A Novel Approach to Apply Different Algorithms to Predict Covid-19 Disease

[1] Utlapalli. Mahesh, [2] Bonela. Syam Jason, [3] S. Nithya Tanvi Nishitha
[4] Jonnalagadda. Surya Kiran

[1,2,3] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P., India-522502.

[1] umahesh987654321@gmail.com , [2] Nithyatanvi9999@gmail.com, [3] syamjasonb777@gmail.com, [4] kiransurya93@kluniversity.in

Abstract: The Covid-19 episode happened in Wuhan in December 2019 and spread wherever on the planet. The Covid-19 transferable infection has an immunization and medication for its treatment. The premier significant components in lessening the spread of the infection are washing hands, utilizing a veil, and diminishing social distance. Today furthermore to clinical investigations, PC-supported examinations likewise are generally regulated for the Covid-19 episode. Computer-based intelligence strategies are effectively applied during this examination. In this investigation, we used different algorithms for the prediction and analysis of Covid-19 everyday cases. Because of the examination, the number of everyday cases was effectively assessed with these various kinds of algorithms.

Keywords: COVID-19, transferable infection, clinical investigations, computer-based intelligence, prediction.

I. INTRODUCTION

The essential case was distinguished in Wuhan, China, in December 2019, passing from bats to people. The infection spread everywhere in the world in a brief period and transformed into a lethal illness and fell the wellbeing frameworks of numerous nations in this world. Coronavirus is an extremely intense respiratory condition and infection movement can have deadly results. The clearest side effects of the illness; It is known as fever, dry hack, sore throat, migraine, shortcoming, muscle torment, looseness of the bowels, and windedness. In further developed cases, it causes serious pneumonia, irritating the lungs because of oxygen distinction and different organ disappointment. Particularly this infection has significantly more hazardous impacts for those with ongoing sicknesses, powerless obstruction or insusceptible framework, smokers, and the older. Man-made brainpower techniques [20] are utilized effectively in the arrangement of numerous issues. In the COVID-19 [11] scourge, numerous examinations have been completed utilizing computerized reasoning strategies. A portion of those are analysis utilizing radiology pictures, following sickness, assessing the patient's wellbeing result, early identification, and conclusion of disease, drug revelation, observing therapy.

II. LITERATURE SURVEY

A. Title: Covid Symptom Severity Using Decision Tree

Authors: Naim Rochmawati,Hanik Badriyah Hidayati, Yuni Yamasari ,Wiyli Yustanti ,Lusia Rakhmawati, Hapsari P. A. Tjahyaningtijas ,Yeni Anistyasari

In this paper, the main aim is to present a simple key whether it is symptoms of covid-19 [21] or normal cough. A clinical symptom dataset has been taken which is used for classifying the symptoms using the Decision Tree algorithm. In this research, the decision tree that has been used is hoeffding and j48. As we know that decision is one of the widely used algorithms for classification methods [12] as it is easy to simply by humans. the main aim is to cover the concept of concerting data into decision trees or decision rules. As a result, j48 has shown more efficiency than hoeffding tree in terms of recall, accuracy, precision. In further, this research work can be also used for the same dataset and different preprocessing. [1]

B. Title: COVIDGR Dataset and COVID-SDNet Methodology for Predicting COVID-19 Based on Chest X-Ray Images

Authors: S. Tabik , A. Gómez-Ríos, J. L. Martín-Rodríguez, I. Sevillano-García, M. Rey-Area, D. Charte, E. Guirado, J. L. Suárez, J. Luengo, M. A. Valero-González, P. García-Villanova, Olmedo-Sánchez, and F. Herrera

Current pandemic covid-19 is detected by taking the help of CT-Scans, X-Rays, CT scans. As we know that RT-PCR tests are mostly not available in medical centres and pharmacy stores hence CXR images have been used alternatively which is the most efficient, less cost, less time tool for assisting in taking the decisions. Here DL Neural networks have been used and shown potential for building [7] the covid-19 triage systems and to identify the covid-19 patients mostly detected patient with less severity. this research around two points i.e., clarified the hypersensitivity which is achieved in covid-19 classifications, well collaborate with most hospitals such as Spain, Granada. The dataset is stabilized by using the covidgr-1.0 that has every level. To use it in the future, it can enhance with covidgr-1.0 with a CXR image taken from multiple hospitals. [2]

