International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

Development Of Machine Learning Techniques To Differentiate COVID-19 Indications From Serious Diseases

Dr.Neelamadhab Padhy[1], Praveen Yadav[2], Dr.Syed Khasim[3], Dr.Shaik Shakeer Basha[4], Purshottam J. Assudani[5], S.Ranjana[6]

[1]Associate Professor, Department of Computer Science and Engineering, GIET University, Gunupur, Odisha 765022.
[2]Assistant Professor, Department of Computer Science and Engineering, University Institute of Technology, RGPV Bhopal Madhya Pradesh-462033.
[3]Professor, Department of Computer Science and Engineering, Dr.Samuel George Institute of Engineering & Technology, Markapur, Prakasam Dt, Andhra Pradesh, 523316.
[4]Assistant Professor, Department of Computer Science and Engineering, Avanthi Institute of Engineering and Technology, Gunthapally, Abdullahpurmet Mandal-501512, Hyderabad, Telangana.
[5]Assistant Professor, Department of Information Technology, Shri Ramdeobaba College of Engineering and Management, Gittikhadan, Katol Road, Nagpur, Maharashtra, India-440014.
[6]Assistant Professor, Department of Computer Science, Anna Adarsh College for Women, A1,II street,9th Main Road,Anna Nagar,Chennai -600040.

Abstract: Considering the identical signs of both covid-19 and influenza, most individuals are unable to distinguish between the two, which can result in a person's death. To control the death rate, several approaches are needed to categories the signs of covid-19 and other diseases. Severe sickness is more likely to hit the elderly and individuals with underlying medical conditions and diseases lung diseases and cancer. In the context of the present outbreak, identification of these diseases is limited to a few clinical studies like RT-PCR and CT-Scan of lung pictures to detect the covid-19. We will develop a method to solve the present issues experienced by people in the outbreak condition, as these examinations take a long time and are highly expensive. Researchers discovered that image processing, data mining, artificial intelligence, and pattern recognition are widely utilized approaches for solving these problems after doing a research study.

Keywords: Covid-19, Lung illness, cardiovascular disorder, diabetes, Data mining, machine learning

1. INTRODUCTION

Since 2019, the Corona Virus has had a significant impact, resulting in several injuries and deaths. The COVID-19 outbreak was proclaimed at the global level. COVID-19 began with flu-like indications, progressed to influenza, and eventually invaded the lungs. For the majority of COVID-19 sufferers, however, a heart attack was the cause of death. Within 2 to 14 days, the virus-infected individual began to exhibit indications. In summary, the COVID-19 revealed a wide range of illnesses. COVID-positive individuals have recovered fully and

have begun living a healthy life in some circumstances, although in the majority of cases, the condition has deteriorated and individuals have been admitted to a hospital.
Designers use several Machine Learning algorithms to discriminate COVID-19 illness apart from flu throughout the research design, and once we have such indicators, we add them to COVID-19 clinical symptoms and divide the dataset into three groups. Slight, Medium and Serious Clients are the three types of patients. The dataset expands as COVID-19 reveals signs of different diseases. For a larger dataset, machine learning techniques are an excellent choice. For separating COVID-19 from flu, designers utilized K-Nearest Neighbors, Linear Regression, and Decision Tree methods. For treatment outcomes, decision trees, Nave Bayes, and K-Means methods are employed.

2. LITERATURE SURVEY

The investigation was conducted out [1]using approaches such as real-time data query, visualization on their site, and then use of the queried data for Susceptible-Exposed-Infectious-Recovered (SEIR) synthesis processes. The author analyzed the information and divided it into pleasant and unpleasant feelings to fully understand the impact of the information on people's choices social - financial behavior. The five top authors talk about teamwork and personal strength in the face of the pandemic, whereas the top five negative stories discuss uncertainty and bad illness outcomes, such as fatalities. Ultimately, it was determined that the infectious illness is still unknown, implying that the author will only be able to make an accurate SEIR forecast once the epidemic is over.
A study[2] looks at theoretical equations for the distribution of a group organized by age. Because illness spreads via social contact and varies with age, it's critical to forecasting disease transmission based on changes in social structure. The COVID-19 was evaluated using a computational formula that combined contact pattern synthesis with empirical evidence. The model shows that a long duration of shutdowns followed by occasional relaxation reduces the number of instances.
It was given the term novel since it was the first time an animal Coronavirus mutation had been detected. Cases range from moderate to severe, with extreme cases resulting in significant medical problems or even death. The virus's incubation period in the human is 2-14 days 3, although the exact duration is unclear. COVID-19 infection is linked with several clinical symptoms, that are classified into Moderate and Severe pneumonia. The CT- SCAN findings are classified into 3 phases: Low, Medium, and Serious ARDS. The last two phases of the findings are extremely hard to define. Cleaning your hands is a typical measure for preventing the COVID-19 virus.
Deep learning algorithms were developed in this study paper [3] for estimating COVID-19 positive cases in India. Long short-term memory cells based on neural networks were utilized for prediction. Convolution LSTM produces the lowest outcomes, whereas bi-directional LSTM produces the highest performance. This study shows [4] the forecast of COVID-19, which has been done since traditional methods had demonstrated low accuracy for long-term forecast cases. The number of people in class S increases with time, which is commonly calculated to use simple equations [5].
The current study was based on readily viewable data of newly confirmed every day reported incidents from the tenth of January to the tenth February in this paper [6]. The key epidemiology metrics, such as the basic reproduction number and case recovery ratios, were estimated using the average scores of the main epidemiological indicators [7]. The magnitude

of the outbreak in Wuhan was assessed with cases imported from Wuhan to all places around the world [8].

3. METHODOLOGY

K-Nearest Neighbors is a Supervised Learning-based Learning Algorithm that is one of the most basic. It saves all of the information and group a data point depending on its resemblance to the available data. This implies that fresh data may be quickly categorized into a well-suited group using the K-NN method as it arises. Below are basic formulae for K-Nearest Neighbour categorization. Distance measure formula

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad \text{-------- (1)}$$

Manhattan Distance equation

$$d(x,y) = \sum_{i=1}^{m}|x_i - y_i| \quad \text{--------- (2)}$$

Minkowski Distance equation

$$\left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p} \quad \text{----------(3)}$$

A Decision Tree is a machine learning algorithm that may be applied to regression and classification issues. Input data contain data characteristics, the organization decision rules, so each leaf node provides the conclusion in this tree-structured classification.
Leaf nodes are the result of those selections, while Choice nodes are utilized to make the decision and contain numerous paths. The decision tree classifier's theoretical formulae are as continues to follow:

1. Information Gain= Entropy(S)- [(Weighted Avg) *Entropy (each feature)
2. Gini Index= $1- \sum_j P_j^2$

The supervised machine learning framework method logistic regression was used to estimate the likelihood of a target variable. The structure of the objective is binary in this case, which implies there are only two groups. There are two sorts of variables: mean and standard deviation. The response variable is binary, with data recorded as 1 (yes) or 0 (no) (no). Logistic regression predicts P(Y=1) as a variable of X Empirically, a linear regression model predicts P(Y=1) as a variable of X.

$$P(X) = P(Y=1|X)$$

Logistic regression equation

$$y = e^{\wedge}(b0 + b1*x) / (1 + e^{\wedge}(b0 + b1*x)) \quad \text{---------- (4)}$$

The Naive Bayes method is based on the Bayesian network, which would be a data mining and machine learning approach. The Naive Bayes designer's following equation can be seen here.

$$P(c|x) = p(x|c)\, p(c) / p(x) \quad \text{---------- (5)}$$

International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 02, 2021

K means is an incremental technique that attempts to divide a dataset into K pre-defined groups, each of which includes just one piece of data. It attempts to maintain based on inter datasets as comparable as feasible and maintaining groups as distinct as feasible. It operates on the idea of computing the sum of squared lengths between two points and ensuring that the cluster's center is as small as possible.

$$J(V) = \sum_{i=1}^{C} \sum_{j=1}^{c} (|Xi - Vj|)\text{^}2 \quad \text{---------- (6)}$$

4. RESULTS ANALYSIS

In this work, researchers tested several of the approaches for predicting covid-19 illness and stratifying the degree of covid-19 illness. We came to the following findings after completing the installation of the supplied machine learning models. The efficiency of the K-NN algorithm is 0.7611 to begin off. It has an accuracy of 0.886, a memory of 0.890, and an f1-measure of 0.89. Secondly, the Decision tree classification model has a 0.905 accuracy. Accuracy is 0.89, recall is 0.91, f1- measure is 0.990, and confidence is 93 for this model. Ultimately, the Regression Model has an effectiveness of 0.889. It has a 0.90 accuracy.

Table 1 Accuracy rate of different methods

Sl. No	Algorithm	Precision
1	K Classifier	0.886
2	Decision Tree Classifier	0.910
3	Logistic Regression	0.849
4	Naive Bayes	0.954
5	K- Means	0.840

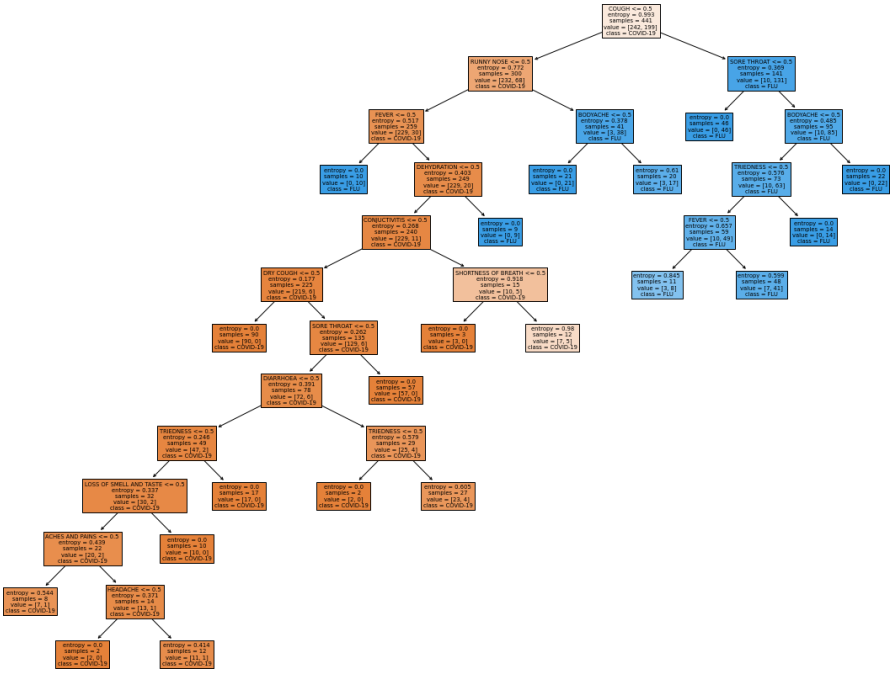

Figure 1: Decision Tree Classification

	Real Values	Predicted Values
0	Moderate	Moderate
1	Moderate	Moderate
2	Moderate	Moderate
3	Moderate	Moderate
4	Mild	Mild
...	...	...
257	Mild	Mild
258	Severe	Severe
259	Mild	Mild
260	Mild	Mild
261	Mild	Mild

Figure 2: Real observations and identified observations

Figure 1 and 2 shows the decision classification trees and Real observations and identified observations form the different method. Table 1 describes the accuracy rate in contrast to a different method. It's impossible to determine the similarity between COVID-19 and influenza since the indications of COVID-19 are possible to specify. In the instance of data segmentation, researchers arrived at the following findings after completing the installation of the supplied machine learning techniques. To begin, the Decision tree model's efficiency is 0.994. It has an accuracy of 0.993, a recall of 0.993, and an f1-measure of 0.892. Secondly, the Naive Bayes classification algorithm has an efficiency of 0.894. Ultimately, the K Means analysis has an efficiency of 0.880.

5. CONCLUSION

According to the modeling levels of accuracy achieved for the data using machine learning algorithms, the decision tree algorithm produces the highest results (0.925), followed by the Regression model and KNN classifier, which are the weakest point. Whenever it comes to information segmentation, the feature selection and Naive Bayes are nearly identical, with the feature selection yielding the greatest results (0.934 accuracies), whereas the K Means method yields the most reliable data. As a result, researchers believe that this project on COVID-19 forecasting and severity distinction is complete and that it might have been utilized to save physicians time & expense when diagnosing the illness and determining which phase the individual.

6. REFERENCES

[1] Fairoza Amira Binti Hamzaha, Corona Tracker Community Research Group , COVID 19 site of WHO , 2020.
[2] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China
[3] Joseph T Wu, Kathy Leung, Gabriel M Leung, Nowcasting and forecasting the

potential domestic and international spread of the nCoV-19 outbreak originating in Wuhan, China: a modelling study, 2020.

[4] Organization WH. WHO Statement Regarding Cluster of Pneumonia Cases in Wuhan, China; 2020. Available from: https://www.who.int/china/news/detail/09-01-2020-whostatement-regarding-cluster-of-pneumonia-cases-in-wuhan-china.

[5] Prabira Kumar Sethy, Detection of coronavirus Disease (COVID-19) based on Deep Features and Support Vector Machine, 2020

[6] S. Towers and Z. Feng, "Social contact patterns and control strategies for influenza in the elderly".

[7] The Johns Hopkins Center for Health Security. Daily updates on the emerging novel coronavirus from the Johns Hopkins Center for Health Security. February 9, 2020; 2020. Available from: https://hub.jhu.edu/2020/01/23/ coronavirus-outbreak-mapping-tool-649- em1-art1-dtd-health/

[8] Varsha Kachroo , Novel Coronavirus (COVID-19) in India: Current Scenario ,International Journal of Research and Review , vol 7;Issue: 3,March 2020.