference of its official websites [10] and some tutorials [11][12].

III. METHODOLOGY

This section provides a description of the proposed method.

A. Diabetes

In the process of building of the diabetes model we have used the support vector machine. The process is same for the SVM model as logistic model. Here we have trained the model with the database of diabetes from the Kaggle. Here firstly we must normalize the data into same intervals. Then the Support vector machine classifier is used to enhance this data so for that we must create the SVM model [1], and we fit this data into that model. As a result, we will get the SVM classifier model. This SVM classifier model [4] is checked with the testing data and after getting a good accuracy.

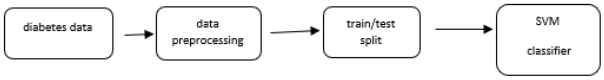


Fig 2. flow of Diabetes model

B. Heart Disease

In the process of building the logistic regression model that detects the person's heart is healthy or not, we first enhance the heart data. The data must be pre-processed like removing null values from the data and split the data in a specific way, in this paper we split the train and test data as 80 percentage and 20 percentage respectively. For creating a Logistic regression model, we must import the logistic regression method from sklearn library [14] in the module of linear model and for the training the model we have to fit the train data into the model so that the model will be trained with our data. Then after enhancing the model, we must check the model accuracy with the test data. In this paper we are getting accuracy of 90 percentage. Now testing the model[15] with the random data, we can enhance the results and the model is saved using pickle library for setting it up for the backend of the interface.
This model is used in the interface for the new inputs of model through the interface

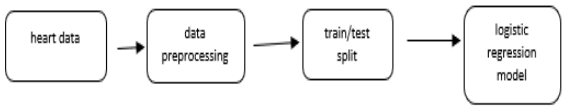


Fig 2. flow of heart disease model

C. Pneumonia

In the training of pneumonia disease model firstly we have the data of two classes of x-ray photos which are matched with having the disease or not. Using those photos we train a CNN model using conv2d method of keras library. Different inner layers are used in the process of building convolution neural network. The summary has been embedded in the image of the CNN model and the different activation functions, pooling layers. After the training of the CNN model and running for certain number of epochs and enhancing the good accuracy. After getting the good accuracy the model has been saved as the pneumonia model. This pneumonia model is used for the backend of interface to enhance the inputs of x-ray images for pneumonia through interface.

D. Interface

The interface is made through the streamlit framework where we have used the option menu streamlit library which enhances different options embedded in it where the selected option is showed on the display page. By using these option menus, the different diseases are embedded in those options where the model are loaded in those options and the user can use those disease through these option menus[10]. Selected disease model will be worked at the backend and the inputs are taken through these interface and respected model give the output for the inputs of the user.

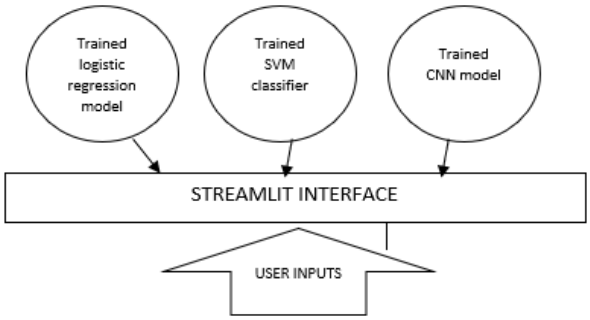


Fig 4. Interface Flow

IV. IMPLEMENTATION

A. Diabetes Prediction

1) Dataset: For performing machine learning the first and most important step is the availability of Dataset. We collected dataset from a public data repository KAGGLE. More attributes are added in our dataset. Figure 5. Is the Screenshot of the dataset. The data set consists of 9 attributes, in which 9th attribute is the class label.0 means having diabetes and 1 means not having diabetes.

```
diabetes_dataset.head()
```

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

Fig 5. dataset first 5 rows

The data does not have any null values, so we directly enhanced with SVM Classifier training after standardizing the data.

