*International Journal of Research in Engineering and Science (IJRES)*
*ISSN (Online): 2320-9364, ISSN (Print): 2320-9356*
*www.ijres.org Volume 09 Issue 10 ǁ 2021 ǁ PP. 86-90*

# Virtual Hospital
# A Machine Learning based Medical Test web which makes predictions about various diseases.

Aniket Jadhav, Dhananjay Rajput, Shubham pawar, Prof.Tanmayee kute
*P.E.S Modern college of engineering Department of Information Technology, Pune-05,India.*

***Abstract:***
*Virtual Hospital is a Machine Learning based Medical Test web which makes predictions about various diseases based on the information/inputs or the symptoms user enter into the system and provides the accurate results based on that information. Many of the prevailing machine learning models for health care analysis are concentrating on one disease per analysis. Like one analysis if for diabetes analysis, one for cancer analysis, one for kidney related diseases like that. There is no common system where one analysis can perform multiple diseases prediction. In this article proposing a system which predict multiple diseases by using flask API. Virtual hospital is a system used to predict cancer detection, diabetes,heart attack detection, liver disease detection, kidney disease detection,malaria detection and pneumonia detection and many more diseases can be included. We also provide some cure and prevention of that disease along with details information about these diseases,also it will sugest some doctors if needed.virtual hospital also contains third party chatbot for virtual medical help.The importance of this analysis to analyse the maximum diseases, so that to monitor the patient's condition and warn the patients in advance to decrease mortality rate.along with it our main motive to develop this project is the user can sit at their convenient place at their convenient time and have check up of their health.*

---
Date of Submission: 06-10-2021 Date of acceptance: 20-10-2021
---

## I. Introduction:

According to research 40% of people who ignores about general diseases which leads to harmful diseases later .The main reason is laziness to consult a doctor and time concern that people involve themselves so much that they have no time to take an appointment and consult doctor which later result into fatal diseases.According to research 70% people in India suffers from general diseases and 25% people face death due to early ignorance.Now-a-days, people face various diseases due to the environmental condition and their living habits. So the prediction of these lethal disease at earlier stage becomes important task. But the accurate prediction on the basis of symptoms becomes too difficult for doctor at some stages hence the role of Artificial Intelligence in healthcare is very crucial, Taking this in mind we build web application called virtual hospital system which predicts disease based on the information or the symtoms and user inputs . we present a website in which the following applications are implemented cancer detection,diabetes ,heart attack detection` liver disease detection, kidney disease detection , malaria detection and pneumonia detection in virtual hospital . For small problems, the users need to go personally to the hospital for check-up which is longer consuming, As well as handling the telephonic calls for appointments is quite hectic. Our aim is to solve Such a problem by using new technology like machine learning and deep learning by giving proper guidance regarding healthy living. We also provide some cure and prevention of that disease along with details information about these diseases,also it will sugest some doctors if needed.

For example for heart disease analysis in many existing systems considered few parameters like Age,gender ,height, weight, ap lo, ap hi,cholesterol ,gloc ,smoke or not,active or not , Final models behaviour will be saved as python pickle file. Flask API is designed. When user accessing this API, the user has got to send the parameters of the disease along side disease name. virtual hospital will invoke the corresponding model and returns the status of the patient. The importance of this analysis to analyse the maximum diseases, so that to monitor the patient's condition and warn the patients in advance,also suggest some doctors contacts if needed to decrease mortality ratio.this method is followed for all seven diseases in virtual hospital are predict diseases.

The main motivation of the virtual hospital is to create a web application using the Flask framework and machine learning technique to predict the multiple diseases.so the user can sit at their convenient place, at their convenient time and have check up of their health.

## II. Literature servey :

1. A deep convolutional neural network to classify pulmonary tuberculosis was developed by Lakhani et al. Transfer learning models like AlexNet and GoogleNet were also used to classify chest X-ray images. The dataset was split into training, testing and validation sets as 68%, 14.9% and 17.1%, respectively.

2. Rubin et al.developed a CNN model to detect common thorax disease from frontal and lateral chest X-ray images. MIMIC-CXR dataset was used to perform large-scale automated recognition of these images. The dataset was split into training, testing and validation sets as 70%, 20% and 10%, respectively.

3. Song et al.explained and described using various factors such as Age, Glucose, BP, BMI, Skin Thickness etc. Diabetes Pedigree function, insulin, and pregnancy parameters not concluded.

4. Translational Lung Cancer Research, June 2018, Lung cancer prediction using machine learning and advanced imaging techniques, Timor Kadir and Fergus Gleeson, The demonstration of a 20% reduction in lung cancer mortality within the USA National Lung Screening Trial (NLST)

5. S.Ramya and Dr.N.Radha worked on diagnosis time and improvement of diagnosis accuracy using various classification algorithms of machine learning. The proposed work deals with classification of various stages of CKD according to its gravity. By analysing different algorithms such as Basic Propagation Neural Network, RBF and RF. The analysis results indicates that RBF algorithm gives better results than the another classifiers and produces 85.3% accuracy.

6. S.Dilli Arasu and Dr. R. Thirumalaiselvi has worked on missing values in a dataset of chronic Kidney Disease. Missing values in dataset will reduce the accuracy of our model also of prediction results. They find solution over this problem that they performed a recalculation process on CKD stages and by doing so they got up with unknown values. They replaced missing values with recalculated values

7. Sajida et al. in discusses the role of Adaboost and Bagging ensemble machine learning methods using J48 decision tree as the basis for classifying the Diabetes Mellitus and patients as diabetic or non diabetic, based on diabetes risk factors. Results achieved after the experiment proves that, Adaboost machine learning ensemble technique outperforms well comparatively bagging as well as a J48 decision tree.

8. Pradhan et al inused Genetic programming (GP) for the training and testing of the database for prediction of diabetes by employing Diabetes data set which is sourced from UCI repository. Results achieved using Genetic Programming,gives optimal accuracy as compared to other implemented techniques. There are often significant improve in accuracy by taking less time for classifier generation. It proves to be useful for diabetes prediction at low cost.

## III. Methodology :

**1.pneumonia detection and maleria disease detection using cnn algorithm:**

Convolutional networks were inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex. A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.

Database collection: Initial step for any image processing based project is acquiring proper dataset which is valid . Most of the time the standard database is preferred but in certain circumstances we do not get proper dataset .Dataset used in this project for **pneumonia detection and maleria disease detection** is from kaggle. Chest X-Ray Images (Pneumonia) dataset of 1.16 GB size has been imported from Kaggle with jpeg images split into Train, Test and Val folders each divided into category Pneumonia and Normal. Data available here is not labeled .So the first task is to clean and label the dataset. There is a huge dataset so basically the images with better resolution and angle are selected. this dataset contains different plant diseases images of multiple plants.

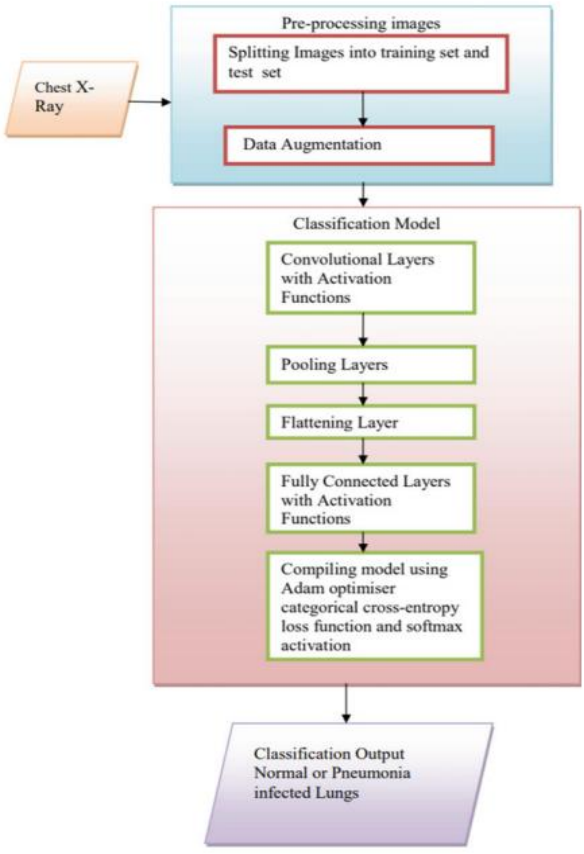
Preprocessing and Training the model (CNN): The dataset is Preprocessed such as Image reshaping ,resizing and conversion to an array form. Similar processing is also done on the test image. A database consisting of different x-ray images is obtained , out of which any image can be used as a test image for the software. The train database is used to train the model (CNN ) so that it can identify the test image and the disease it has .

After the cleaning and preprocessing of dataset is done ,main algorithm is implemented as follows

**CNN Architecture :**

CNN models are feed-forward networks with convolutional layers, pooling layers, flattening layers and fully connected layers employing suitable activation functions.

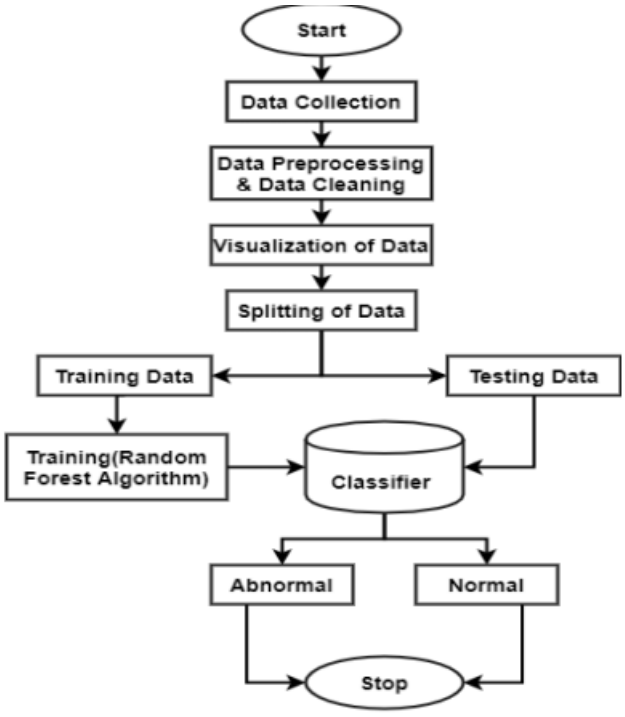
- Convolutional layer: It is the building block of the CNNs. Convolution operation is done in mathematics to merge two functions . In the CNN models, the input image is first converted into matrix form. Convolution filter is applied to the input matrix which slides over it, performing element-wise multiplication and storing the sum. This creates a feature map. 3 × 3 filter is generally employed to create 2D feature maps when images are black and white. Convolutions are performed in 3D when the input image is represented as a 3D matrix where the RGB color represents the third dimension. Several feature detectors are operated with the input matrix to generate a layer of feature maps which thus forms the convolutional layer.
- Activation functions: All four models presented in this paper use two different activation functions, namely ReLU activation function and softmax activation function. The ReLU activation function stands for rectified linear function. It is a nonlinear function that outputs zero when the input is negative and outputs one when the input is positive. ReLU activation function has many variants such as Noisy ReLUs, Leaky ReLUs and Parametric ReLUs. Advantages of ReLU over other activation functions are computational simplicity and representational sparsity. Softmax activation function is used in all four models presented in this paper. This broadly used activation function is employed in the last dense layer of all the four models. This activation function normalizes inputs into a probability distribution. Categorical cross-entropy cost function is mostly used with this type of activation function.
- Pooling layer: Convolutional layers are followed by pooling layers. The type of pooling layer used in all four models is max-pooling layers. The max-pooling layer having a dimension of 2 × 2 selects the maximum pixel intensity values from the window of the image currently covered by the kernel. Max-pooling is used to down sample images, hence reducing the dimensionality and complexity of the image . Two other types of pooling layers can also be used which are general pooling and overlapping pooling..
- Flattening layer and fully connected layers: After the input image passes through the convolutional layer and the pooling layer, it is fed into the flattening layer. This layer flattens out the input image into a column, further reducing its computational complexity. This is then fed into the fully connected layer/dense layer. The fully connected layer has multiple layers, and every node in the first layer is connected to every node in the second layer. Each layer in the fully connected layer extracts features, and on this basis, the network makes a prediction . This process is known as forward propagation. After forward propagation, a cost function is calculated. It is a measure of performance of a neural network model. The cost function used in all four models is categorical cross-entropy. After the cost function is calculated, back propagation takes place. This process is repeated until the network achieves optimum performance.



**2. Heart disease detection and diabetes detection,kidney disease detection using Random forest algorithm:**

RANDOM FOREST (RF) It is one of the prediction algorithms in the machine learning area. It is more adaptable to ensemble approach. It can easily tackle large datasets.

Data Collection is the major step as the quality and quantity of the data that we gather for the proposed system will directly determine how good the results of the predictive model are. We have collected the dataset for heart diseses and kidney diseses from the Kaggle. For better and accurate prediction, we consider parameters like Age, Sex, Resting blood pressure (in mm Hg), serum cholesterol (in mg/dl), Fasting blood sugar Rest ECG results and Chest pain type in heart disease prediction and Pregnancies,glucose,blood pressure,insuline,BMI,skin thickness,age this inputs for diabetes detection. Classification is the task of approximating the mapping function from the input variable to the discrete output variable. It is the classification where the outcome is either true or false (1 or 0). Random forest is an estimator that fits several decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. In the regression model, the prediction is based on the independent variable. Random forest in regression operates on constructing a multitude of decision trees at training time and outputting the class that is the mode of mean prediction of the individual trees. After reading all the data the data visualization method is performed. In this method, the large data sets are transformed into a statistical and graphical representation.




Data processing is defined as the collection and manipulating of data to produce the desired meaning and understandable data for prediction. In this stage regression and classification, techniques are used for the process. The main steps that included in data processing are as :

- Import the libraries and datasets.
- Data cleaning.

Data Transformation.

Data visualization is the method of transforming large data sets into a statistical and graphical representation. It is an important task of data science and these are the useful techniques that make the data less confusing and more accessible. Data Visualization takes a huge complex amount of data and then represents them in the form of charts or graphs for better understanding. Analysis using a Heat map, bar graphs, and pair plots .

The splitting of data is classified into training and testing of data. Training is applied to the 75 parts of the data set. Testing is applied to the 25 parts of the dataset. Testing the data is used to evaluate the performance of the model using a few algorithms. Based on the training data and testing data the best model is selected. The training data is different from testing data; the obtained data is applied to the algorithm. The flowchart of the proposed methodology shown above. The work aims to predict the diagnosis of different disease and its stages..

Random Forest is a supervised learning algorithm. It is an ensemble learning method for classification, regression, and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the class or mean prediction of the individual trees. This algorithm creates decision trees on data samples. There is a direct relationship between the number of trees in the RF and the results it can get. After creating the decision trees, it gets the prediction from each of them (tree), and then finally it selects the best solution employing voting. The larger the number of trees, the more would be the accuracy of the result. The larger the trees the better will be accuracy. The main advantage of using Random forest is that it has high accuracy and less variance than a single decision tree. The Data Pre-processing is done and then based on the parameters in the dataset; the number of trees formed as shown in figure 4. The accuracy of the result directly depends on the number of trees, so if the number of trees is more the accuracy would be high. In the proposed model, we split the data into 75% training data and 25% testing data. Accordingly, the data is trained and tested in all possible combinations and finally, it gives the best model. Now, the model is trained with the help of the Random Forest Classifier. From each tree, we get a predicted result and have entropy which is calculated individually. Voting (which includes combining the predictions) is performed for all the predicted results and finally, the most voted prediction result is selected.

### 3. liver disease using KNN algorithm:

KNN

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.KNN is a semi-supervised and competitive learning method that belongs to the family of instance based algorithms. It creates its model based on training dataset and predicts a new data case by searching training data for the most similar cases. It retains all observations selected at the time of training. This prediction data case of k-most similar cases is recapitulated and returned as the forecast for a new case. The selection of distance metric functions for finding similarity measure depends on structure of data. Available functions are euclidean, cityblock, cosine, correlation and hamming out of which correlation performed best for this study and hamming was not supportive as it can only be used for categorical or binary data. K-Nearest Neighbour algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.

## IV. Conclusion :

In this project, we make a machine learning based web system called virtual hospital, which can be effectively use for diagnostic purposes for early detection of multiple diseases with appropriate parameter inputs.also it will give detail information about perticular disease as well as able to help you with some doctors contact. Virtual Hospital will be simple to use and understand so that every user can use it at their convenient time and place for their check-ups. It will help to reduce the mortality rate , early detection of fatal diseases,eventually contributing to healthcare industry.

## References :


[1]. A. Gavhane, G. Kokkula, I. Pandya, and K. Devadkar, "Prediction of heart disease using machine learning," in 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 1275–1278

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

[3]. Y. Amirgaliyev, S. Shamiluulu, and A. Serek, "Analysis of chronic kidney disease dataset by applying machine learning methods," in 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT), 2018, pp. 1–4.

[4]. A. Charleonnan, T. Fufaung, T. Niyomwong, W. Chokchueypattanakit, S. Suwannawach, and N. Ninchawee, "Predictive analytics for chronic kidney disease using machine learning techniques," 2016 Management and Innovation Technology International Conference, MITiCON 2016, pp. MIT80–MIT83, 2017.

[5]. P. P. Sengar, M. J. Gaikwad, and A. S. Nagdive, "Comparative study of machine learning algorithms for cancer prediction," Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020, pp. 796–801, 2020.

[6]. World Health Organization, Household Air Pollution and Health [Fact Sheet], WHO, Geneva, Switzerland, 2018, http:// www.who.int/newa-room/fact-sheets/detail/household-airpollution-and-health.

[7]. I. Rudan, L. Tomaskovic, C. Boschi-Pinto, and H. Campbell, "Global estimate of the incidence of clinical pneumonia among children under five years of age," in Bulletin of the World Health Organization, vol. 82, pp. 85–903, 2004.

[8]. R. Olaf, F. Philipp, and B. .omas, U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI Springer, New York, NY, USA, 2015.

[9]. B. Vijay, K. Alex, and C. Roberto, "Segnet: Deep convolutional encoder-decoder architecture for image segmentation," 2015, http://arxiv.org/abs/1511.00561.

[10]. M. Aliasghar, K. Rashed, R. Kawal, B. Jeremy, and B. Ulas, Cardiacnet, Segmentation of Left Atrium and Proximal Pulmonary Veins from MRI using Multi-View CNN, MICCAI, Springer, New York, NY, USA, 2017.

[11]. Z. Xue, D. You, S. Candemir et al., "Chest x-ray image view classification," in Proceedings of the Computer-Based Medical Systems IEEE 28th International Symposium, Sao Paulo, Brazil, June 2015.