World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:16, No:12, 2022

Machine Learning Techniques for COVID-19 Detection: A Comparative Analysis

Abeer Aljohani

***Abstract*—**The COVID-19 virus spread has been one of the extreme pandemics across the globe. It is also referred as corona virus which is a contagious disease that continuously mutates into numerous variants. Currently, the B.1.1.529 variant labeled as Omicron is detected in South Africa. The huge spread of COVID-19 disease has affected several lives and has surged exceptional pressure on the healthcare systems worldwide. Also, everyday life and the global economy have been at stake. Numerous COVID-19 cases have produced a huge burden on hospitals as well as health workers. To reduce this burden, this paper predicts COVID-19 disease based on the symptoms and medical history of the patient. As machine learning is a widely accepted area and gives promising results for healthcare, this research presents an architecture for COVID-19 detection using ML techniques integrated with feature dimensionality reduction. This paper uses a standard University of California Irvine (UCI) dataset for predicting COVID-19 disease. This dataset comprises symptoms of 5434 patients. This paper also compares several supervised ML techniques on the presented architecture. The architecture has also utilized 10-fold cross validation process for generalization and Principal Component Analysis (PCA) technique for feature reduction. Standard parameters are used to evaluate the proposed architecture including F1-Score, precision, accuracy, recall, Receiver Operating Characteristic (ROC) and Area under Curve (AUC). The results depict that Decision tree, Random Forest and neural networks outperform all other state-of-the-art ML techniques. This result can be used to effectively identify COVID-19 infection cases.

***Keywords*—**Supervised machine learning, COVID-19 prediction, healthcare analytics, Random Forest, Neural Network.

I. Introduction

THE healthcare sector is quite wide and requires a lot of data for analysis. Additionally, the issue of data handling is central to this sector, as timely medical care requires real-time prediction and the transmission of data to experts. The great havoc is created by the COVID-19 pandemic. COVID-19 is a wide-ranging family of virus that produces varying symptoms such as flu and common cold to critical respiratory issues. As stated by NCBI, “In past two decades, several viral epidemics like Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) have been registered” [1]. According to Gambhir et al., WHO declared pandemics in 2002-2004 and H1N1 influenza in 2009 [1]. In 2019, a few instances of unfamiliar low respiratory infections were spotted in Wuhan city of China around December 2019. This spotted mysterious infection having diseases like pneumonia was termed “COVID-19” by WHO. By 30th January, 2020 this strain had impacted nearly 20 countries and hence WHO pronounced this infection as a Public Health Emergency of International Concern (PHEIC) [1]. The common signs are dry cough, fatigue and fever. Other less common symptoms are loss of smell or taste, headache, joint pain, nasal congestion, vomiting, diarrhea, chills, dizziness and sore throat. It rapidly became pandemic due to a huge upsurge in the cases as well as mortality rate. So far, specific treatments are still not available for this virus. Hence, the only possible way to decrease the spread of the virus is by maintaining hygiene (i.e., washing hands, sanitizing), social distancing and taking vaccinations. Currently, there have been several recoveries throughout the globe and active cases are reducing. The pandemic, however, is still not under control. Seniors with medical issues such as diabetes, heart illness, cancer, lung disease, etc. have a higher risk through this virus [2]. Also, the virus is continuously bringing up new strains by mutation. In 2021, Omicron, detected in South Africa is a big threat to the world as it has the highest mutation in contrast to its previous mutations. The world's economy, hygiene, and most importantly, health has all been affected negatively by this virus. Many countries have announced lockdown and sealed their borders. As major economic activities are shut down, the normal routine of people is affected to a major extent. Hence, the primary requirement is to realize the features as well as characteristics of this disease.

Methods to detect COVID-19 disease include rapid tests, RTPCR tests. Once COVID-19 has been detected, the effect of the virus in the body can be detected using several measures such as Chest CT images, blood tests and X-ray images. In the blood test, Complete Blood Count (CBC), D-Dimer and C-reactive Protein (CRP), S-ferritin, Alanine Aminotransferase (ALT), Lactate dehydrogenase (LDH) are measured. CBC showcases the comprehensive health and also detects infection’s range [3]. Inflammation in the human body is detected using CRP which is a protein made by the liver [4]. Blood clot dissolution is detected using D-dimer [5]. To measure the iron in the body, S-ferritin is utilized which is a blood protein [6]. Liver damage can be checked using the ALT test [7]. The harmed tissues of the human body can be estimated using LDH which is an enzyme discovered in body’s major cells [8]. Thus, the paper aims to predict whether the person has COVID-19 or not based on several symptoms such as breathing problems, dry cough, fever, sore throat, asthma, running nose, chronic lung disorder, heart disease, headache and diabetes.

