Proc. of Int. Conf. on Intelligence Computing and Information Technology , ICIT

Metrics based Study of Machine Learning Model for COVID-19

Priya Mishra[1], Dr,Shikha Verma[2], Dr.Devendra Kumar [3], Meghna Gupta[4] and Vinod Kumar[5]
[1-5]ABES Eng. College/MCA, Ghaziabad, India
Email: priya.mishra@abes.ac.in, {Shikha.verma, hodmca}@ abes.ac.in, {Meghna.guptar, vinod.kumar}@ abes.ac.in

***Abstract*—The biggest health issue in the world is COVID-19. Many variants came into existence in 2019 due to changing behavior of the virus. The virus is easily converted in a new variant or strain, which behaves differently from the previous variant in terms of transmission rate, infection rate, severity, and death rate. With the fast spread and mutating behavior of the virus, the healthcare system faced a lot of pressure to deal with it. Predicting the transmission growth of virus and the number of new, confirmed, recovered, and dead cases in the future have become a big issue for the healthcare sector so that they can increase medical facilities to cope with this virus. The worsening conditions need immediate execution of control schemes to stop its transmission. This paper introduces the concept of machine learning algorithms and accuracy-based Evaluation of Machine Learning Models.**

***Index Terms*— Machine learning, COVID, SVM, RMSI.**

I. INTRODUCTION

COVID-19 has become the biggest health issue all over the world. This virus came in existence in, China. It is highly transmittable and communicable disease which caused by severe acute respiratory syndrome and known as corona virus (SARS-CoV-2). From Chine, It transmitted all over the world and declared as pandemic. Many variants came into existence since 2019 due to changing behaviour of the virus. The virus easily converted in new variant or strain, which behaves different from the previous variant in terms of transmission rate, infection rate, severity and death rate. The fast spread and mutating behaviour of the virus, healthcare system faced a lot of pressure to deal with it. To predict the transmission growth of virus and number of new, confirmed, recovered and death cases in future have become a big issue for the health-care sector so that they can increase medical facilities to cope with this virus. The worsening conditions need immediate execution of control schemes to stop its transmission. Although there is continuous research is going in the direction of forecasting, detection, treatment and recovery of COVID-19 yet health care system and governments must need effective planning and arrangements to manage the adverse effect of COVID. To handle the explosion of COVID-19, there is a great need to estimate and analyze the medical records. Therefore, some prediction tools are required to predict the possibility and cause of existing cases so that it can help the effective and rapid arrangement to overcome the pandemic. For this reason, the research community utilizes the concept of artificial intelligence to develop precise models to predict COVID-19 cases in different countries. Machine learning (ML) is one of the most emerging concepts of artificial intelligence (AI) which provide a strategic approach to develop computerized, complex and objective algorithmic techniques to analyze and detect data. It has already found the potential method for identifying, discovering, encompassing, and beneficial motoring of many Diseases [10–12].The ML

Grenze ID: 02.ICIT.2022.6.512

algorithm creates interferences out of labelled as well as unlabeled input datasets , this feature can be helpful to deal with the irregular pattern of data for prediction of future cases of corona. These methods also evaluate the risk as age, social habits, location, and weather.
ML techniques can be mainly classified in three categories.

- **Supervised learning**
- **Unsupervised learning**
- **Reinforcement learning**

These techniques are very helpful to forecast, diagnosis, medication and identify the transmission factors of virus etc. Deep learning is a branch having efficient algorithms like ANN, CNN, RNN, LSTM etc., are used to predict the number of corona cases, rate of recovery as well as mortality rate. In general, to identify the COVID cases image and patterns recognition are used and they detect symptoms of corona in patient successfully. Unfortunately, there are many places in the world where effective medical facilities are not adequate and clinical testing is very low or not available in such situations or unaffordable X-ray or CT scans are the only option to diagnose the virus. The research works show that machine based learning approaches can helpful to diagnose to corona virus in comparison with X-ray and other image recognition methods. These models are more accurate than other medical approach to deal the transmission, predication as well as taking preventive measures, which help the medical experts to trace infected persons. A number of software solutions based on artificial intelligence are currently in use to trace spread of the virus and lots of research is going on in this direction which help to discover new solutions to cope up with pandemic. The paper aims to present a review of the literature relating to machine learning methods for the study of COVID-19.
This survey covers the papers, which we believe are the key and unique papers in the area of ML and COVID-19. These works are based on all the area of COVID-19 such as transmission rate, frequency of new cases, recovery rate and death rate. We evaluate the significance of proposed machine learning models in context of COVID-19 by using the performance metric such as MSE, RSME, Accuracy and RSME, MAPE etc. We hope this literature survey paper add value to understand the role of machine learning techniques to cope with this pandemic. Section 2 introduces the concept of machine learning algorithms and performance metrics. The comparative study of research work is discussed in section 3.

