How to Cite:
Vedaraj, M., Saravanan, K., Srinivasan, V. P., Balachander, K., & Jaithunbi, A. K. (2022). Prediction of COVID 19 using marching learning techniques. *International Journal of Health Sciences*, *6*(S3), 9467–9474. https://doi.org/10.53730/ijhs.v6nS3.8315

Prediction of COVID 19 using marching learning techniques

M. Vedaraj
Assistant professor, Department of CSE, R.M.D. Engineering College

K. Saravanan
Associate professor, Department of IT, R.M.D. Engineering College

V. Prasanna Srinivasan
Associate professor, Department of IT, R.M.D. Engineering College

K. Balachander
Associate professor, Department of CSE, Velammal Institute of Technology

A. K. Jaithunbi
Assistant professor, Department of CSE, R.M.D. Engineering College

***Abstract*---**Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. Most people infected with the virus will experience mild to moderate respiratory illness and recover without requiring special treatment. However, some will become seriously ill and require medical attention. Older people and those with underlying medical conditions like cardiovascular disease, diabetes, chronic respiratory disease, or cancer are more likely to develop serious illness. Supervised machine learning models for COVID-19 infection were developed in this work with learning algorithms which include support vector machine, naive Bayes, random Forest, GNB using epidemiology labeled dataset for positive and negative COVID-19 cases of Mexico. The correlation coefficient analysis between various dependent and independent features was carried out to determine a strength relationship between each dependent feature and independent feature of the dataset prior to developing the models. The 80% of the training dataset were used for training the models while the remaining 20% were used for testing the models. The result of the performance evaluation of the models showed that GNB prediction model has the highest accuracy of 98% compared to other existing ML techniques.

***Keywords*---**COVID, SARS, artificial neural network (ann), dataset.

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
Manuscript submitted: 27 March 2022, Manuscript revised: 9 April 2022, Accepted for publication: 18 May 2022

Introduction

Coronavirus disease 2019 (COVID-19) is a contagious disease caused by a virus, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first known case was identified in Wuhan, China, in December 2019.[1] The disease spread worldwide, leading to the COVID-19 pandemic[2]. Symptoms of COVID-19 are variable, but often include fever,[3] cough, headache,[14] fatigue, breathing difficulties, loss of smell, and loss of taste.[5][6][7] Symptoms may begin one to fourteen days after exposure to the virus. At least a third of people who are infected do not develop noticeable symptoms.[8] Of those people who develop symptoms noticeable enough to be classed as patients, most (81%) develop mild to moderate symptoms (up to mild pneumonia), while 14% develop severe symptoms (dyspnea, hypoxia, or more than 50% lung involvement on imaging), and 5% develop critical symptoms (respiratory failure, shock, or multiorgan dysfunction).[9] Older people are at a higher risk of developing severe symptoms. Some people continue to experience a range of effects (long COVID) for months after recovery, and damage to organs has been observed.[10] Multi-year studies are underway to further investigate the long-term effects of the disease. The confirmed and death cases for top 15 countries shown in figure.