C. *Title: A Novel Parametric Model for the Prediction and Analysis of the COVID-19 Casualties*

Authors: ONDER TUTSOY , ŞULE ÇOLAK , ADEM POLAT , AND KEMAL BALIKCI

This papers mainly cover the concepts of multi-dimensional, PSID, strongly coupled model. The research on mathematical analytics has shown the results on using region, instability region, stable region. the information has been analysed and lifting the restriction shows that the covid-19 [14] is moderately unstable which in the future will increase if precautions are not taken. the model evaluates many dead and contaminated will meet zero in three hundred days, but the no. of uncertainty discrete requires 1000 days to decrease under the present conditions. Allowing the model was used to evaluate the corresponding covid-19 casualties. So, in further, it can be used more efficiently without taking the uncertainty, death causalities, infected into consideration. we must also avoid pharmacological policies, non-pharmacological policies, and intubation casualties. [3]

D. *Title: Prediction of Covid-19 pandemic based on Regression*

Authors: Ashish U Mandayam,Rakshith.A.C,Siddesha S, S K Niranjan

As we are in the COVID-19 pandemic, it would be useful to anticipate the forthcoming number of positive cases for better assessment and control. We have chipped away at two managed learning models to gauge the future utilizing the time-arrangement dataset of COVID-19. To examine the arranging of the forecast, the correlation between Linear Regression [13] and Support Vector Regression is performed. We chose to utilize these two models as the information was almost straight. at the point when we make progress toward the improvement of our outcomes, we came to see that SVR contrasted with Linear Regression with time-arrangement information, the Linear Regression calculation performs better since the informational collection utilized here is likewise straight and the SVR can't deal with enormous direct datasets quite well. [4]

E. *Title: Covid-19 Disease Simulation using GAMA platform*

Authors: *Tran Quy Ban, Phan Lac Duong, Nguyen Hoang Son, Tran Van Dinh*

As the Covid-19 pandemic is advancing, information is gathered from different sources. From one perspective, specialists permit to make educated acclimations to the current and arranged intercessions and uncover them. Then again, a dire requirement for devices and strategies that empower speedy investigation, getting, correlation, and anticipating of the viability of the reactions against COVID-19 across various networks and settings. In this point of view, computational displaying shows up as priceless influence as it permits us to investigate in silico an assortment of mediation procedures before the likely period of field execution. For future work, we intend to advance the reproduction with more functionalities. We additionally need to think about among GAMA and other recreation structures to have an appropriate view of the re-enactment system as a rule. [5]

F. *Title: A Novel Deep Learning Approach for Classification of COVID-19 Images*

Authors: Malaya Kumar Nath,Aniruddha Kanhe, Madhusudhan Mishra

In this paper, the capability of man-made brainpower [19] has been researched to build up a profound neural organization model for fast, exact, and compelling COVID-19 location from the CT and X-beam pictures. The proposed strategy gives a strong profound learning method to parallel (COVID versus NON-COVID) and multi-class (COVID versus NON-COVID versus Pneumonia) grouping from X-ray and CT pictures. A 24-layer CNN network has been proposed for the grouping. It achieves an exactness of 99.68% and 71.81% on X-beam and CT pictures, individually. For both datasets, the Sgdm streamlining agent has been utilized with a learning rate of 0.001. As in the future to enhance this we will make These simulations to work on almost every system. [6]

G. *Title: Fuzzy Rule-Based System for Predicting Daily Case in COVID-19 Outbreak*

Authors: Pinar Cihan

As we know that covid-19 is a new virus and hasn't proven clinically or had a drug/vaccine for the treatment. A most prime factor to reduce the increase of virus is to maintain Social distance, always wearing up masks (N95 preferably), sanitizing hands regularly. As we know today's computer-aided technology is commonly used for covid-19. In that AI [18] is one of the best successful applications in epidemic studies. In this paper, fuzzy concepts have been used such as FRBs which is used for prognostication. The number of covid cases increasing daily. As result the number of covid cases has been successfully estimated i.e., R2 = 0.96, MAE = 186, RMSE = 254. [8]