2) SVM Classifier Training: The dataset was first split into 60:40 that is 60% of data was used for training the machine learning model and 40% of data was used for testing. After evaluating the performance, it was concluded that the model showed 70% of Accuracy. To improve the accuracy of the proposed model data was split into 80:20, 80% of data is used for training the Support Vector Machine and 20% is used to test the model. After evaluating the model, the accuracy of 80% is achieved.

B. Heart Disease Prediction

1) Dataset: The data from UCI machine learning repository was collected. The dataset contains 303 records, and 14 attributes. Thirteen parameters were used as the eigenvalues for the forecast of heart disease and one of the parameters is the output value or the forecast value of the patients with heart disease. 0 refers to no heart disease and 1 refers the heart is not healthy.

```
heart_data.head()
```

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Fig 6. dataset first 5 rows

2) *Logistic Model Training:* The Data was split into 80-20 percentage, where the training data is standardized and fitted into the logistic model. After training the model. The model is then tested with the test data and we gt the accuracy of 80 percentage where it can be used for the raw data. now the model is deployed at the backend of the interface.

C. Pneumonia Detection

1) *Dataset:* The dataset used for this study originates from one of Kaggle's many deep learning competitions.1 The dataset depicts lung X-ray images of infants aged one through five. The lung X-ray images taken from Guangzhou Women and Children's Medical Center were verified by medical experts. All chest X-ray imaging were performed as part of patients' routine clinical care. The dataset consisted of 5856 labeled images, of which 4273 showed pneumonia, while the remaining 1583 were negatives. Due to the imbalance in the dataset, a generative adversarial network was used to generate further images for the minority class (which were used exclusively during training). No generated image was used for the evaluation of the algorithm. All scans were single-channel intensity images and their dimensions varied from 1346 * 1044 to 2090 * 1858 pixels. All images were transformed to the 224 * 224 * 3 format to comply with the expected input of most CNN network architectures.

2) *CNN Model Training:* The model follows the pattern of convolutional layer followed by pooling layer[8][9]. This pattern is repeated several times, because it yet again proves the power of the convolutional layers and its ability to learn specific features, which are often referred to as feature maps when many of them are stacked together . Filters are essentially the elements that enable the neural network to pick up on specific patterns such as edges. Filters are often initialized in many ways but ultimately that can be omitted, because the purpose of the convolutional layer is to learn the specific features during the training; they are basically the mechanism needed for feature extraction. At last, the matrices are flattened[7] and are then connected to the dense fully connected layers, which are responsible for the classification. It is worth noting that within each convolutional block in the model's architecture there is a batch normalization layer followed by a ReLU activation function. The given pictures shows the flow of CNN layers in the development of the CNN model and the epochs are implemented and model is saved after getting an accuracy of 97 percentage.

D. Streamlit Interface

The implementation of the interface happens through streamlit-option-menu [10] library where the three options are diabetes, heart disease, Pneumonia and the models are loaded on their respected options. The diabetes takes the input of 8 attributes with specified values mentioned and the button get result shows the result after giving all inputs[12]. Same as diabetes the heart disease takes 13 attributes as specified names and gives the result. Whereas the Pneumonia model asks to upload the x-ray photo and gives the result whether it is normal or pneumonia. Fig shows the interface of three diseases as options at the left of the interface.