There are several areas where AI has given promising results. It includes computer vision [9], [10], ML [11] and healthcare

A. Aljohani is with the Computer Science Department, Applied College, Taibah University, Madinah 46537, Saudi Arabia (e-mail: aahjohani@taibahu.edu.sa).

[12]. Hence, AI techniques can also be utilized to predict COVID-19 infection. As the COVID-19 cases start rising heavily, a huge burden is caused on the hospitals. Also, early diagnosis is very crucial for this virus. COVID-19 cases are increasing heavily and there are lack of awareness and increased fear. Due to this, people go through the test for verifying whether they are infected or not. For the detection of COVID-19 infection, there are currently just a few analytical tests available. Therefore, there is a requirement to employ an automatic prediction mechanism to efficiently predict the infection and reduce the utilization of test kits. This will also help in detecting the infection at an early stage and prevent spreading amongst individuals by providing proper treatment to them at an initial stage.

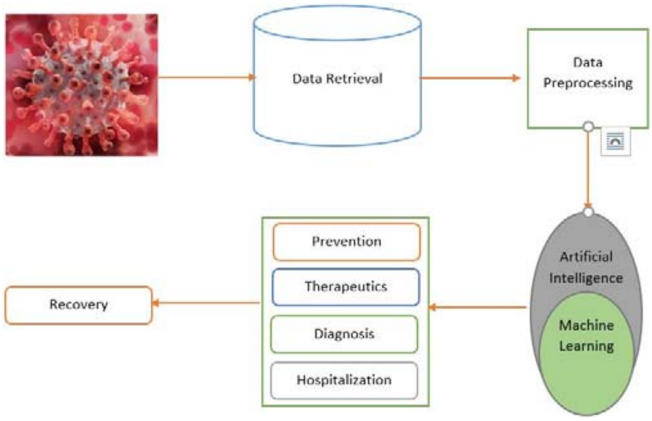

Fig. 1 Architecture for generalized COVID-19 detection to recovery process using AI and ML

Fig. 1 shows how AI and ML can help in predicting several roles such as prevention, diagnosis, hospitalization, etc. It also provides a brief overview of stages a person goes through from infection to recovery and how ML helps in this phase. ML is indeed a potential technique that can be very helpful in prevention against the COVID-19. ML can be utilized to handle enormous data and efficiently forecast the infection through several symptoms. It can help in diagnosing as well as predicting COVID-19. Three categories of ML include unsupervised, supervised and reinforcement learning. Unsupervised ML is a technique which does not have labelled data. In this technique, training samples are provided to the machine and the machine identifies the hidden patterns from the dataset. Using labelled datasets, supervised ML trains the system. The samples are labeled according to the class which they belong [13]. The machine analyses the labelled data and predicts new instances based on results retrieved from the past data. In reinforcement learning, a trial-and-error approach is utilized with an objective to determine the most suitable steps by a machine that acts as an agent [14]. Whenever the machine executes a task correctly, it is rewarded and its state is increased. In case, a machine does not execute its task it is punished and its state is also decreased. This task is continuously repeated numerous times until the machine performs a specific task successfully or reaches the specified number of states. Reinforcement learning is utilized in training robots to provide personal assistance and also to present tasks like humans.

The major contribution of the paper is that it proposes an efficient ML architecture that provides high accuracy for predicting the COVID-19 based on symptoms. Also, it pre-processes the data before providing it for training to increase the efficiency of the proposed architecture.

Section II showcases the related literature works carried out in this field. In Section III the proposed architecture and description of the same is shown. Section IV shows experimental results and analysis and finally, the conclusion and future scope are presented at last.