H. *Title: COVID-19 Pandemic Prediction using Time Series Forecasting Models*

Authors: Naresh Kumar,Seba Susan

In this paper, the aim is to classify the overlong affected covid cases, growth of infection is evaluated to reduce and take the precautions within the health care service ignorer to avoid the deaths. As we know that predicting the growth/speed of covid-19 is the most backbreaking real-world problem hence here the day level information has been taken from the top 10 most affected covid-19 countries such as Spain, us, Italy, Germany, Iran, Russia, UK, Turkey, India. The data has been taken from January 22, 2020, to May 20, 2020. Then according to the data, the model of evaluation of covid-19 is predicted by models such ARIMA and prophet time series forecasting [16]. The AME, RMSE, RRSE, and MABE also has been calculated. From the output understand that the ARIMA model is suitable for prognosticating [17] covid-19 similarity. [9]

I. Title: A Comparative Study of Machine Learning Models for COVID-19 prediction in India

Authors: Vartika Bhadana, Anand Singh Jalal, Pooja Pathak

The paper shows the capacity of ML models to evaluate the proportion of approaching COVID-19-affected patients that is as of now saw as a certified risk to human progress. During this paper, we have administered comparative research on five machine learning [15] standard models like LR, DT, LASSO, RF, and SVM to forecast the frightening variables of COVID-19. Each of the models makes three sorts of forecasts, i.e., the total active cases, the entire deaths, and the total recovery in the coming 5 days. The findings provided by the paper recommend that the utilization of those techniques for the present COVID-19 pandemic scenario may be a promising strategy. For greater accuracy, we have used a six-degree polynomial. Experiment outputs illustrate that poly LR, and poly LASSO gives the best results followed by LR, LASSO, random forest, and decision tree. SVM shows a poor outcome in the prediction of COVID-19. [10]

III. BLOCK DIAGRAM

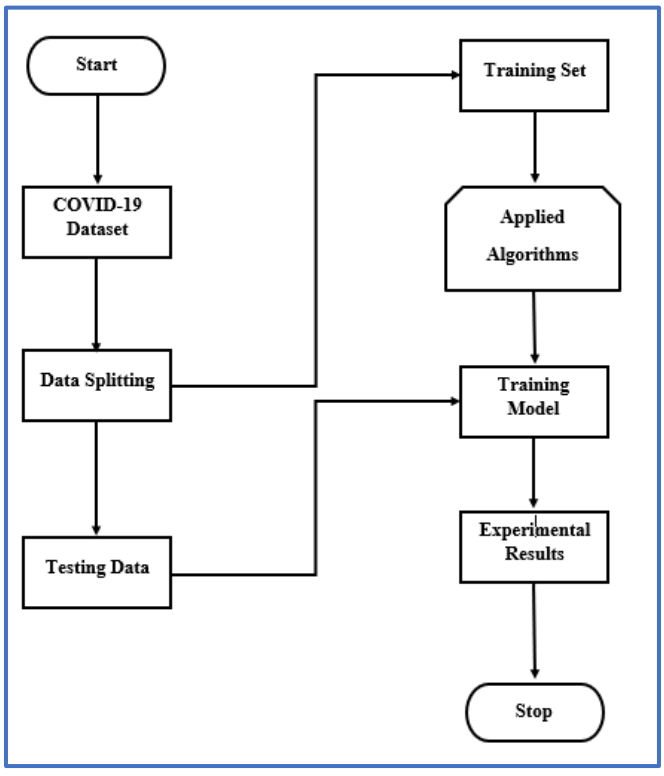

Figure 1: Block Diagram of the Proposed System

Algorithm of the Proposed System:

Step 1: Start

Step 2: Required dataset collection using Kaggle and GitHub repositories.

Step 3: Pre-processing the dataset to remove the noisy data.

Step 4: Splitting the data into training and testing data.

Step 5: Creation and training the models using the training data.

Step 6: Calculating accuracy of the models using the test data.

Step 7: Comparison of experimentation results.

Step 8: Stop

IV. IMPLEMENTATION

A. Python: Tools, Libraries

Python 3 or Above version is used for our work and have used the many libraries for different algorithms.

B. Libraries

Pandas: In order to perform Data analysis and preprocessing of data in python we have a built in "pandas" library, which provides various inbuilt functions that performs operations for manipulating numerical tables and time series.

Matplotlib: To perform Data Visualization in python, we use Matplotlib which is an inbuilt library that performs 2D visualization of data. Matplotlib generate high quality bar charts, scatter plots, histograms, and many more.

NumPy: NumPy is a python library that works on Numerical data, which contains a multi-dimensional arrays and matrix data structures. It is mainly used to perform mathematical operations on Numerical data during analysis.

