DETECTING COVID-19 FROM QT DATA SET USING CLASSIFICATION

1st OMI, SHAHED CHOWDHURY
BSc. CSE
American International University-Bangladesh
Dhaka, Bangladesh
shahedchowdhury120@gmail.com

2nd AHMED, RAIHAN
BSc. CSE
American International University-Bangladesh
Dhaka, Bangladesh
raihanahmed@gmail.com

3rd RAHMAN, MD. SADMAN
BSc. CSE
American International University-Bangladesh
Dhaka, Bangladesh
sadmanrman@gmail.com

4th AHMED, ABIR
BSc. CSE
American International University-Bangladesh
Dhaka, Bangladesh
abir64844@gmail.com

***Abstract*—During the early stages of COVID 19, the world was short of COVID test kits due to a supply shortage that did not meet requirement. Due to a lack of kits, testing for COVID 19 was quite expensive. Even though the cost was lower in the public hospital, people had to wait a long time to be tested. If there was a system that could accurately predict whether a patient was COVID positive or negative, people could easily test their COVID status at home, which would be free of charge and relieve the patient of report tampering. We are developing a machine learning based model which will help people to get to know about their possibility of catching COVID.**

Keywords: COVID-19, CoronaVirus, Machine Learning, SVM Classifier, Prediction, Infections

I. Introduction

People throughout the world were astonished by the Noble Coronavirus which occurred in late 2019, better known as COVID-19. Many nations started implementing temporary lockdown due to rising death rates, but when the rate grew quickly, full lockdown were imposed, and people began to adapt to working at home. During the beginning phase of COVID 19, the world was short of test kits for COVID, due to the shortage of supply which did not meet the demand. Due to the shortage of kits, it was quite expensive to test for COVID 19, though in the public hospital the cost was lesser people had to wait very long to get a chance for testing. Many private hospitals took advantage of this situation and charged extra money from the patient and a few also admitted patients with fake reports.

If there was a system that could suggest if the patient was either COVID positive or negative with high accuracy, people could easily test their COVID status at home, which would be free of cost and relieve the patient from report tampering. Therefore, we are building a project whose objective is to classify COVID as positive or negative using a Support Vector Machine classifier by using the features, Temperature, Pulse rate, and Oxygen saturation.

II. Literature Review

A.

The status of COVID-19 in cough audio signals is automatically identified using a new framework that is given in this study. The suggested architecture is demonstrated and comprises of a hybrid GA-ML with six different machine learning algorithms: decision tree, support vector machine, logistic regression, Naive Bayes, and linear regression. The framework also incorporates various crucial steps to carry out the innovative Coronavirus diagnostic process based on cough audio signals [1].

B.

The authors of the second paper that we have studied developed a CNN model for COVID-19 detection from chest radiography images under the assumption that in order to correctly diagnose and identify COVID-19, radiologists must first distinguish COVID-19 X-rays from normal chest X-rays, and then they must distinguish COVID-19 X-rays from other viral and bacterial infections in order to isolate and properly treat the patient. As a result, they decide to select CNN's layout to state one of the following predictions: Normal, COVID-19, viral and bacterial infections, and normal. The purpose of using these four examples is to assist radiologists in selecting COVID-19 patients for PCR testing and directing treatment toward infections with certain etiology. These requirements guided the creation of their straightforward CNN architecture [2].

C.

Experiments were carried out in the third paper that we studied to classify the input images into three categories: healthy, Pneumonia, and COVID. The images obtained from the dataset as input are pre-processed. The training dataset was divided into 5 folds, with 20 percent of the training data used for validation and the remaining 80 percent used to train

the model at each fold. This process is repeated five times, each time with a different set of image samples for training and validation. For 50 epochs, the validation and training accuracy of a Convolutions Neural Network (CNN) model were obtained. The results show that as the number of epochs increases, so does the training accuracy [3].

III. Methodology

This project was built in order to detect if a person is covid positive or negative. A dataset from Kaggle named QT dataset was used to train the machine learning model. This data set contained 10,000 samples where 4990 was positive and 5010 was negative. The target was based on three attributes, Oxygen, Pulse Rate and Temperature.