II. Related Work

Monika et al. [15] offered three ML algorithms for evaluating COVID19 cases: Decision Tree Regressor, Polynomial Regression, and Random Forest. For Polynomial Regression, their model obtains 90% accuracy. Gupta et al. used ML approaches including the Regression model and the Susceptible Exposed Infectious Removed (SEIR) model to predict variance in COVID-19 disease contamination [16]. The RMS log error was calculated to be 1.52 for the RMS model and 1.75 for the Regression model. Iwendi et al. combined the Random Forest and AdaBoost techniques in their research [17]. The model has a 0.86 F1 score and a 94% accuracy rate. For forecasting human intelligence, S. Makridakis et al. offered statistical approaches and ML training [18].

Several surveys mention the use of ML for COVID-19 diagnosis. Batista et al. utilized the Kaggle dataset [19] to precisely classify COVID-19 patients of Brazil with 85% accuracy utilizing the Support Vector Machine (SVM) method [20]. Mondal utilized multilayer perceptron (MLP) and XGBoost and achieved 91% classification accuracy on the same dataset [19]. Meza precisely classified COVID-19 patients with 91% cases on UCLA Health System dataset of USA. Meng et al. and Sun et al. measured COVID-19 with accuracies of 89% and 91% respectively [21], [22].

In last two years, several studies have been proposed in which AI was used to identify, prevent, and predict the COVID-19 pandemic. Wang et al. [23] used Chest X-Ray images (CXR) and presented a Convolutional Neural Network (CNN) architecture for COVID-19 prediction in patients. The ImageNet model is trained using an open-source dataset of CXR pictures to apply transfer learning. Pal et al. used a Long Short Term Memory (LSTM) model on weather and trends data from a given country to anticipate the country's specific risk of COVID-19 [24]. For anticipating the spread of the pandemic in China, Liu et al. used ML on data containing internet activity, papers from health organizations, media activity, and news reports [25]. For the next 70 days, Bayes and Valdivieso employed the Bayesian approach to forecast deaths in Peru [26]. Beck et al. used Bidirectional Encoder Representations from the Transformers (BERT) to categorize commercial drugs which can be given to COVID-19 patients [27]. Tang et al. utilized the Random Forest-based ML method on CT scans images to examine patients infected with patients [28]. A

Generative Adversarial Network (GAN) model is recommended to identify pneumonia from CXR scan images at early stage by Khalifa et al. [29]. Sujatha et al. applied MLP, linear regression and Vector autoregression model for predicting the epidemiological pattern of COVID-2019 cases [30]. Waheed et al. utilized a deep learning approach for production of CXR images by introducing a model such as Auxiliary Classifier Generative Adversarial Network (ACGAN). This is also called COVIDGAN [31].

Wu et al. presented a system to detect COVID-19 and classify whether it is COVID-19 or Pneumonia [32]. There is a challenge to quickly diagnose the disease. Hence, a Random forest-based ML model is used which had 95.95% accuracy. The dataset compromises 253 records from 169 suspected patients. Each record comprises 49 parameters. However, only 11 parameters were chosen as final indicators. To ensure reliability, several validation methods were used. Bastug et al. proposed an architecture that forecasts the severe prognosis of COVID-19 [33]. They utilized laboratory as well as clinical data to handle COVID-19 patients and recognize patients that require Intensive Care Unit (ICU). Brinati et al. presented a Random Forest-based ML system to detect COVID-19 [34]. The architecture was compared with other ML techniques. Random Forest has the highest accuracy on their two proposed model with 82% and 86%. Hussein et al. proposed an analysis intending to evaluate clinical features as well as radiological features of COVID-19 patients. The prime aim is to analyze whether a patient requires ICU or not [35]. For analysis, laboratory data of 72 COVID-19 patients were utilized. The data covered CBC, D-dimer, seru, ferritin and CRP. The limitation is that dataset has a very less information of COVID-19 patients.

This research primarily aims on identifying the existence of COVID-19 in a person. Hence, a supervised ML model has to be developed. This paper predicts COVID-19 symptoms using several ML methods such as Neural Network, SVM, K-Nearest Neighbor (KNN), Decision Tree and Random Forest (RF) along with feature reduction done by PCA.

III. COVID-19 Detection Using ML Techniques

The proposed architecture for detection of COVID-19 based on symptoms is given in Fig. 3.

A. Data Collection

The dataset used for the detection of COVID-19 using ML techniques is a freely available dataset from Kaggle [36]. The dataset consists of various symptoms of the patients i.e., features used for training and the outcome i.e., the target variable. There are 20 features and one target variable. The dataset consists of 5434 instances. The features and target variable are given below [37].