V. EXPERIMENTAL SET UP AND RESULTS

A. Dataset

We have collected the data from the Kaggle website and some other GitHub repositories.

B. Procedure

Step 1: Firstly, the collected dataset is taken and then we undergone some pre-processing techniques to clean the raw data.

Step 2: Then the dataset is splitted into training data and testing data.

Step 3: Then we need to apply different types of machine learning algorithms on the training data.

Step 4: Then we need to find the accuracy of the testing data with the help of those models.

Step 5: Lastly, we need to compare the experimental results obtained by those models.

C. Results

```
print("accuracy of training dataset is{:.2f}".format(classifier.score(x_train,y_train)))
print("accuracy of test dataset is {:.2f}".format(classifier.score(x_test,y_test)))

accuracy of training dataset is0.96
accuracy of test dataset is 0.88
```

Figure 2 Description: The above figure depicts the accuracy of Decision Tree.

```
In [113]: print("Error rate is",1- accuracy_score(tar_test, predictions, normalize = True))

          Error rate is 0.09392265193370164
```

Figure 3 Description: The above figure depicts the Error Rate of Decision Tree.

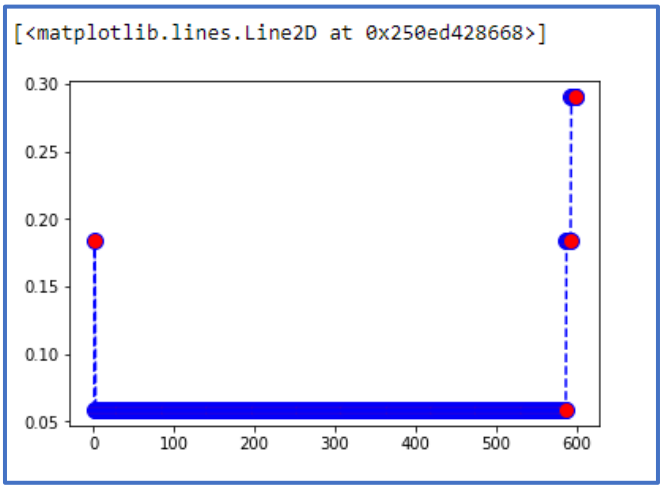

Figure 4 Description: Fig 3: The above figure represents the Error Rate vs K value of the K Nearest Neighbor algorithm.

```
print(accuracy_score(y_test,pred))

0.9411764705882353
```

Figure 5 Description: The above figure represents the Accuracy Rate of the KNN algorithm.

```
print("Error rate is",1- accuracy_score(y_test, pred, normalize = Tru

Error rate is 0.08018433179723505
```

Figure 6 Description: The above figure depicts the Error Rate of the KNN Algorithm.

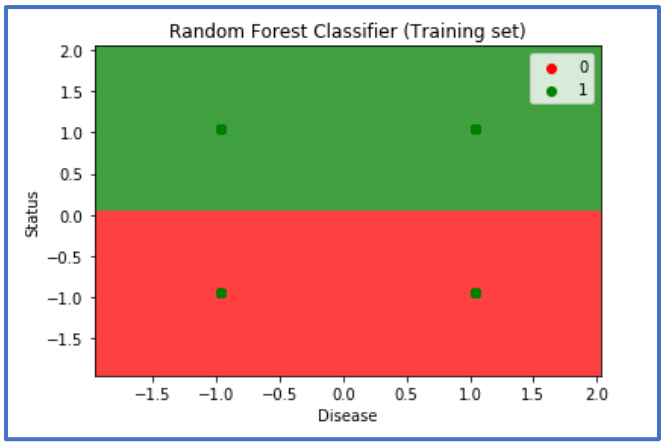

Figure 7 Description: The above figure visualizes the Training set results of the Random Forest algorithm.

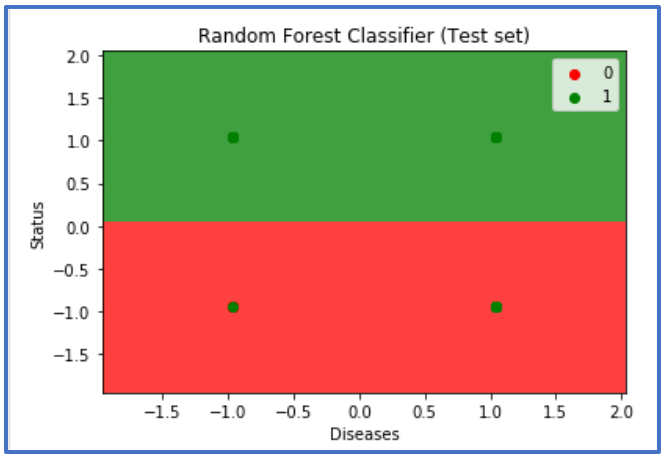

Figure 8 Description: The above figure visualizes the Testing set results of the Random Forest algorithm.

```
from sklearn import metrics

print('Mean Absolute Error:', metrics.mean_absolute_error(Y_Test, Y_Pred))
print('Mean Squared Error:', metrics.mean_squared_error(Y_Test, Y_Pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(Y_Test, Y_Pred)))

Mean Absolute Error: 0.0752212389380531
Mean Squared Error: 0.0752212389380531
Root Mean Squared Error: 0.27426490650109264
```

Figure 9 Description: The above figure depicts the MAE, MSE, RMSE values of the RF Algorithm.

```
print(accuracy_score(Y_Test, Y_Pred))

0.9247787610619469
```

Figure 10 Description: The above figure depicts the Accuracy Rate of the Random Forest Algorithm.

```
print("Error rate is",1- accuracy_score(Y_Test, Y_Pred, normalize = True))

Error rate is 0.08517699115044253
```

Figure 11 Description: The above figure depicts the Error Rate of the Random Forest Algorithm.

```
accuracy_score(y_test,y_pred)
0.925414364640884
```

Figure 12 Description: The above figure depicts the Accuracy Rate of the Naïve Bayes Algorithm.

```
print("Error rate is",1- accuracy_score(y_test, y_pred, normalize = True))
Error rate is 0.074585635359116
```

Figure 13 Description: The above figure depicts the Error Rate of the Naïve Bayes Algorithm.

```
print("accuracy of training dataset is{:.2f}".format(classifier.score(x_train,y_train)))
print("accuracy of test dataset is {:.2f}".format(classifier.score(x_test,y_test)))
accuracy of training dataset is0.92
accuracy of test dataset is 0.92
```

Figure 14 Description: The above figure depicts the Accuracy Rate of the SVM Algorithm.

```
print("Error rate is",1- accuracy_score(y_test, y_predict, normalize = True))
Error rate is 0.0755760368663595
```

Figure 15 Description: The above figure depicts the Error Rate of the Naïve Bayes Algorithm.

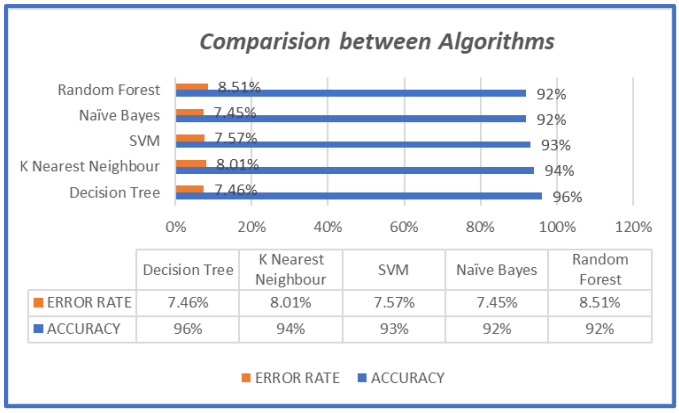

Figure 16 Description: The above figure illustrates the comparison between algorithms with the help of accuracy and error rate used for the Covid-19 prediction during experimentation.

VI. CONCLUSION

In this paper, we predicted the COVID-19 cases by taking the data records of 3617 and by using the Decision Tree, SVM, Naïve Bayes, Random Forest, and KNN algorithms out of all the algorithms Decision Tree algorithm got the highest accuracy rate when compared with other algorithms.

VII. FUTURE SCOPE

Currently, we predicted whether the person is cured or not cured due to COVID in future we can create models in such a way to predict the possible affected regions Based on the speed of increase of cases and the places that the affected people have visited and their geolocation we can be able to track what may be the affected areas in the future. And we can also use Hadoop and the concept of HDFs in such a way that we can be able to process huge datasets.

VII. REFERENCES

[1] Chaurasia, V., & Pal, S. (2020). COVID-19 pandemic: ARIMA and regression model based worldwide death cases predictions.