II. MACHINE LEARNING

Machine learning is branch of artificial intelligence which focuses to train machines to do tasks just like the humans do and are used to solve various real-world problems. These techniques allow computer programs to automatically improve through experience and provide a solution to deal with real scenarios. ML techniques can be divided into three main categories:

- *Supervised Learning*: This technique learn from labeled data (trained data) and act on test data to find the solution. This approach works under two branches.
  - Classification: Process to classify the data into different classes.
  - Regression: To find the relation between dependent and independent variable.
- *Unsupervised Learning: Unlabeled data* is used for learning to learning set to extract features and recognize the patterns.
  - Clustering
  - Dimension reduction of large and high-dimensional data
- *Reinforcement Learning*: This mechanism learn from the environment through trial and error by applying the reward and punishment.
  - Policy Based
  - Value Iteration based

Your goal is to simulate the usual appearance of papers in a Conference Proceedings or Journal Publications. We are requesting that you follow these guidelines as closely as possible.

A. Performance Matrices

In this paper, the comprehensive survey work is studied and grouped according the metrics used to evaluate the performance of existing research work . In this section, we study only those metrics, which are highly used to evaluate the performance of models, but only some of them are best suited for machine learning models. Therefore, this study focuses on three most important performance metrics, which are commonly used for ML models. These metrics are:

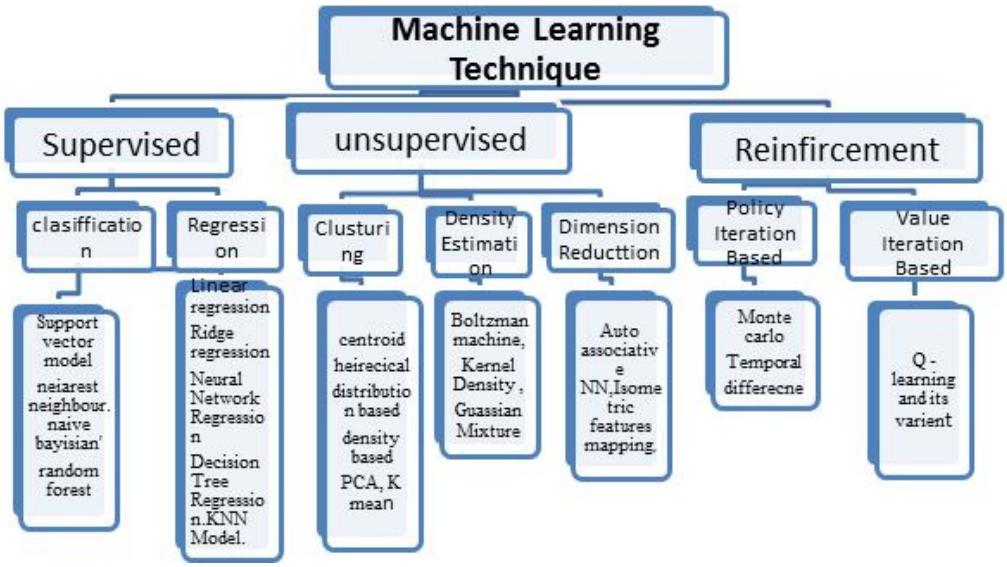

Figure 1. Machine learning Techniques

- Root Square Mean Error (RSME): This error is calculated as the square root of Mean Square Error (MSE) is broadly used as an absolute measurement of the goodness for the fitted graph. By applying this metric, we can find the best regression model. To calculate the MSE, we find the difference between the actual output with desired output, which is dividing by all the data points to select the best regression model.