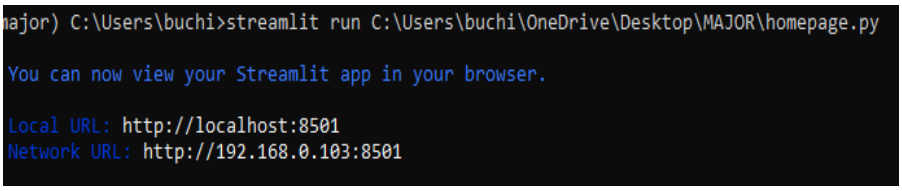


Fig 7.running streamlit in terminal

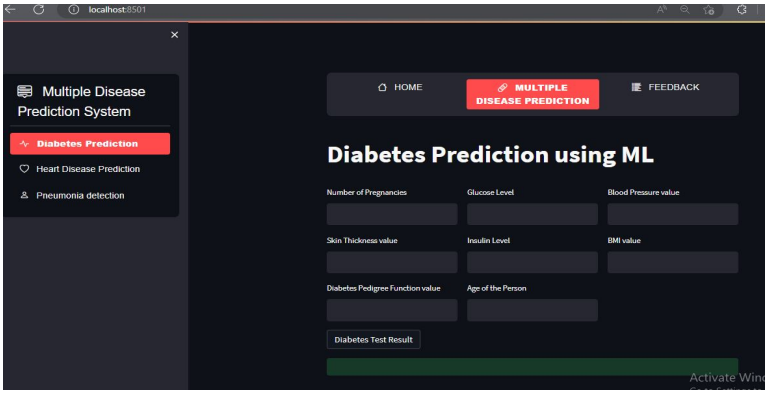


Fig 8. Streamlit Interface

V. EXPERIMENTAL RESULTS

A. Results

1) The below figure shows the output of the diabetes model in the interface.

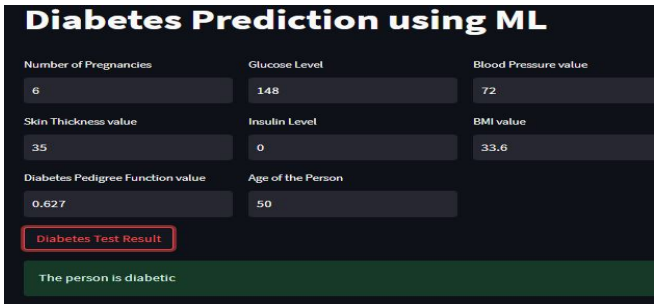


Fig 9. diabetes result

2) The below figure shows the output of the heart disease model in the interface

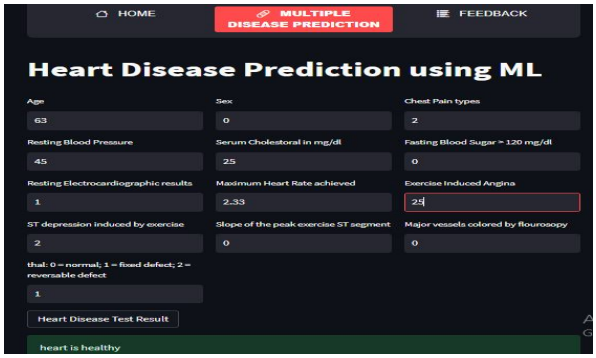


Fig 10. heart result

3) The below figure shows the output of the heart disease model in the interface

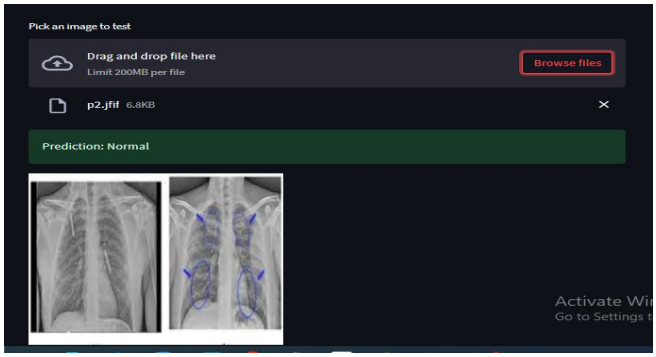


Fig 11. Pneumonia result

VI. CONCLUSION AND FUTURE SCOPE

In this paper we concluded that the three different disease model has been embedded into a single interface where the user can interact with all of them at a time as per their need .

The future scope of this project can be in the two different parts where the first part is developing the models with other machine learning techniques and enhancing different good accuracies than the ones which we deployed. When it comes logistic, SVM models the classification can be happened in many ways it has scope different good algorithms . The pneumonia model can be enhanced through more data and can have a scope of building a independent model with a better accuracy. The second part is about the interface where it has the scope of embedding many more disease prediction models using different machine learning techniques and the interface can be made more colorful and interactive using advance element options of streamlit. This can also be deployed into some official clouds for the universal URL access.