[2] Tabik, S.,Gomez-Rios, A., Martin-Rodriguez, J. L., Sevillano-Garcia, I.,Rey-Area, M.,Charte, D., Guirado, E., Suarez, J. L., Luengo, J., Valero-Gonzalez, M. A., Garcia-Villanova, P., Olmedo-Sanchez, E., & Herrera, F. (2020). Covidgr dataset and COVID-sdnet methodology for predicting COVID-19 based on chest X-ray images. IEEE Journal of Biomedical and Health Informatics, 24(12), 3595-3605.

[3] Rochmawati, N., Hidayati, H. B., Yamasari, Y., Yustanti, W.,Rakhmawati, L.,Tjahyaningtijas, H. P ., & Anistyasari, Y. (2020). Covid symptom severity using decision tree. 2020 Third International Conference on Vocational Education and Electrical Engineering (ICVEE).

[4] Mandayam, A. U., A.C, R., Siddesha, S., & Niranjan, S. K. (2020). Prediction of COVID-19 pandemic based on regression. 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN).

[5] Ban, T. Q., Duong, P. L., Son, N. H., & Dinh, T. V. (2020). COVID-19 disease simulation using GAMA platform. 2020 International Conference on Computational Intelligence (ICCI).

[6] Nath, M. K., Kanhe, A., & Mishra, M. (2020). A novel deep learning approach for classification of COVID-19 images. 2020 IEEE 5th international conference on computing communication and automation (ICCCA).

[7] Gambhir, E., Jain, R., Gupta, A., & Tomer, U. (2020). Regression analysis of COVID-19 using machine learning algorithms. 2020 international conference on smart electronics and communication (ICOSEC).

[8] Fuzzy rule-based system for predicting daily case in COVID-19 outbreak. (2020). 2020 4th international symposium on multidisciplinary studies and innovative technologies (ISMSIT).

[9] Kumar, N., & Susan, S. (2020). COVID-19 pandemic prediction using time series forecasting models. 2020 11th international conference on computing, communication and networking technologies (ICCCNT).

[10] Bhadana, V., Jalal, A. S., & Pathak, P. (2020). A comparative study of machine learning models for COVID-19 prediction in India. 2020 IEEE 4th conference on information & communication technology (CICT).

[11] Latif, S., Usman, M., Manzoor, S., Iqbal, W., Qadir, J., Tyson, G., Castro, I., Razi, A., Boulos, M. N., Weller, A., & Crowcroft, J. (2020). Leveraging data science to combat COVID-19: A comprehensive review.

[12] Bhati, A., & Jagetiya, A. (2020). Prediction of COVID-19 outbreak in India adopting Bhilwara model of containment. *2020 5th International Conference on Communication and Electronics Systems (ICCES)*.

[13] Singh, S., Raj, P., Kumar, R., & Chaujar, R. (2020). Prediction and forecast for COVID-19 outbreak in India based on enhanced epidemiological models. *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*.

[14] Istaiteh, O., Owais, T., Al-Madi, N., & Abu-Soud, S. (2020). Machine learning approaches for COVID-19 forecasting. *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*.

[15] Alwaeli, Z. A., & Ibrahim, A. A. (2020). Predicting COVID-19 trajectory using machine learning. *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*.

[16] Kumar, N., & Susan, S. (2020). COVID-19 pandemic prediction using time series forecasting models. *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*.

[17] Maurya, S., & Singh, S. (2020). Time series analysis of the COVID-19 datasets. *2020 IEEE International Conference for Innovation in Technology (INOCON)*.

[18] Laguarta, J., Hueto, F., & Subirana, B. (2020). COVID-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open Journal of Engineering in Medicine and Biology*, *1*, 275-281.

[19] Liew, C., Quah, J., Goh, H. L., & Venkataraman, N. (2020). A chest radiography-based artificial intelligence deep-learning model to predict severe COVID-19 patient outcomes: The CAPE (COVID-19 AI predictive engine) model.

[20] Pham, Q., Nguyen, D. C., Huynh-The, T., Hwang, W., & Pathirana, P. N. (2020). Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic: A survey on the state-of-the-Arts.

[21] Latif, S., Usman, M., Manzoor, S., Iqbal, W., Qadir, J., Tyson, G., Castro, I., Razi, A., Boulos, M. N., Weller, A., & Crowcroft, J. (2020). Leveraging data science to combat COVID-19: A comprehensive review.