Major Project-II
ABSTRACT
Multiple Disease Predictions using ML and IP

Review -1 Report

Division: 8TC3 **Submission Date: 08/01/2022**

Team Details

Sr. No.	Enrolment No.	Student Name	Division
1	91800103168	Yash Nawani	TC3
2	91800103145	Priyankkumar Gondha	TC3

ABSTRACT

The wide adaptation of computer-based technology in the health care industry resulted in the accumulation of electronic data. Due to the substantial amounts of data, medical doctors are facing challenges to analyze symptoms accurately and identify diseases at an early stage. However, supervised machine learning (ML) algorithms have showcased significant potentialin surpassing standard systems for disease diagnosis and aiding medical experts in the early detection of high-risk diseases. In this literature, the aim is to recognize trends across various types of supervised ML models in disease detection through the examination of performance metrics. The most prominently discussed supervised ML algorithms were Naïve Bayes (NB), Decision Trees (DT), K-Nearest Neighbor (KNN). As per findings. The Logistic Regression (LR) performed highly at the prediction of heart diseases. Finally, Random Forest (RF), and Convolutional Neural Networks (CNN) predicted in precision of brain tumor diseases and common diseases, respectively.

We have referred various books regarding different components working. I have taken ideas of the project from different website and I have also seen YouTube videos and used Google search engine for completing my project.

BACKGROUND WORK

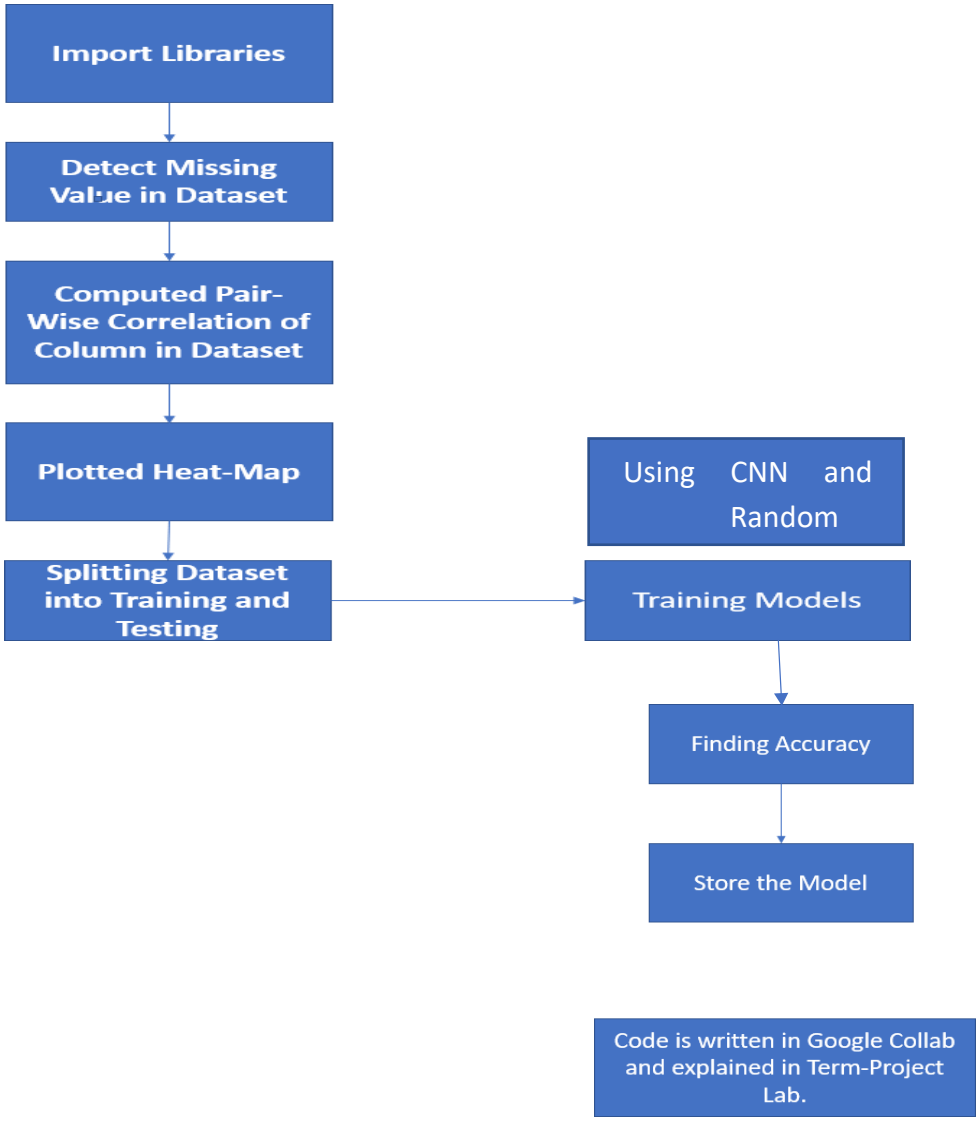
The Earth is going through a purplish patch of technology where the demand of intelligence and accuracy is increasing behind it. Today's people are likely addicted to internet but they are not concerned about their physical health. People ignore the small problem and don't visit to visit hospital which turn into serious disease with time. Taking the advantage of this growing technology, our basis aim is to develop such a system that will predict the multiple diseases in accordance with symptoms put down by the patients without visiting the hospitals / physicians. Machine Learning is a subset of AI that is mainly deal with the study of algorithms which improve with the use of data and experience. Machine Learning has two phases i.e., Training and Testing. Machine Learning provides an efficient platform in medical fieldto solve various healthcare issues at a much faster rate. There are two kinds of Machine Learning – Supervised Learning and Unsupervised Learning. In supervised learning we frame a model with the help of data that is well labelled. On the other hand, unsupervised learning model learn from unlabeled data. The intent is to deduce a satisfactory Machine Learning algorithm which is efficient and accurate for the prediction of disease. In this paper, the supervised Machine Learning concept is used for predicting the diseases. The main feature will be Machine Learning in which we will be using algorithms such as Decision Tree, Random Forest, Naïve Bayes and KNN which will help in early prediction of diseases accurately and better patient care.