REFERENCES

[1] Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, et al. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham offspring study. Arch Intern Med. 2007;167:1068–74 Montu Saw, Tarun Saxena, Sanjana Kaithwas, Rahul Yadav, Nidhi Lal. Estimation of Prediction for Getting Heart Disease Using Logistic Regression Model of Machine Learning

[2] Mashayekhi M, Prescod F, Shah B, Dong L, Keshavjee K, Guergachi A. Evaluating the performance of the Framingham diabetes risk scoring model in Canadian electronic medical records. Can J Diabetes. 2015;39(30):152–6 Mashayekhi M, Prescod F, Shah B, Dong L, Keshavjee K, Guergachi A. Evaluating the performance of the Framingham diabetes risk scoring model in Canadian electronic medical records. Can J Diabetes. 2015;39(30):152–6

[3] Cortes, C., Vapnik, V., "Support-vector networks", Machine Learning, 20(2),pp. 273-297, 1995.

[4] Meng XH, Huang YX, Rao DP, Zhang Q, Liu Q. Comparison of three data mining models for predicting diabetes or prediabetes by risk factors.

[5] Herron P., "Machine Learning for Medical Decision Support: Evaluating Diagnostic Performance of Machine Learning Classification Algorithms", INLS 110, Data Mining, 2004

[6] Jain R, Nagrath P, Kataria G, Kaushik VS, Hemanth DJ. Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning. Measurement 2020;165 . https://doi.org/10.1016/j. measurement.2020.108046 108046.

[7] Dey N, Zhang YD, Rajinikanth V, Pugalenthi R, Sri Madhava Raja N. Customized VGG19 Architecture for Pneumonia Detection in Chest X-Rays. Patt. Recogn Lett 2021;143(67–74). https://doi.org/10.1016/j.patrec.2020.12.010.

[8] Brunese L, Mercaldo F, Reginelli A, Santone A. Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays. Comput Meth Prog Biomed 2020;196 . https://doi.org/10.1016/j.cmpb.2020.105608 105608.

[9] Panwar H, Gupta PK, Siddiqui MK, Morales-Menendez R, Bhardwaj P, Singh V. A deep learning and grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images. Chaos, Solitons & Fractals 2020;140 . https://doi.org/10.1016/j. chaos.2020.110190 110190.

[10] https://docs.streamlit.io/ -website to implement its componenets

[11] https://towardsdatascience.com/rational-ui-design-with-streamlit-61619f7a6ea4

[12] https://www.section.io/engineering-education/streamlit-ui-tutorial/ by Rahul Banerjee -tutorial for building an UI using streamlit interface

[13] Sana Bharti, Dr.Shaliendra Narayan Singh,Analytical Study of Heart Disease Prediction Comparing With Different Algorithms.