- Breathing Problem: whether the patient is suffering from shortness of breath.
- Fever: whether the patient is having a fever or not.
- Dry Cough: whether the patient is suffering from continuous coughing without mucus.
- Sore Throat: whether the patient is suffering from a sore throat.

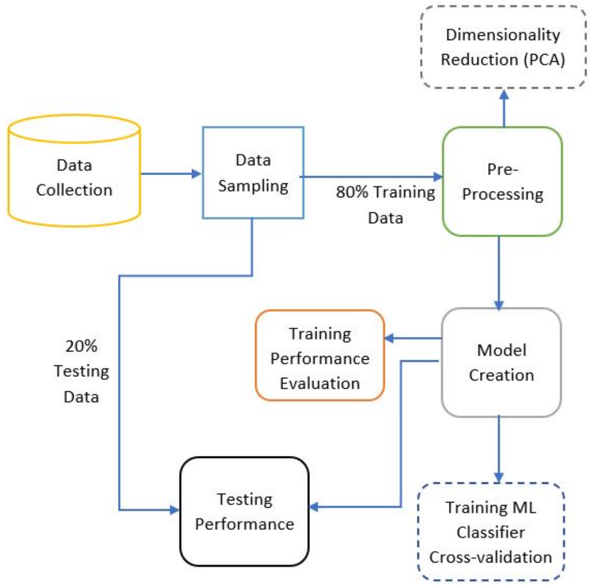

Fig. 2 Architecture for detection of COVID-19 using ML

- Runny Nose: whether the patient is suffering from a runny nose.
- Asthma: whether the patient is suffering from asthma or not.
- Chronic Lung Disease: whether the patient is suffering from lung disease.
- Headache: whether the patient is suffering from a headache.
- Heart Disease: whether the patient has cardiovascular disease
- Diabetes: whether the patient has diabetes.
- Hypertension: whether the patient has high blood pressure.
- Fatigue: whether the patient is experiencing tiredness.
- Gastrointestinal: whether the patient has any gastrointestinal problems.
- Abroad Travel: whether the patient has gone abroad recently.
- Contact with COVID-19 Patient: weather the patient had been in contact with COVID-19 infected person.
- Attended Large Gathering: whether the patient or any family member had appeared in a public meeting.
- Visited Public Exposed Places: whether the patient had visited mosques or temples, public gardens and other such religious places recently.
- Family working in public: whether any family member of the patient is working in an office, market area, etc.
- Exposed places: whether the patient is exposed to crowded areas.
- Wearing mask: whether the patient is wearing a mask all the time.
- Sanitization from market: whether the things brought from the market were sanitized before using.
- COVID 19: whether the patient is detected as COVID positive or negative.

B. Data Sampling

Training and testing data are separated from the dataset for measuring performance of the trained model. The trained model receives 80% of the dataset, while the tested model receives 20% of the dataset.

C. Pre-processing

The dataset contains some features which are not so relevant as others for training the model. So, those features need to be removed to prevent the curse of dimensionality issue while training the ML algorithm. Here, "dimensionality" denotes the count of features in the dataset. In this, the curve of dimensionality is defined as "when the number of variables is large as compared to the number of instances in the dataset, the ML algorithms are unable to train effective models". Thus, to solve this problem, the paper used PCA for feature reduction.

D. Model Creation

The model is created by training the original as well as the reduced features selected and extracted by PCA. Multiple supervised ML techniques along with the 10-fold cross-validation are used for training the model. To generalize the architecture on the selected dataset, k-fold cross-validation is utilized. This provides the solution to the overfitting issue that increases because of several features or containing noisy data. For training, the proposed work uses various classification algorithms such as neural network, KNN, SVM, Decision Tree, Naïve Bayes and RF.

E. Evaluation Parameters

The training and testing assessment of the proposed architecture is done by utilizing standard parameters that include Confusion matrix, Classification accuracy (CA), F1 measure, precision, recall, ROC and AUC.

Confusion Matrix: This parameter is utilized for the binary classification which is shown in Table I.

TABLE I
CONFUSION MATRIX

	Actual Yes	Actual No
Predicted Yes	True Positive (TP)	False Negative (FN)
Predicted No	False Positive (FP)	True Negative (FN)

In Table I, True Positive (TP) are the occurrences of the dataset that are predicted as well as classified as true. The occurrences that are predicted as well as classified as false are called True Negative (TN). False Positives (FP) are occurrences that are false but are predicted as true by the classifier. Finally, the occurrences that are true, however, classifier predicts them as false are called False Negative (FN).

CA: Accuracy is the number of correct predictions made by the classification model to the total number of predictions made.

$$Classification\ Accuracy\ (CA) = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

Recall: Recall is the percentage of predicted positives by the model out of the total positive values in the dataset.

$$Recall\ (True\ Positive\ Rate) = \frac{TP}{TP+FN} \tag{2}$$

Precision. Precision is the percentage of true positives in the dataset out of all the positive predicted values by the model.

$$Precision\ (Positive\ Predictive\ value) = \frac{TP}{TP+FP} \tag{3}$$

F1 Score. F1 score is the harmonic mean of precision and recall. Mathematically, it is the weighted average of precision and recall as in (4):

$$F1\ score = \frac{2*Precision*Recall}{Precision+Recall} \tag{4}$$

ROC curve: The ROC curve is used to measure the performance so that ability of classifiers can be examined by varying its discrimination threshold [12]. ROC is a graph that plots the FP rate (specificity) versus TP rate (sensitivity). Here, at (0,1) point, all the occurrences are precisely classed for all the classifiers in their specific group. This showcases the efficiency of a classifier. Generally, a classifier has a parameter that varies in such a way that the TP is in proportionate to FP. A point (FP, TP) is retrieved by varying the parameter value. To plot the ROC curve, a series of such points are deployed. A single point (FP, TP) also called ROC point is used to describe a classifier that does not have fluctuating parameter and it is showcased through a single point.

AUC: This parameter estimates the likelihood which showcases the proficiency of the classifier to differentiate amongst several classes. In this curve when the utilized features are normalized, high preference is given to the positive observation in contrast to a negative observation. Hence, the most prominent classifier has an AUC score of 1.

$$\text{AUC} = \text{P}\ (m_1 > m_2) \tag{5}$$

Here, m1 is the score of the observation that is positive and m2 is the score of the observation that is negative.

IV. Experimental Results and Analysis

The proposed approach is simulated in the Orange Data Mining tool (version:3.24). The tool is open-source and widely used for Data mining, data visualization, ML and image processing.

Table II shows the training results with 10-fold cross validation with PCA. Decision Tree and RF outperform with the highest F1 score of 98.0%.

TABLE II
RESULTS WITH TRAINING 10-FOLD CROSS-VALIDATION PCA

Model	AUC	CA	F1	Precision	Recall
SVM	0.935	0.892	0.894	0.896	0.892
Naïve Bayes	0.989	0.965	0.964	0.964	0.965
kNN	0.996	0.974	0.974	0.975	0.974
Neural Network	0.998	0.978	0.978	0.978	0.978
RF	0.998	0.980	0.980	0.981	0.980
Decision Tree	0.998	0.980	0.980	0.980	0.980

Table III shows training results with 10-fold cross validation without PCA. Decision Tree, RF and neural networks perform well with a 98.2% F1 score.

TABLE III
TRAINING 10-FOLD CROSS VALIDATION WITHOUT PCA

Model	AUC	CA	F1	Precision	Recall
SVM	0.931	0.877	0.882	0.891	0.877
Naïve Bayes	0.990	0.966	0.965	0.965	0.966
kNN	0.959	0.967	0.966	0.968	0.967
Neural Network	0.998	0.982	0.982	0.982	0.982
RF	0.998	0.982	0.982	0.982	0.982
Decision Tree	0.998	0.982	0.982	0.982	0.982

Tables IV and V show testing results with and without PCA respectively. Decision Tree, RF and neural network perform well with 98.7% F1 score with PCA while RF and neural network perform well with 99% F1 score without PCA.

TABLE IV
TESTING WITH PCA

Model	AUC	CA	F1	Precision	Recall
SVM	0.955	0.936	0.935	0.935	0.936
Naïve Bayes	0.992	0.970	0.969	0.969	0.970
kNN	0.999	0.983	0.982	0.982	0.983
Neural Network	0.999	0.987	0.987	0.987	0.987
RF	0.999	0.987	0.987	0.987	0.987
Decision Tree	0.999	0.987	0.987	0.987	0.987

TABLE V
TESTING WITHOUT PCA

Model	AUC	CA	F1	Precision	Recall
SVM	0.973	0.909	0.913	0.921	0.909
Naïve Bayes	0.992	0.970	0.969	0.969	0.970
kNN	0.945	0.972	0.971	0.973	0.972
Neural Network	0.999	0.990	0.990	0.990	0.990
RF	0.999	0.990	0.990	0.990	0.990
Decision Tree	0.999	0.987	0.987	0.987	0.987

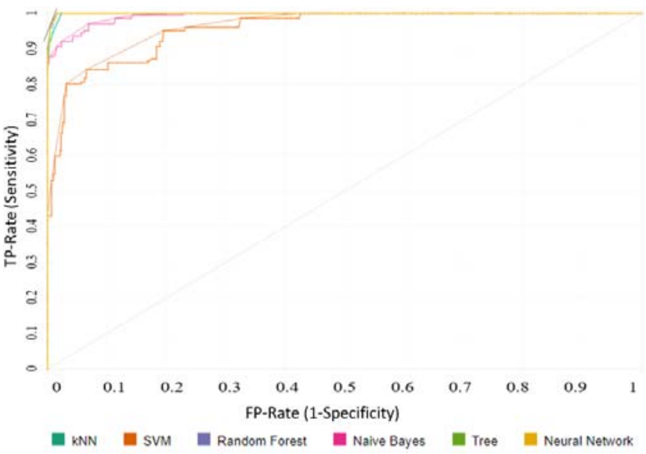

Fig. 3 ROC curve testing with PCA

Fig. 3 depicts the ROC curve of test data with PCA and Fig. 4 depicts the ROC curve of test data without PCA. From Fig. 3, it is clear that SVM and Naïve Bayes do not perform well as their graph is quite far from the ideal point (0,1) while KNN, Decision Tree, RF and neural network perform well as their graphs are too close to the point (0,1). From Fig. 4, it is clear that the Decision Tree, RF and neural networks perform well while SVM and KNN do not perform well.

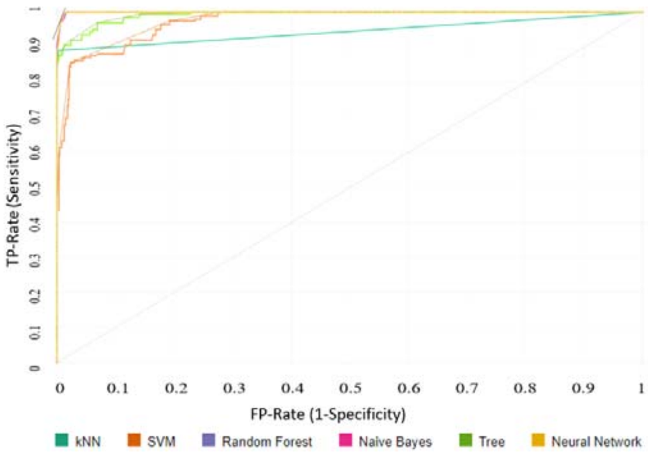

Fig. 4 ROC curve testing without PCA

V. CONCLUSION AND FUTURE SCOPE

COVID-19 has caused a big panic worldwide. All sectors are facing challenges to handle this situation. Government organizations and researchers are worried that several percentages of the world's population will be affected by this pandemic. Also, the virus is continuously mutating and coming up with a new strain. AI has emerged as a savior by providing impactful solutions to almost all sectors. This paper showcases an architecture for COVID-19 disease identification in a patient by using symptoms and medical history. In this paper, the ML approach is used for identifying the occurrence of COVID-19. Hence, several ML algorithms such as KNN, SVM, RF, Naïve Bayes, Decision Tree, neural network are applied in the architecture. The paper utilized the standard UCI dataset which has symptoms and medical history for 5434 patients. The model is evaluated using standard parameters such as Confusion matrix, accuracy, precision, recall, F1 score, ROC and AUC. The evaluation parameters depict that the Decision Tree, RF and neural networks outperform other state-of-the-art ML techniques. In the future, ML techniques can be utilized in tracing COVID-19 cases, forecasting, producing dashboards, diagnosing and giving suitable medications, producing alerts to manage social distancing and several other processes for controlling the spread of the virus.