$$\boldsymbol{MSE} = \frac{\boldsymbol{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}(\boldsymbol{Y_i} - \boldsymbol{Y'_i})$$

To calculate the RSME, we take square root of MSE

$$\boldsymbol{RMSE} = \frac{\boldsymbol{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}}\sqrt{(\boldsymbol{Y_i} - \boldsymbol{Y'_i})}$$

Where, N= total number of data inputs, Yi =Actual output, Yi' =Predicted outputs

The other metric is mean absolute error (MAE) which is same as the MSE and RSME the only difference is that works with direct calculation of sum of error. It also gives more penalization to large prediction error by square it while MAE treats all

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|(Y_i - Y'_i)|$$

- **Accuracy:** To find the correct estimation of output, we use this metric. This is calculated as the as correct prediction over the total number of input data. It works well in those cases where equal number of data belonging to each class. This is calculated as :

$$Accuracy = \frac{Correct\ Prediction\ (estimation)}{Total\ number\ of\ inputs}$$

- **Mean Absolute Percentage Error (MAPE)**: It is most commonly used to measure forecast accuracy. It is considered as the average of the percentage error. To calculate it, the sum of the discrete absolute errors divided by the demand is required for each period.

III. Comparative Study

Corona is a major problem that health-care systems are currently dealing with. Due to the risk of the corona virus mixing with other lung infections, diagnosis and categorization of corona patients is a major challenge. COVID-19 instances are now identified using reverse transcriptase quantitative polymerase chain reaction (RT-qPCR). Unfortunately, this is a time-consuming procedure that requires medical experts, but medical expertise is not

always available. Sometimes testing is not accurate and shows high false-positive rate, according to a few studies [23-25]. To address this dilemma, scientists from the fields of virology, medicine, and artificial intelligence (AI) must collaborate

To deal with this situation, AI experts are attempting to create substantial solutions that can aid in the efficient identification, forecasting, and treatment of COVID-19 [26]. We investigate and compare various machine and deep learning-based approaches to dealing with COVID-19 instances in this research. The survey delves into the performance criteria that are used to assess machine-learning models. Table 1 presents a comparison of machine learning models with accuracy as the performance metric, whereas table 2 displays machine learning methods using RMSE as the performance measure.

TABLE I. ACCURACY BASED EVALUATION OF MACHINE LEARNING MODEL IN COVID-19

STUDY	**OBJECTIVE**	**Data Source**	**Used Model**	**Metrics**	**Results**
Gomathi **et.al.[5]**	Propose a model, which help to forecast the transmission of corona virus taking age, gender, travel, and contact with patients, clinical indicator and critically as inputs.	India	Automated machine learning (AutoML), Naïve Bayes, logistic regression, KNN, AutoML method of Cox proportional hazards (CoxPH).	Accuracy	Rate of Auto-ML algorithm is high among all the applied machine learning models
Yadav et.al. **[6]**	To Propose a model to predict the numbers of existing cases, new reported cases, confirmed and recovery cases of COVID-19	India(On daily Basis)	Mathematical modeling and ANN	Accuracy	transmission rate of virus will exponentially grow
Yadav et.al.[11]	To predict the ,transmission rate and mitigation rate of corona virus also find dependency and correlation with weather condition	Positive cases, recoveries, deaths for three months in different countries	SVM	Accuracy	SVR achieves maximum accuracy among all the models.
Malki. At.el.[13]	To find the correlation and dependency between mortality rate, weather data and virus	Kaggle , Johns Hopkins Center for Systems Science	Variant of regression model are used	Accuracy	On increasing the humidity, the regression models perform well.
Sedik, at.el.[14]	To increase the accuracy of predication of corona cases by using the deep learning models	Health and COVID images are used	CNN Model, Binary classification, LSTM	Accuracy	Provide rapid and consistent Corona virus diagnosis
Xu, Xiaowei, et al.[15]	A model is developed to extract specific graphical features of COVID-19 and enhance clinical diagnosis to save critical time for disease control.	World wide	Images are used from different age groups of COVID patients and healthy persons healthy cases having mean age 39 years	Accuracy	Accuracy rate of model is 89.5% to predict the situation.
[16]	To study AI system for differentiating COVID-19 and other pneumonia at chest CT and assessing radiologist performance without and with AI assistance.	CT images from 1186 patients (132,583 CT slices).	Fully connected DL model Binary classification	Accuracy	Accuracy is tested with area under the receiver operating characteristic curve of 0.90 and area under the precision-recall curve of 0.87.
Sethy at.al.[17]	To detect of coronavirus infected patient using X-ray images.	Repository of GitHub and Kaggle.	SVM and CNN	Accuracy	The classification model, i.e. ResNet50 plus SVM achieved highest accuracy,
Zhang at.el,[18]	To propose a model to diagnose NCP(novel coronavirus pneumonia) In addition, differentiate it from other common pneumonia and normal controls.	deep learning methods	large computed tomography (CT) database from 3,777 patients	Accuracy	The proposed model accurately differentiate NCP with normal pneumonia and cold)