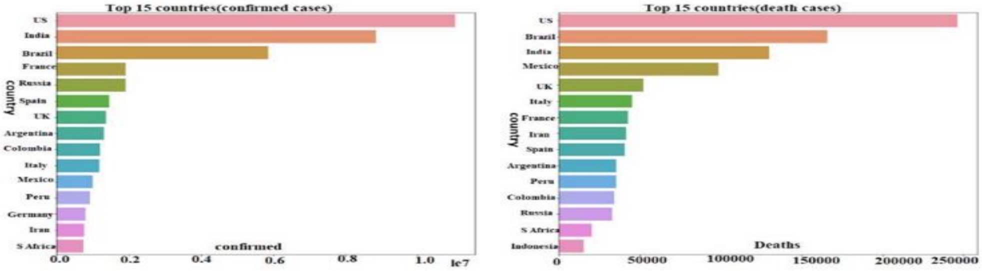


Figure 1. The confirmed and death cases for top 15 countries

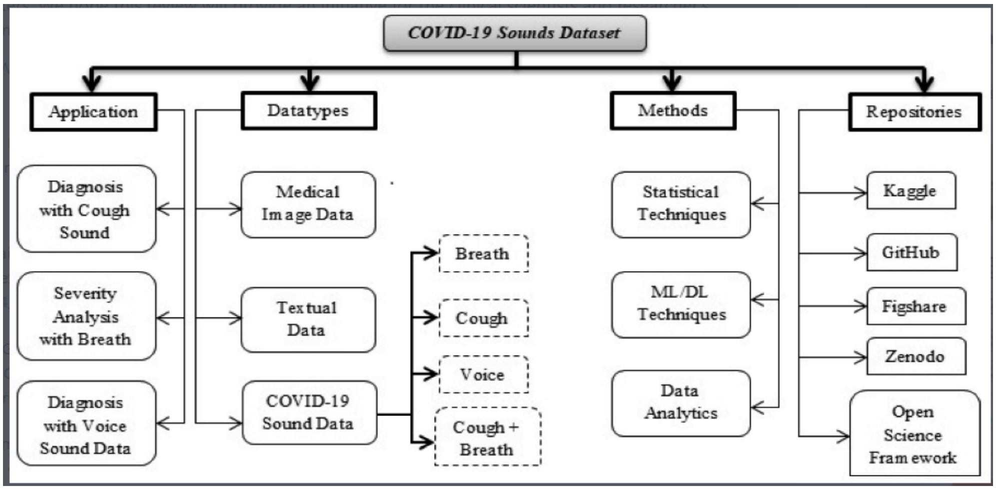


Figure 2. *COVID-19 sounds taxonomy*

Related Works

Early detection and diagnosis using AI techniques help to prevent the spread and to combat the COVID-19 pandemic using different data such as CT scans, X-ray, clinical data, and blood sample data. Sun et al. [12] developed a prediction model using the support vector machine (SVM) to predict the severe cases of COVID-19 patients. In the study, they used the clinical and laboratory features that are significantly associated with these cases. Using 336 cases of COVID-19 patients, 26 severe/critical cases and 310 noncritical, they found that the main features to discriminate the mild and severe cases are age, growth hormone secretagogues (GHSs), immune feature cluster of differentiation 3 (CD3) percentage, and total protein. They found that the proposed model was effective and robust in predicting patients in severe conditions with up to 0.775 accuracy.

Another research conducted by Yao et al. [13] also applied the SVM model to classify the COVID-19 patients according to the severity of the symptoms. 'ey applied SVM for the binary class label on a total of 137 records including urine and blood test results and combining both severely ill patients and patients with mild symptoms. 'e results showed that around 32 factors have high correlations with severe COVID-19, with an accuracy of 0.815. It is worth mentioning that, amongst all factors, age and gender had mostly affected the classification of cases between severe and mild. Patients aged around 65 had more severe cases than others. Moreover, male patients were at a higher risk of developing severe COVID-19 symptoms. In terms of the urine and blood test samples, blood test result features show more significant differences between severe and mild cases than urine test result features.

Hu et al. [14] used the logistic regression (LR) model to identify the COVID-19 patients' severity. They used a dataset containing demographic and clinical data for 115 COVID-19 patients under the nonsevere condition and 68 COVID-19 patients under the severe condition. Four features have been selected as the most significant features to discriminate the mild and severe cases: age, high-sensitivity C-reactive protein level, lymphocyte count, and d-dimer level. 'is model was evaluated, and the results showed that the prediction was effective with area under the receiver operating characteristic (AUROC) of 0.881, sensitivity of 0.839, and specificity of 0.794, respectively.

Proposed Model

Fig.1 Shows covid 19 prediction model. The proposed Covid 19 prediction model consist of five different steps such as data collection, data pre-processing, future extraction, classification, and performance evaluation. we prepared three classes of chest X-ray images dataset COVID-19 patients, normal people, and pneumonia patients. It contains 520 images (120 COVID-19 images, 200 pneumonia images and 200 healthy images). COVID-19 is a new disease, so, the number of COVID-19 chest x-ray images is limited. The datasets were collected from Kaggle website. Then, the datasets are divided into 80% for training and 20% to test the classifiers. Fig. 2 shows an example of chest X-ray image Datasets. The dataset is already preprocessed by and resizing it to 64×128 pixels so that it is ready for training and testing our model. After pre-processing, features are extracted from

images by using Histogram of oriented gradients. The basic advantages are describing the shape and contour properties of an image. Support vector machine, naive Bayes, random Forest, GNB are used for disease classification.

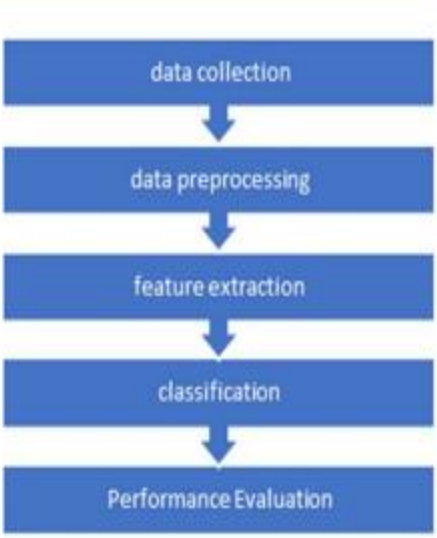


Figure 3. Covid 19 Prediction Model

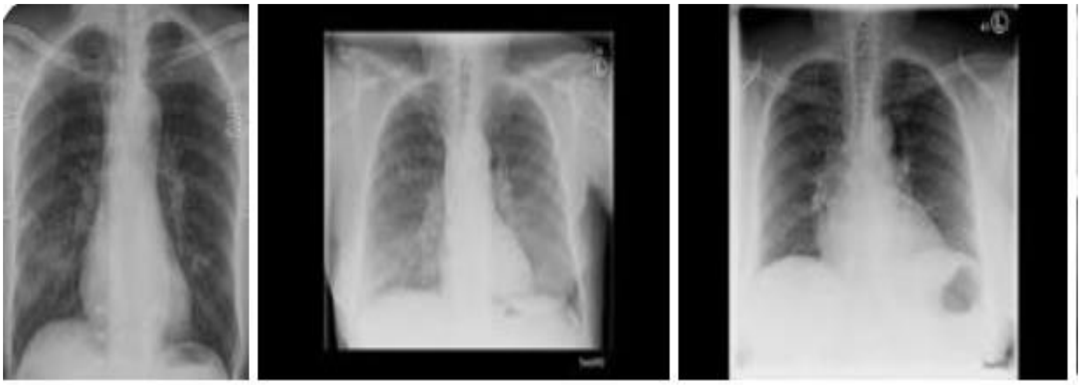
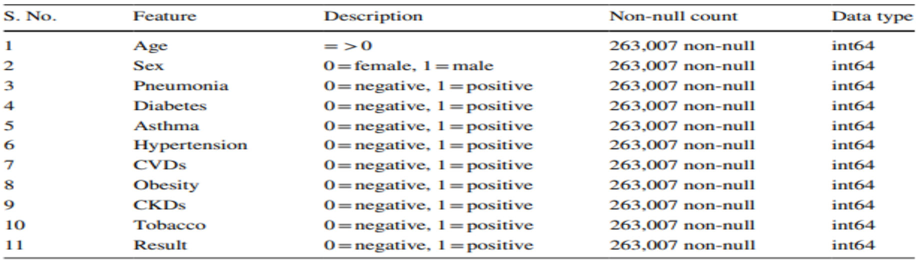


Figure 4. Covid 19 X ray Images

Dataset Description



S. No.	Feature	Description	Non-null count	Data type
1	Age	= > 0	263,007 non-null	int64
2	Sex	0 = female, 1 = male	263,007 non-null	int64
3	Pneumonia	0 = negative, 1 = positive	263,007 non-null	int64
4	Diabetes	0 = negative, 1 = positive	263,007 non-null	int64
5	Asthma	0 = negative, 1 = positive	263,007 non-null	int64
6	Hypertension	0 = negative, 1 = positive	263,007 non-null	int64
7	CVDs	0 = negative, 1 = positive	263,007 non-null	int64
8	Obesity	0 = negative, 1 = positive	263,007 non-null	int64
9	CKDs	0 = negative, 1 = positive	263,007 non-null	int64
10	Tobacco	0 = negative, 1 = positive	263,007 non-null	int64
11	Result	0 = negative, 1 = positive	263,007 non-null	int64

Experimental Setup

The supervised machine learning algorithms were executed using a python programming language in window operating system environment deployed HP Branded computer system (Laptop), Corei5 with 8 GB of Ram and 2.8 GHz processor speed. All the necessary libraries were installed on python notebook and used for the data analysis including correlation analysis and development of the models.

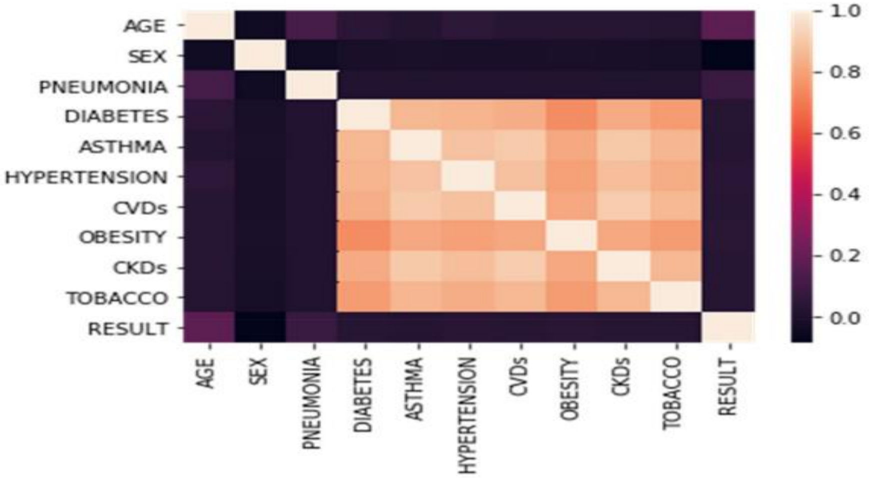


Figure 5. Scatterplot correlation coefficient of the feature of the dataset

Confusion Matrix

A confusion matrix is an extremely useful tool to observe in which way the model is wrong. It is a matrix that compares the number of predictions for each class that are correct and those that are incorrect. In a confusion matrix, there are 4 numbers to pay attention to.

- **True positives:** The number of positive observations the model correctly predicted as positive.
- **False-positive:** The number of negative observations the model incorrectly predicted as positive.
- **True negative:** The number of negative observations the model correctly predicted as negative.
- **False-negative:** The number of positive observations the model incorrectly predicted as negative.

The performance of the proposed Covid 19 prediction model can be evaluated using following performance metrics such as accuracy, precision, Recall, F1 Score, and Specificity.

Accuracy=(TP+TN)/(TP+FP+FN+TN) ....(1)
Precision=TP/(TP+FP) ....(2)
Recall = TP/(TP+FN) ....(3)

F1 Score = 2*(Recall * Precision) / (Recall + Precision) ....(4)
Specificity = TN/(TN+FP) ....(5)

Table 1
Prediction of Covid-19

ML Algorithm	Accuracy	Precision	Specificity	F1 Score
SVM	94	85	87	87
NB	95.5	86	89	89
RF	97	87	86	89.9
GNB	98	92	91	90

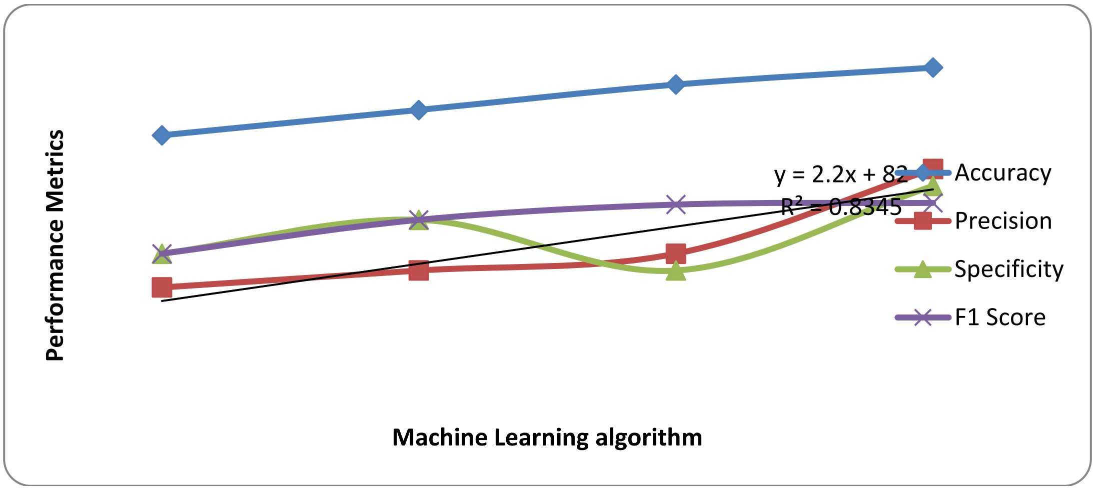


Figure 6. Covid 19 Prediction

Conclusion

Coronaviruses are important human and animal pathogens. At the end of 2019, a novel coronavirus was identified as the cause of a cluster of pneumonia cases in Wuhan, a city in the Hubei Province of China. It rapidly spread, resulting in an epidemic throughout China, followed by a global pandemic. In February 2020, the World Health Organization designated the disease COVID-19, which stands for coronavirus disease 2019. The virus that causes COVID-19 is designated severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2); previously, it was referred to as 2019-nCoV. Supervised ML models for COVID-19 infection were developed in this work with a support vector machine, naive Bayes, random Forest, GNB ML algorithms using an epidemiology labeled dataset of positive and negative COVID-19 cases in Mexico. The models were trained with 80% training data and tested with the remaining 20% of the data. The model developed with decision tree happened to be the best model among all models developed in terms of accuracy with 98%.

References

1. Page J, Hinshaw D, McKay B (26 February 2021). "In Hunt for Covid-19 Origin, Patient Zero Points to Second Wuhan Market – The man with the first confirmed infection of the new coronavirus told the WHO team that his parents had shopped there". The Wall Street Journal. Retrieved 27 February 2021.
2. Zimmer C (26 February 2021). "The Secret Life of a Coronavirus – An oily, 100-nanometer-wide bubble of genes has killed more than two million people and reshaped the world. Scientists don't quite know what to make of it". The New York Times. ISSN 0362-4331. Archived from the original on 28 December 2021. Retrieved 28 February 2021.
3. Islam MA (April 2021). "Prevalence and characteristics of fever in adult and paediatric patients with coronavirus disease 2019 (COVID-19): A systematic review and meta-analysis of 17515 patients". PLOS ONE. 16 (4): e0249788. Bibcode:2021PLoSO..1649788I. doi:10.1371/journal.pone.0249788. PMC 8023501. PMID 33822812.
4. Islam MA (November 2020). "Prevalence of Headache in Patients With Coronavirus Disease 2019 (COVID-19): A Systematic Review and Meta-Analysis of 14,275 Patients". Frontiers in Neurology. 11: 562634. doi:10.3389/fneur.2020.562634. PMC 7728918. PMID 33329305.
5. Saniasiaya J, Islam MA (April 2021). "Prevalence of Olfactory Dysfunction in Coronavirus Disease 2019 (COVID-19): A Meta-analysis of 27,492 Patients". The Laryngoscope. 131 (4): 865–878. doi:10.1002/lary.29286. ISSN 0023-852X. PMC 7753439. PMID 33219539.
6. Saniasiaya J, Islam MA (November 2020). "Prevalence and Characteristics of Taste Disorders in Cases of COVID-19: A Meta-analysis of 29,349 Patients". Otolaryngology–Head and Neck Surgery. 165 (1): 33–42. doi:10.1177/0194599820981018. PMID 33320033. S2CID 229174644.
7. Agyeman AA, Chin KL, Landersdorfer CB, Liew D, Ofori-Asenso R (August 2020). "Smell and Taste Dysfunction in Patients With COVID-19: A Systematic Review and Meta-analysis". Mayo Clin. Proc. 95 (8): 1621–1631. doi:10.1016/j.mayocp.2020.05.030. PMC 7275152. PMID 32753137.
8. Oran DP, Topol EJ (January 2021). "The Proportion of SARS-CoV-2 Infections That Are Asymptomatic : A Systematic Review". Annals of Internal Medicine. 174 (5): M20-6976. doi:10.7326/M20-6976. PMC 7839426. PMID 33481642.
9. "Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19)". U.S. Centers for Disease Control and Prevention (CDC). 6 April 2020. Archived from the original on 2 March 2020. Retrieved 19 April 2020.
10. Jump up to:$^{a\ b}$ CDC (11 February 2020). "Post-COVID Conditions". U.S. Centers for Disease Control and Prevention (CDC). Retrieved 12 July 2021.
11. A Vasantharaj, Pacha Shoba Rani, Sirajul Huque, KS Raghuram, R Ganeshkumar, Sebahadin Nasir Shafi "Automated Brain Imaging Diagnosis and Classification Model Using Rat Swarm Optimization with Deep Learning Based Capsule Network" International Journal of Image and Graphics https://doi.org/10.1142/S0219467822400010

12. L. Sun, F. Song, N. Shi et al., “Combination of four clinical indicators predicts the severe/critical symptom of patients infected COVID-19,” Journal of Clinical Virology, vol. 128, p. 104431, 2020.
13. H. Yao, N. Zhang, R. Zhang et al., “Severity detection for the coronavirus disease 2019 (COVID-19) patients using a machine learning model based on the blood and urine tests,” Frontiers in Cell and Developmental Biology, vol. 8, pp. 1–10, 2020.
14. Vamsidhar Enireddy, R P Shobha Rani ,Anitha, Sugumari Vallinayagam, T Maridurai, T Sathish, E Balakrishnan "Prediction of human diseases using optimized clustering techniques" 2021Materials Today: ProceedingsVolume 46Pages 4258-4264 PublisherElsevier
15. Hu, Z. Liu, Y. Jiang et al., “Early prediction of mortality risk among patients with severe COVID-19, using machine learning,” International Journal of Epidemiology, vol. 49, no. 6, pp. 1918–1929, 2020.