REFERENCES

[1] E. Gambhir, R. Jain, A. Gupta and U. Tomer, "Regression analysis of COVID-19 using machine learning algorithms," *In: International conference on smart electronics and communication (ICOSEC)*, Tamil Nadu, India, 2020, pp. 65-71.

[2] S. Sakib, "Dl-crc: Deep learning-based chest radiograph classification for COVID-19 detection: A novel approach," *IEEE Access*, 8, 2020, pp. 171575-171589.

[3] D. Weatherby and S. Ferguson, "Blood chemistry and CBC analysis," *Weatherby & Associates, LLC*, 4, 2002, pp. 1-312.

[4] W. Ling, "C-reactive protein levels in the early stage of COVID-19," *Medecine et maladies infectieuses*, 50(4), 2020, pp. 332-334.

[5] L. Zhang, X. Yan, Q. Fan, H. Liu, X. Liu, Z. Liu, Z. Zhang, "Ddimer

levels on admission to predict in-hospital mortality in patients with COVID-19," *Journal of Thrombosis and Haemostasis*, 18(6), 2020, pp. 1324-1329.

[6] S. TuralOnur, S. Altin, S. N. Sokucu, B. I. Fikri, T. Barca, E. Bolat and M. Toptas, "Could ferritin level be an indicator of COVID-19 disease mortality?," *Journal of medical virology*, 93(3), 2020, pp. 1672-1677.

[7] H. Nyblom, U. Berggren, J. Balldin and R. Olsson, "High ast/alt ratio may indicate advanced alcoholic liver disease rather than heavy drinking," *Alcohol and alcoholism*, 39(4), 2004, pp. 336-339.

[8] I. Serin, N. D. Sari, M. H. Dogu, S. D. Acikel, G. Babur, A. Ulusoy, M. I. Onar, E. C. Gokce, O. Altunok and F. Y. Mert, "A new parameter in COVID-19 pandemic: initial lactate dehydrogenase (ldh)/lymphocyte ratio for diagnosis and mortality," *Journal of Infection and Public Health*, 13(11), 2020, pp. 1664-1670.

[9] D. Israni and H. Mewada, "Identity Retention of Multiple Objects under Extreme Occlusion Scenarios using Feature Descriptors," *Journal of Communications Software and Systems (JCOMSS)*, 14(4), 2018, 290-301.

[10] D. Israni and H. Mewada, "Feature descriptor based identity retention and tracking of players under intense occlusion in soccer videos," *International Journal of Intelligent Engineering and Systems*, 11(4), 2018, pp. 31-41.

[11] N. Purohit and D. Israni, "Vehicle classification and surveillance using machine learning technique," In: 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), India, 2017, pp. 910-914.

[12] P. Israni, "Breast cancer diagnosis (BCD) model using machine learning," International Journal of Innovative Technology and Exploring Engineering, 8(10), 2019, pp. 4456-4463.

[13] Supervised vs. Unsupervised Learning: Key Differences. Available online: https://www.guru99.com/supervised-vsunsupervised-learning.html (accessed on 30 November 2021).

[14] L. P. Kaelbling, M. L. Littman and A. W. Moore, "Reinforcement Learning: A Survey," *Journal of Artificial Intelligence Research*, 4, 1996, pp. 237–285.

[15] G. Monika and M. Bharathi Devi, "Using Machine Learning Approach to Predict COVID-19 Progress," *International Journal for Modern Trends in Science and Technology*, 6(8S): 2020, pp. 58-62.

[16] R. Gupta, G. Pandey, P. Chaudhary and S. K. Pal, "Machine Learning Models for Government to Predict COVID-19 Outbreak," *Digital Government: Research and Practice*, 1(4), 2020, pp. 1-6.

[17] C. Iwendi, A. K. Bashir, A. Peshkar, R. Sujatha, J. Chatterjee, S. Pasupuleti, R. Mishra, S. Pillai, and O. Jo, "COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm," *Frontiers in Public Health*, 8, 2020, pp. 1-9.

[18] S. Makridakis, E. Spiliotis and V. Assimakopoulos, "Statistical and machine learning forecast ing methods: Concerns and ways forward," PloS one, 13(3), 2018, pp. 1-26.

[19] https://www.kaggle.com/einsteindata4u/COVID19, last accessed: 2020/10/15.

[20] A. F. M. Batista, J. L. Miraglia, T. H. R. Donato and A. D. P. C. Filho, "COVID-19 diagnosis prediction in emergency care patients: a machine learning approach," *medRxiv*, (2020).

[21] Y. Sun, V. Koh, K. Marimuthu, O. T. Ng, B. Young, S. Vasoo, M. Chan, V. J. Lee, P. P. De, T. Barkham and R. T. Lin et al., "Epidemiological and clinical predictors of COVID-19," *Clinical Infectious Disease*, 71(15), 2020, pp. 786-792.

[22] Z. Meng, M. Wang, H. Song, S. Guo, Y. Zhou, W. Li, X. Song, Y. Zhou, and Q. Li, "Development and utilization of an intelligent application for aiding COVID-19 diagnosis," *medRxiv*, 2020, pp. 1-21.

[23] L. Wang and A. Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images," *Scientific Reports*, 10(1), 2020, pp. 1-12.

[24] R. Pal, A. A. Sekh, S. Kar and D. K. Prasad, "Neural network-based country wise risk prediction of COVID-19," *Applied Sciences*, 10(18), 2020, pp. 1-16.

[25] D. Liu, L. Clemente, C. Poirier, X. Ding, M. Chinazzi, J. T. Davis, A. Vespignani and M. Santillana, "A machine learning methodology for real-time forecasting of the 2019–2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models," *arXiv*, 2020, pp. 1-23.

[26] C. Bayes and L. Valdivieso, "Modelling death rates due to COVID-19: a Bayesian approach," *arXiv*, (2020).

[27] B. R. Beck, B. Shin, Y. Choi, S. Park and K. Kang, "Predicting commercially available antiviral drugs that may act on the novel coronavirus (2019-nCoV), Wuhan, China through a drug-target interaction deep learning model," *Computational and Structural Biotechnology Journal*, 18, 2020, pp. 784-790.

[28] Z. Tang, W. Zhao, X. Xie, Z. Zhong, F. Shi, T. Ma, J. Liu and D. Shen, "Severity assessment of COVID-19 using CT image features and laboratory indices," *Physics in Medicine and Biology*, 66(3), 2020, pp. 035015.

[29] N. E. M. Khalifa, M. H. N. Taha, A. E. Hassanien, and S. Elghamrawy, "Detection of coronavirus (COVID-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest X-ray dataset," *arXiv*, (2020).

[30] R. Sujatha, J. M. Chatterjee, and A. E. Hassanien, "A machine learning forecasting model for COVID-19 pandemic in India," Stochastic Environmental Research Risk Assessment. 34, 2020, pp. 959-972.

[31] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, "COVIDgan: data augmentation using auxiliary classifier GAN for improved COVID-19 detection," *IEEE Access*, 8, 2020, pp. 91916–91923.

[32] J. Wu, P. Zhang, L. Zhang, W. Meng, J. Li, C. Tong, Y. Li, J. Cai, Z. Yang and J. Zhu et al., "Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results," *MedRxiv*, 2020, pp. 1-12.

[33] A. Bastug, H. Bodur, S. Erdogan, D. Gokcinar, S. Kazancioglu, B. D. Kosovali, B. O. Ozbay, G. Gok, I. O. Turan, G. Yilmaz and C. C. Gonen, "Clinical and laboratory features of COVID-19: Predictors of severe prognosis," *International immunopharmacology*, 88, 2020, 106950.

[34] D. Brinati, A. Campagner, D. Ferrari, M. Locatelli, G. Banfi, F. Cabitza, "Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study," *Journal of medical systems*, 44(8), 2020, pp. 1-12.

[35] A. Mohamed-Hussein, I. Galal, M. M. A. R. Mohamed, H. A. Elaal and K. M. Aly, "Is there a correlation between pulmonary inflammation index with COVID-19 disease severity and outcome?," *medRxiv*, (2020).

[36] Symptoms and COVID Presence (May 2020 data), https://www.kaggle.com/hemanthhari/symptoms-and-COVID-presence?select=COVID+Dataset.csv, last accessed: 2021/11/30.

[37] C. N. Villavicencio, J. J. E. Macrohon, X. A. Inbaraj, J. H. Jeng, and J. G. Hsieh, "COVID-19 Prediction applying supervised machine learning algorithms with comparative analysis using WEKA," *Algorithms*, 14(7), 2021, pp. 201.