TABLE II. RMSE BASED EVALUATION OF MACHINE LEARNING MODEL IN COVID-19

STUDY	OBJECTIVE	Data Source	Used Model	Metrics	Results
Kumari,at.el.[4]	Aims to find growth and transmission rate of infection.	Data of four highly impacted states of India on weekly basis	Hybrid variants of LSTM and CNN, models	RMSE, MAPE	Hybrid variant of LSTM and CNN most accurately find the confirmed cases.
Dutta at.el, [2]	Develop a generalized model to identify positive, negative , death and recovery of COVID patient	Kaggle (online data)	hierarchical Bayes theorem, LSTM ,RNN and GRU	RMSE	The hybrid model of LSTM shows better results than others
Ardabili at. el**[1]**	Predication of burst of corona cases	Data of 75 days from 5 different countries Italy China ,Iran,USA ,Germeny	MLP and ANFIS models Then optimized by GA PSO GWO	**RMSE and Correlation**	GWO provides the highest accuracy than with PSO and GA
Punn et.al.[7]	To deal with Confirmed Deaths Recovery	World wide	SVR,LSTM,PR DNN	RMSE	polynomial regression (PRPR) yield a minimum RSME value than others
Melin, at.el **[8]**	To forecast the predict the number of confirmed cases and deaths before 10-day	Data from Maxico Mexico Government	Nonlinear autoregressive networks (NAR) with different parameters and function fitting network (FITNET). The output of the modules is combined via a fuzzy integrator	RMSE	95.32% RMSE value is calculated
Kapoor at.el..[22]	To forecast daily new cases in the US.	Data from NYT and online data of US over the period of Jan-Apr, 2020.	Temporal graph NN learns relationship mobility data. The model is implemented via GNN framework	RMSE, Correlation coefficient	RSME 99.5%.
Rustam et.al.[9]	To forecast data on basis of 3 previous days	Worldwide.	Variants of Regression, LASSO, SVM,ES	RMSE	ES and Lasso perform better
Direkoglu et.al. [10]	To aim to deal with forecasting of data.	WHO, CCDCP, and world meter. .	LSTM	RMSE	1.5% RMSE.

IV. CONCLUSIONS

Researchers are always striving to uncover the most effective and best strategies to deal with the negative effects of the COVID-19 pandemic on human life, as seen by the literature examined. We may conclude from this vast research that machine learning algorithms are emerging as a valuable tool for investigating and evaluating the potential effects of the corona virus and developing methods to combat it. In this research, we attempt to examine all types of proposed works that improve virus spread prediction, transmission rate, outbreak forecasting, recovery, and mortality rate, as well as their implications using acceptable performance measures. Finally, there are still many gaps and obstacles in terms of future study directions due to the inconsistent and fragmented data. and the virus's changing nature As a result, a deeper knowledge of the corona virus is required in the future so that academics and scientists may develop far more robust and trustworthy mechanistic models, especially when dealing with virus variations that continue to mutate and thwart mitigation efforts and varying nature of virus. Therefore, in future we need a better understanding of the corona virus so that the researchers and scientists will provide vastly more robust and reliable mechanistic models, particularly when dealing with virus variants that continue to mutate and dampen mitigation efforts.

REFERENCES

[1] Ardabili, Sina F., et al. "Covid-19 outbreak prediction with machine learning." Algorithms 13.10 (2020): 249

[2] Dutta, Shawni, and Samir Kumar Bandyopadhyay. "Machine learning approach for confirmation of covid-19 cases: Positive, negative, death and release." Iberoamerican journal of medicine 2.3 (2020): 172-177

[3] Pinter, Gergo, et al. "COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach." Mathematics 8.6 (2020): 890.

[4] Kumari, Pratima, and Durga Toshniwal. "Real-time estimation of COVID-19 cases using machine learning and mathematical models-The case of India." 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS). IEEE, 2020.

[5] Gomathi, S., et al. "Pattern analysis: predicting COVID-19 pandemic in India using AutoML." World Journal of Engineering (2020).

[6] Yadav, Ramjeet Singh. "Data analysis of COVID-2019 epidemic using machine learning methods: a case study of India." International Journal of Information Technology 12.4 (2020): 1321-1330.

[7] Punn, Narinder Singh, Sanjay Kumar Sonbhadra, and Sonali Agarwal. "COVID-19 epidemic analysis using machine learning and deep learning algorithms." MedRxiv (2020).

[8] Melin, Patricia, et al. "Multiple ensemble neural network models with fuzzy response aggregation for predicting COVID-19 time series: the case of Mexico." Healthcare. Vol. 8. No. 2. Multidisciplinary Digital Publishing Institute, 2020.

[9] Rustam, Furqan, et al. "COVID-19 future forecasting using supervised machine learning models." IEEE access 8 (2020): 101489-101499.

[10] Direkoglu, Cem, and Melike Sah. "Worldwide and regional forecasting of coronavirus (covid-19) spread using a deep learning model." medRxiv (2020).

[11] Yadav, Milind, Murukessan Perumal, and M. Srinivas. "Analysis on novel coronavirus (COVID-19) using machine learning methods." Chaos, Solitons & Fractals 139 (2020): 110050.

[12] Tian, Yuan, Ishika Luthra, and Xi Zhang. "Forecasting COVID-19 cases using Machine Learning models." MedRxiv (2020).

[13] Malki, Zohair, et al. "Association between weather data and COVID-19 pandemic predicting mortality rate: Machine learning approaches." Chaos, Solitons & Fractals 138 (2020): 110137.

[14] Sedik, Ahmed, et al. "Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections." Viruses 12.7 (2020): 769.

[15] Xu, Xiaowei, et al. "A deep learning system to screen novel coronavirus disease 2019 pneumonia." Engineering 6.10 (2020): 1122-1129.

[16] Artificial Intelligence Augmentation of Radiologist Performance in Distinguishing COVID-19 from Pneumonia of Other Origin at Chest

[17] Sethy, Prabira Kumar, and Santi Kumari Behera. "Detection of coronavirus disease (covid-19) based on deep features." (2020).

[18] Zhang, Kang, et al. "Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography." Cell 181.6 (2020): 1423-1433.

[19] Shavlik, Jude W., Thomas Dietterich, and Thomas Glen Dietterich, eds. Readings in machine learning. Morgan Kaufmann, 1990.

[20] Walther, Bruno A., and Joslin L. Moore. "The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance." Ecography 28.6 (2005): 815-829.

[21] Li, Zhilin. "On the measure of digital terrain model accuracy." The photogrammetric record 12.72 (1988): 873-877.

[22] Kapoor, Amol, et al. "Examining covid-19 forecasting using spatio-temporal graph neural networks." arXiv preprint arXiv:2007.03113 (2020)

[23] Wu, X.; Wang, Z.; He, Z.; Li, Y.; Wu, Y.; Wang, H.; Liu, Y.; Hao, F.; Tian, H. A follow-up study shows that recovered patients withre-positive PCR test in Wuhan may not be infectious. BMC Med. 2021, 19, 1–7. [CrossRef]

[24] Armstrong, S. Covid-19: Tests on students are highly inaccurate, early findings show. BMJ 2020, 371, m4941. [CrossRef]

[25] Roy, S. Physicians' Dilemma of False-Positive RT-PCR for COVID-19: A Case Report. SN Compr. Clin. Med. 2021, 3, 255–258.

[26] Alserehi, H.A.; Alqunaibet, A.M.; Al-Tawfiq, J.A.; Alharbi, N.K.; Alshukairi, A.N.; Alanazi, K.H.; Bin Saleh, G.M.; Alshehri