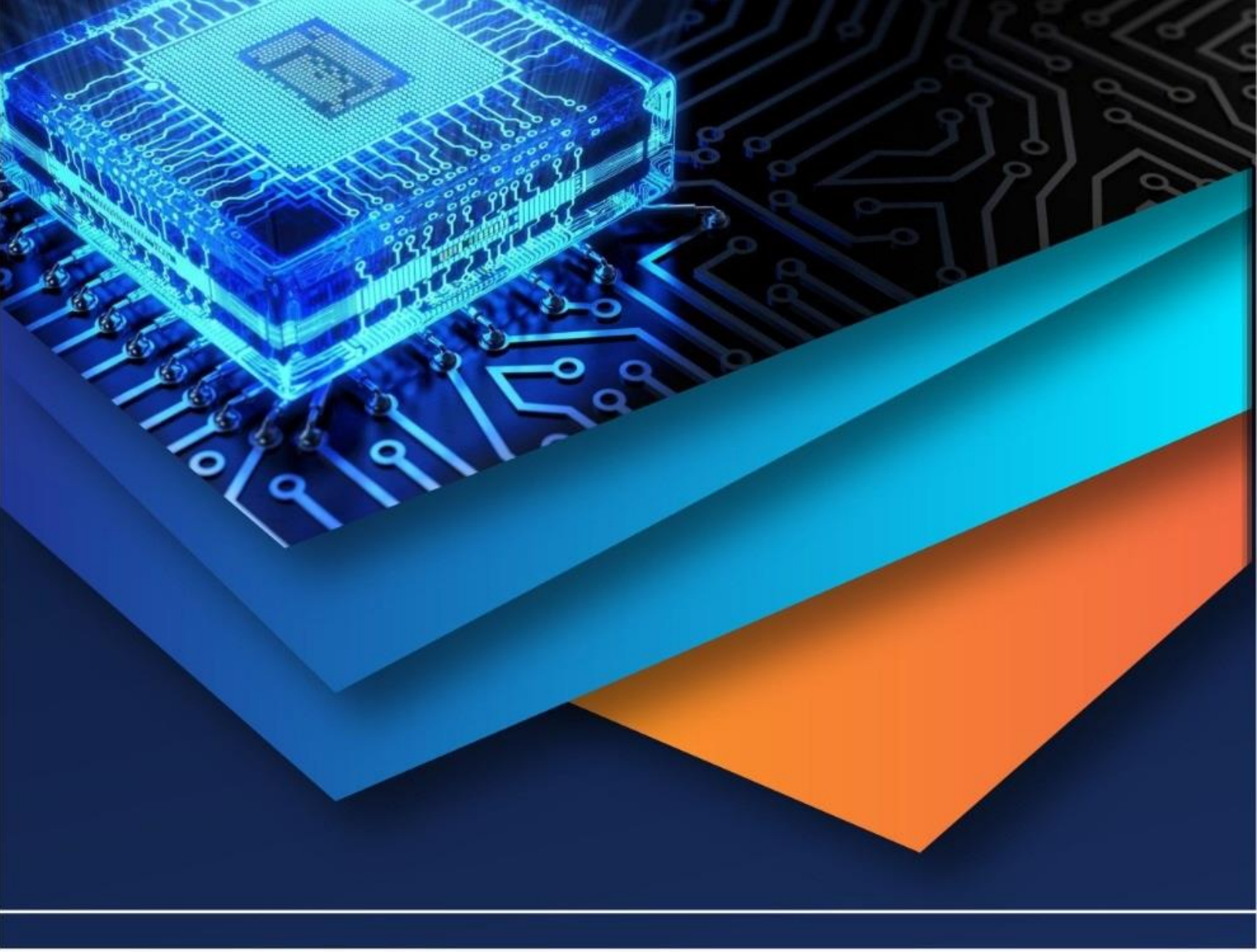
[14] Cincy Raju, Philipsy E, Siji Chacko, L Padma Suresh, Deepa Rajan S, A Survey on Predicting Heart Disease using Data Mining Techniques.

[15] Mohammad Shafenoor Amin, Yin Kia Chiam, Kasturi Dewi Varathan. Identification of significant features and data mining techniques in predicting heart disease.

[16] Jin WQ, Dong SQ, Dong CZ, Ye XD. Hybrid ensemble model for differential diagnosis between COVID-19 and common viral pneumonia by chest X-ray radiograph. Comput Biol Med 2021;131 . https://doi.org/10.1016/j.compbiomed.2021.104252 104252

[17] Karthik R, Menaka R, Hariharan M. Learning distinctive filters for COVID-19 detection from chest X-ray using shuffled residual CNN. Appl Soft Comput 2021;99 . https://doi.org/ 10.1016/j.asoc.2020.106744 106744

[18] Quan H, Xu XS, Zheng TT, Li Z, Zhao MF, Cui XY. DenseCapsNet: Detection of COVID-19 from X-ray images using a capsule neural network. Comput Biol Med 2021;133 . https://doi.org/10.1016/j.compbiomed.2021.104399 104399.



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)