The data was preprocessed and cleaned up before use. It contained two rows with null (NaN) value which was removed. After preprocessing the data, the features and target columns were separated in 'X' and 'y' variables respectively. The visualization of the dataset is given below:

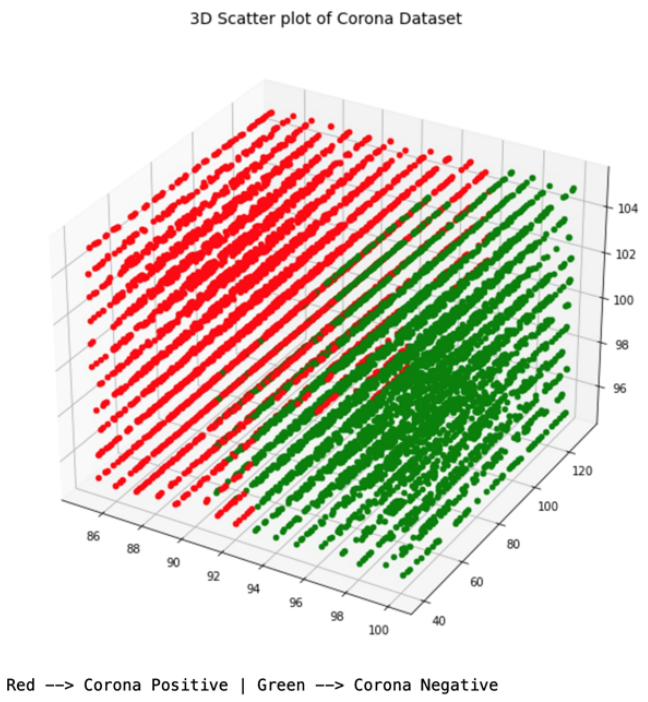

Fig. 1. 3D Scatter plot of COVID Dataset

Here,
X-axis represents Oxygen
Y-axis represents Pulse Rate
Z-axis represents Temperature

The dataset was split using k-fold cross validation, where the value of k was taken to be 10. 10,000 data folded in 10 gives 1000 data in each fold. So, for each iteration 1000 data was used for testing and the remaining 9000 data were used for training.

To build the model built-in SVM function from scikit-learn library was used which supports four different kernel functions, linear, polynomial, radial basis function (rbf) and sigmoid. All four kernels were used in addition to four different values of C parameter, 1, 2, 5 and 10 to evaluate the performance of different models.

Finally, the best model was extracted to get the highest accuracy possible from this dataset using svm classifier. It was found to be 94.36

IV. Table of Accuracy

Kernel	**C=1**	**C=2**	**C=5**	**C=10**
Linear	92.12%	92.12%	92.12%	92.12%
Polynomial	93.93%	93.95%	93.93%	93.93%
Radial basis function	94.26%	94.28%	94.28%	94.36%
Sigmoid	50.10%	50.10%	50.10%	50.10%

V. Significance of the Study

Globally, the COVID-19 epidemic has spread. Experts have categorized it into five categories: asymptomatic, mild, moderate, severe, and critical. It is a contagious disease that spreads quickly from one person in close contact with another. As of December 5, 2020, there were already over 66 million infections worldwide, with over 22 million actively unwell patients. Around the world, more than 1.5 million patients (or around 2.5% of all documented cases) perished. People might readily test their COVID status at home, which would be cost-free and alleviate the patient from report manipulation if there existed a system that could accurately predict whether the patient was COVID positive or negative. So, utilizing the features of temperature, pulse rate, and oxygen saturation, we are developing a project whose goal is to categorize COVID as positive or negative using a Support Vector Machine classifier.

VI. Discussion and Conclusion

After almost two years from the beginning of COVID-19 pandemic, it is clear that this virus is not going to stop spreading anytime soon. To minimize the testing requirement of testing kits, our system will help people to know whether they need to test for COVID or not. This will put a significant positive impact as the patients can check their possibility of catching COVID easily from their home. Most of the households keep thermometer and oxymeter nowadays. So checking the body temperature, oxygen saturation and pulse rate can be done at home. Our system will analyze these data and suggest if that person need to test for COVID or not. This will save time and money aswell as from the extra hassle that people face from testing for COVID with minor symptoms. We have worked on a dataset with 10000 samples and 3 attributes. But we are planning on expanding our scope and work on bigger datasets with more attributes in the near future.

References

[1] Sethy, P. K., Behera, S. K. (2020). Detection of coronavirus disease (covid-19) based on deep features.

[2] Iwendi, C., Mahboob, K., Khalid, Z., Javed, A. R., Rizwan, M., Ghosh, U. (2021). Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. Multimedia Systems, 1-15.

[3] Feng, K., He, F., Steinmann, J., Demirkiran, I. (2021, March). Deep-learning based approach to identify covid-19. In SoutheastCon 2021 (pp. 1-4). IEEE.