Limitations

- Large number of datasets is required for getting higher accuracy.
- Time taken by brain tumor predictor to predict disease is more.
- We have to fill all the blocks for diabetes and heart disease prediction in webapp, if any of the block is left or not filled then it will show invalid data.
- We have only able to add 3 types of disease due to time constraint

METHODOLOGY

Flow Chart:



Working:

Heart and diabetes disease working:

First, we import libraries which is necessary for prediction of diabetes and heart diseases. Thenwe add the csv file to Google Collab. Then we check for null values in rows and columns. If there is any null value, then it will be replaced by zero. Then we plotted the Heat-Map to find correlation between various parameter. Then we have divided the dataset into training and testing part. Test=0.2 and training=0.8. Then we apply Random Forest Algorithm to predict the disease. Later, we plotted confusion matrix and calculated the accuracy of the prediction. After that we save the model.

Brain tumor disease working:

First, we import libraries which is necessary for prediction of Brain Tumor. Then we mount from google drive like images of training and testing. There are 4 types of tumors like glioma tumor, no tumor, meningioma tumor, pituitary tumor.

Then we have done Image Data Augmentation. Image data augmentation is a technique that can be used to artificially expand the size of a training dataset by creating modified versions ofimages in the dataset. It uses techniques such as flipping, zooming, padding, cropping, etc. Then we Performed One Hot Encoding on the labels after converting it into numerical values:A one hot encoding is appropriate for categorical data where no relationship exists between categories. After that we have performed Transfer Learning.

Deep convolutional neural network models may take days or even weeks to train on very large datasets. A way to short-cut this process is to re-use the model weights from pre-trained modelsthat were developed for standard computer vision benchmark datasets, such as the ImageNet image recognition tasks. After that we have done GlobalAveragePooling2D. This really helpsin decreasing the computational load on the machine while training. Then we have divided thedataset into training and testing part. Then we apply CNN Algorithm to predict the disease. Later, calculated the accuracy of the prediction. After that we save the model. We have also created a widget or a button to upload the image and get the result.

Trouble Shooting:

Verify the metrics for model underfit or overfit. Using various algorithm for increasing accuracy of the model like decisions tree, linear regression,random forest, CNN etc.

Data quality Machine learning systems rely on data. Machine Learning heavily depends upon dataadded to it. If there is good data, then accuracy increases and vice-versa.

TECHNOLOGY TO BE USED

Software requirements

- Google Collab
- Visual studio
- Python
- CSS, HTML

Hardware requirements

- 64- bit or 32-bit Processor
- 2 GB RAM 32-bit or 64-bit
- 16 GB Hard disk space
- NVIDIA GT 730 Graphics

REFERENCES

D. Dahiwade, G. Patle, and E. Meshram, “Designing disease predictionmodel using machine learning approach,” Proceedings of the 3rd Inter-national Conference on Computing Methodologies and Communication,ICCMC 2019, no. Iccmc, pp. 1211–1215, 2019.

S. Jadhav, R. Kasar, N. Lade, M. Patil, and S. Kolte, “Disease Predictionby Machine Learning from Healthcare Communities,” InternationalJournal of Scientific Research in Science and Technology, pp. 29–35,2019.

R. Saravanan and P. Sujatha, “A state of art techniques on machinelearning algorithms: A perspective of supervised learning approachesin data classification,” in 2018 Second International Conference onIntelligent Computing and Control Systems (ICICCS), 2018, pp. 945–949.

STUDENT DETAILS WITH SIGNATURE

Sr. No.	Enrolment No.	Student Name	Signature
1	91800103168	Yash Nawani	
2	91800103145	Priyankkumar Gondha	

INTERNAL GUIDE WITH SIGNATURE

Name:

Institute Email ID:

Contact:

Signature: