

## Research Article

# COVID-19 Tweets Classification during Lockdown Period Using Machine Learning Classifiers

Syed Ali Jafar Zaidi <sup>1</sup>, Indranath Chatterjee <sup>2</sup>, and Samir Brahim Belhaouari <sup>3</sup>

<sup>1</sup>*Institute of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan*

<sup>2</sup>*Department of Computer Engineering, Tongmyong University, Busan 48520, Republic of Korea*

<sup>3</sup>*Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*

Correspondence should be addressed to Indranath Chatterjee; [indranath.cs.du@gmail.com](mailto:indranath.cs.du@gmail.com) and Samir Brahim Belhaouari; [sbelhaouari@hbku.edu.qa](mailto:sbelhaouari@hbku.edu.qa)

Received 9 April 2022; Revised 31 May 2022; Accepted 15 June 2022; Published 7 July 2022

Academic Editor: Imran Ashraf

Copyright © 2022 Syed Ali Jafar Zaidi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, COVID-19 has been regarded as the most dangerous pandemic for several countries. On various social media platforms, such as Twitter, Facebook, and Instagram, a variety of rumours, hypes, and news are published. This might have a detrimental impact on people's life. As a result, social media platforms have always had a difficult time authenticating this fake information. Different machine learning (ML) and deep learning (DL) classifiers were used in this work to categorize the continuing impacts of tweets and forecast their after-effects. Support vector machine (SVM), random forest (RF), decision tree (DT), and k-nearest neighbor (KNN) were used for classification, while AdaBoost and convolutional neural network (CNN) were utilized for future effects. The tweets dataset from Kaggle was used to train the SVM, RF, KNN, and DT models, which were then assessed on multiple evaluation criteria such as accuracy, precision, recall, and F1-score, using a 70:30 ratio. The CNN and AdaBoost, on the other hand, have been taught to detect the mean square error, root mean square error, and mean absolute error. With 0.74 and 0.73 percent score out of 1, respectively, RF and SVM exhibit the best accuracy in impact when classifying the outcomes on the obtained dataset. In terms of a regression problem, CNN beat the ADA Regressor across the board.

## 1. Introduction

Since the end of 2019, a pandemic illness, also named as coronavirus disease 2019 (COVID-19) by the World Health Organization (WHO), has spread fast, affecting people all over the world. COVID-19 infection originated in China and has spread around the world [1]. Since it is a pandemic illness, it spreads from person to person. Thousands of people have died as a result of this strange virus, which has been documented in millions of instances. This is a terrible news for everyone, especially for governments all across the world [2]. The challenge for all countries was how to stop the virus from spreading and how to save their country, but the countries most affected by the pandemic, such as China, America, India, and others, chose to place their countries on absolute

lockdown [3]. All nations decided to adopt a partial or total lockdown all of a sudden, and an estimated 10 million people were infected [4].

This quarantine may have slowed the spread of the COVID-19 virus, but it has also caused a plethora of other problems, such as unemployment, poverty, inflation, and other socioeconomic issues [5]. The authors were prompted to investigate the influence of lockdown on human behaviour due to the abundance of categorization using various machine learning methods on this issue. According to some estimates, the new virus and lockdown have resulted in a large increase in unemployment, with an estimated 10 million unemployed people globally [6]. The lockdown had an impact on the human brain since everyone was coping with numerous domestic and societal issues. Apart from this,

numerous people enjoyed spending time with their families during the lockdown and welcomed the government's choice. Meanwhile, not everyone had access to the Internet or had the ability to interact with their family and friends [7].

Making decisions to address the issues was incredibly difficult in this situation. Keeping in mind that this research has concentrated on people's views and on determining the impact of the lockdown, this might be relevant in decision-making. People use social media sites such as Twitter, Facebook, Instagram, and others to express themselves [8, 9]. In terms of health, COVID-19 has posed a threat to everyone's life and health. Despite extensive research on COVID-19 vaccination, the health hazards remain the same [10]. Despite the fact that COVID-19 is a relatively new topic, several contributions have been made to various elements of COVID-19 [11]. COVID-19 has now made to conduct a substantial study into deep learning, artificial intelligence, and machine learning in order to answer a variety of real-world problems. The following is the study's contribution to the literature:

- (1) To present a system for the tweeter to use on Twitter to recognise and resolve stress or panic situations
- (2) To develop cutting-edge machine learning classifiers and determine the impact of COVID-19 on humans
- (3) To determine the public's attitude regarding the epidemic in order to make decision-making considerably easier
- (4) The suggested approach might assist other areas in identifying the overall feedback in a single position
- (5) To debunk falsehoods and purge misleading material from social media or to forecast people's behaviour in the event of a pandemic

The goal of this study is to deliver the data quickly so that the decision-making process can be improved. For this reason, the issue has been handled as a regression and classification challenge. Since the goal of this study is to determine the effect and trend of COVID-19, machine learning classifiers such as support vector machine (SVM) [12] and random forest (RF) [13], k-nearest neighbor (KNN), and decision tree (DT) were used. For feature extraction, the emotions of the tweets were acquired in terms of negative, positive, and neutral tweets. To examine the trend and impact, the researchers used data from the tweets on COVID-19 that were sent out during the epidemic. Twitter has been used to classify and arrange the thoughts because it is considered one of the most authentic social media outlets. Another aspect is that many remained at home throughout the lockdown and were continuously on the Internet. Figure 1 shows a sample graph of a huge number of tweets that have been sent on Twitter from various regions.

Despite the fact that the dataset includes tweets related to COVID-19, it reflects people's opinions in a different context and from a variety of backgrounds. The data collection contains tweets from all across the world, as previously indicated. As a result, we have about 179,108 tweets from

throughout the world that were tweeted throughout the outbreak. 70% of the tweets were trained using SVM, KNN, RF, and DT. In contrast, 30% of the data was evaluated for classification issues using assessment metrics such as accuracy, recall, precision, and F1-score. The regression analysis was evaluated using the mean square error, square error, and the mean absolute error evaluation parameters.

## 2. Related Work

People of different ages and backgrounds, as well as those from diverse geographical places, are affected by the negative impacts of this rare disease. Many techniques, such as deep learning, machine learning, and other AI approaches, have been developed to estimate the impact of COVID-19. The machine learning classifiers are essential for resolving a variety of regression and classification issues. Different classifiers have different benefits and drawbacks. To categorize the data into numerous classifications, the SVM and the random forest algorithm are investigated. The most important aspect of categorization is grouping the material into several different categories [14]. Although predictive analysis is considered cutting-edge, it still requires human involvement to gather the data and other resources that the classifier will use to make predictions. The Titanic survival prediction is 83.5 percent accurate using SVM, logistic regression, and linear regression [15]. The support vector machine was shown to be the best prediction approach for the future lever sickness prediction when compared to the naive Bayes in another study evaluating multiple classifiers for predicting the future lever illness [16].

Binary categorization, on the other hand, is done with the help of support vector machine (SVM) [17]. First, the RF algorithm extracts subsamples from the original data using the bootstrap resampling method and creates decision trees for each sample. The random forest (RFs) nomenclature suggests that it consists of multiple decision trees that are utilized to produce predictions, as previously stated. Second, the approach categorizes the decision trees and performs a simple vote, with the classification receiving the most votes determining the prediction's outcome [17].

A classification is fundamentally a sort of machine learning, and the random forest classifier has been utilized all around the world. Apart from this, a number of other ways and strategies have been tested in a variety of settings. For example, the wavelet transform (WT) has been widely used to extract features [18]. These techniques extract properties from signals based on their frequency of occurrence.

The world has witnessed how the technology revolution has changed people and their surroundings [19, 20]. Deep neural networks [21] have had a considerable impact on the real-world applications that spans a larger region and are more complicated. In the machine learning discipline, deep learning has been widely utilized to target natural language sentiment analysis and natural language processing [20]. Deep neural networks, on the other hand, break issues down into layers and are regarded as a great tool for extracting valuable clues for more accurate future predictions [21]. The sentiment analysis [22] was also utilized to measure recall,

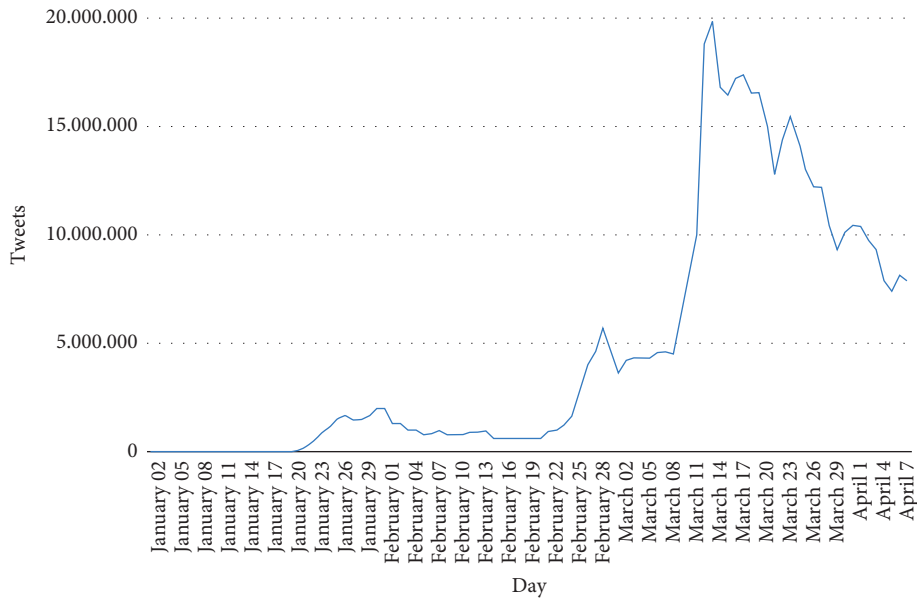


FIGURE 1: Tweets about COVID-19 during the last 4 months.

precision, and accuracy. The accuracy and precision of the findings are 0.86 and 0.827, respectively.

In this research [23], the naive Bayes, support vector machine, and linear regression were implemented on real-time data collection. The MSE and MAE were employed as assessment metrics in this study. At the same time, studies revealed that the naive Bayes approach produces results that are almost identical to real-time COVID-19 illness information. They were investigating the influence of COVID-19 on picture data sets using various artificial intelligence (AI) approaches, in addition to textual data. Finding the impact and disease through the images is a significant gap in the literature to address in the study [24]. The Pyspark machine learning model has also been used to determine the impact by using different classifiers in the study [25]. The study's findings suggest that logistic regression is one of the top classifiers, outperforming others such as the naive Bayes, random forest, and decision tree.

COVID-19 has spread so quickly that it has wreaked havoc on the lives of everyone in the world. As a result, people began to spread varied information on social media based on misconceptions, faulty information, and fake facts in order to portray depressed people in poor light. As a consequence, this study proposes a categorization-based model for determining people's influence on various social media sites and a model based on current data for predicting the future effect utilizing a regression problem. As a result, identifying and correcting the cause of the hype may become easy.

### 3. Materials and Methods

**3.1. Data Details.** The dataset for this study came from Kaggle [26], and it contains 179108 tweets from a variety of people. The users name, tweet date, and COVID-19 impact tweet were all included in the data collection. The user's tweets were

preprocessed before the experiment to remove stop words, special characters, and symbols that might cause the polarity of the tweets to worsen. The polarity of the tweets was tested using Python's Textblob module after all of the preprocessing and reorganisation of the data collection was completed. The polarity rate can be used for sentiment analysis.

**3.1.1. Used Platform for Implementation.** The key experiments were carried out on the Google Collaborator, which is a cloud-based Google product with 2.2 GHz CPU, 13 GB RAM, and 108 GB ROM.

**3.2. Methodology.** This study's main purpose is to provide a mechanism for social media platforms such as Twitter, Facebook, Instagram, and Tumbler to counter social media users' hatred and enthusiasm about a certain topic or event [27]. This is done using traditional machine learning classifiers including the k-nearest neighbor (K-NN), support vector machine (SVM), random forest (RF), decision tree (DT), and artificial neural network (ANN). For testing reasons, the entire dataset has been divided into two parts: training and testing datasets. 70 percent of the dataset was utilized to train the model, while 30 percent was used for testing with the assistance of several categorization assessment criteria for machine learning techniques such as accuracy, precision, recall, and F1-score. Although, for a regression task, the mean square error (MSE), R-square, root mean square error (RMSE), and mean absolute error (MAE) have been used to examine deep learning algorithms (MAE). Following the testing, a comparison study was conducted comparing all the machine learning and deep learning methodologies used in this literature to solve regression and classification issues. Figure 2 depicts a simplified visual depiction of the technique, along with various phases of the proposed system.

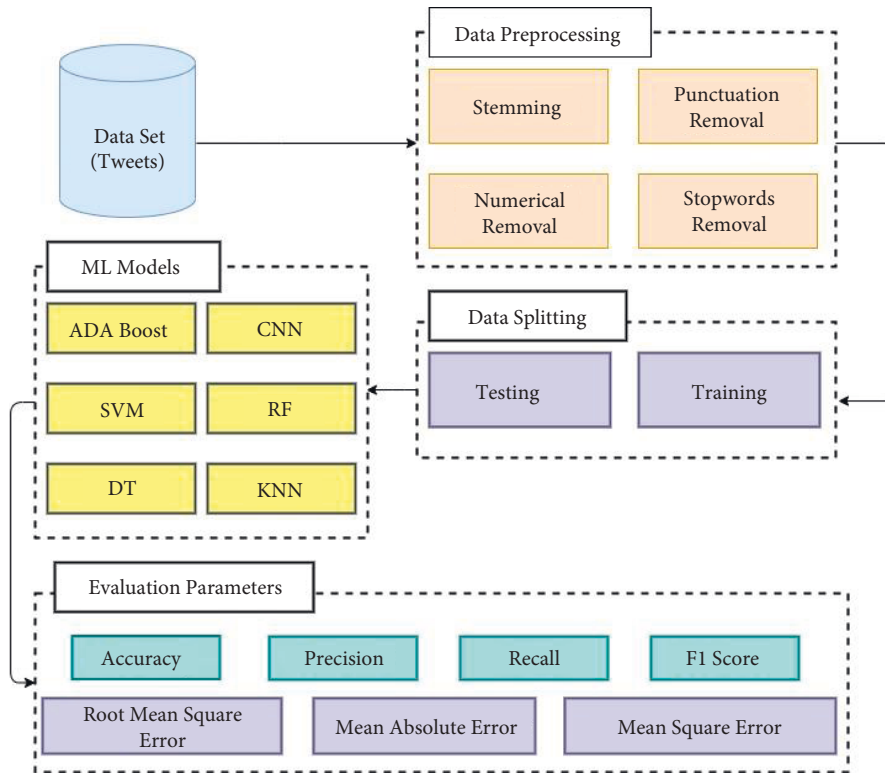


FIGURE 2: Methodology diagram.

3.2.1. *Theoretical Background.* In this literature, machine learning methods have played a critical role [28]. The study of exploiting patterns and experiences to improve results is known as machine learning. Machine learning approaches, both supervised and unsupervised, have been used extensively for many possibilities [29]. For future impacts, this work has used supervised machine learning and two deep learning algorithms. The five different models utilized for the classification task were the k-nearest neighbor (K-NN), the support vector machine (SVM), the random forest (RF), and the decision tree (DT). For regression challenges, the AdaBoost regressor and CNN classifier were employed; all of them are listed as follows.

(1) *K-Nearest Neighbor (KNN).* The K-NN supervised machine learning classifier [30] is a well-known supervised machine learning classifier that may be used to address regression and classification problems. K-NN works by assuming that all data points in close proximity are of the same type. The KNN classifier's purpose is to find the closest neighbour class in order to forecast the target value. Though KNN has the advantage of being simple to understand and deal with nonlinear data, it has a lower accuracy rate than other approaches and uses more storage space since it requires all of the training data to be present [31].

(2) *Support Vector Machine (SVM).* SVM is a versatile and powerful classifier that may be applied to regression and classification issues. As a consequence, numerous items are divided into discrete classes and categories by a hyperplane [32].

(3) *Decision Tree.* One of the most often used classifiers for regression and classification challenges is the decision tree. The CART (classification and regression tree) is used in decision trees to produce judgments based on the requirements and attributes. In a decision tree, inside nodes are viewed as conditions, whereas leaf-nodes are treated as decisions. A decision tree can be useful for showing all of the tree's various choices graphically [33].

(4) *Random Forest.* Random forest, like other traditional machine learning algorithms, is a supervised machine learning approach. It is one of the most widely used machine learning algorithms due to its efficiency and adaptability. The decision trees which are used to forecast individual outcomes are collected by the random forest classifier. The random forest collects the predictions from each decision tree and concludes by voting [11].

(5) *AdaBoost Regressor.* The AdaBoost regressor (ABR) is both a regression and an ensemble classifier. It uses weak learners that are very simple yet have some dataset skills to anticipate occurrences. While prediction employs the addition of a decision tree to the model in a sequential manner, each model employs the prediction of the prior model before the current model [34].

(6) *Convolution Neural Network (CNN).* CNN is a deep learning classifier that is primarily known for its image implementation, while text may also be used to improve the results [35]. In terms of future prediction, our dataset has also been trained on CNN. CNN is far more successful at text

categorization since it uses various sizes and shape filters to compress the original sentence matrix to a smaller size matrix [36].

(7) *Evaluation Matrices for Classification.* There are several categorization assessment matrices [37]. For the assessment mentioned in the following, we used accuracy, precision, recall, and F1-score.

3.2.2. *Accuracy.* It is an evaluation parameter that is defined as the data forecast from the complete dataset. False positives (FP), false negatives (FN), true positives (TP), and true negatives (TN) are split by true positives (TP) and true negatives (TN), respectively [38]:

$$\text{accuracy} = \frac{\text{true positives} + \text{true negatives}}{\text{all samples}}. \quad (1)$$

3.2.3. *Precision.* The precision is the number of positive predictions that are unmistakably in the positive category. Precision [39, 40] can be defined as the ratio of true positive to all positive class values in the data.

$$\text{Precision} = \frac{\text{true positives (TP)}}{\text{true positives (TP)} + \text{false positives (FP)}}. \quad (2)$$

3.2.4. *Recall.* The recall can also be called as sensitivity which is the ratio of positive class prediction and all the positive instances in the class or the dataset [41].

$$\text{Recall} = \frac{\text{true positives (TP)}}{\text{true positives (TP)} + \text{false negatives (FN)}}. \quad (3)$$

3.2.5. *F1-Score.* The F1-score is also known as the F-score or F-measure. Precision and recall are both considered by the F1-score. However, the F1-score is the harmonic mean of precision and recall. It gives the best results when there is some balance in precision and recall [42].

$$F1 - \text{score} = 2 * \frac{\text{precision} * \text{recall}}{\text{recall} + \text{precision}}. \quad (4)$$

3.2.6. *Evaluation Parameters for Regression.* This section determines the evaluation factors used to estimate the regression of the future impact of new COVID-19 illness. For this purpose, the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and R-squared score were used.

3.2.7. *Mean Square Error (MSE).* This is the most important evaluation parameter in the regression problem. To anticipate the predictor's quality, the mean square error is employed to assess the difference between the forecast and the ground actuality. The resultant value was squared, and an average over the dataset was calculated [43, 44]. Because it is

always a square root, the outcome cannot be zero. The mean square error equation is as follows:

$$\text{MSE} = \frac{1}{N} \sum_{i=0}^n (y - \hat{y})^2. \quad (5)$$

3.2.8. *Root Mean Square Error (RMSE).* The root mean square mistake, also known as the root mean square deviation, is a straightforward technique to assess the model's error in predicting the data's residuals. The residual standard deviation (RMSE) can be defined as the standard deviation of the residuals. It determines how near the data is to the best fit line. For regression analysis, forecasting, and climatology, RMSE is commonly utilized [45].

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=0}^n (y - \hat{y})^2}. \quad (6)$$

3.2.9. *Mean Absolute Error.* It is the difference between the expected and the actual numbers, and it is used to figure out where the forecast went wrong. The MAE is used to anticipate and predict the deep learning classifiers, with the resultant value ranging from 0 to infinite [45].

$$\text{MAE} = \frac{1}{N} \sum_{i=0}^n |y - \hat{y}|. \quad (7)$$

## 4. Experiment and Results

The goal of this study is to find out how the unique COVID-19 sickness affects people's thoughts. After the impact has been discovered, dealing with any future hype that may arise on social media platforms will be straightforward. Despite the fact that everyone is familiar with COVID-19, it has had distinct impacts on human brains all around the world and will surely cause serious difficulties in the future. Several academics have contributed to the literature on the COVID-19's influence, which may be measured by the sentiment analysis of tweets or feedback, as well as artificial intelligence implementation. Classic machine learning techniques and deep learning approaches were used in a variety of studies. Precision, recall, F1-score, and accuracy have all been used to train and assess the traditional machine learning classifiers. The regression issue dataset, on the other hand, was examined using typical deep learning classifiers and the mean square error (MSE), the R-square, the root mean square error (RMSE), and mean absolute error (MAE) metrics. Finally, all of the findings were compared, and one of the best classifiers was chosen.

4.1. *Accuracy, Precision, Recall, and F1-Score.* The accuracy, precision, recall, and F1-score were analyzed and tested on the trained model using typical machine learning classifiers in this part. Figure 3 depicts the overall results obtained as a consequence of this research.

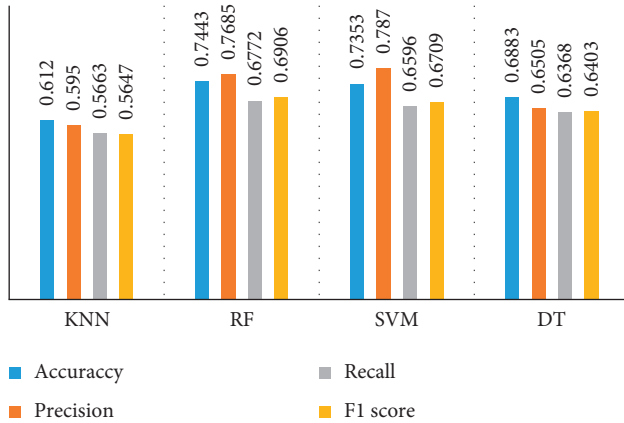


FIGURE 3: Results after evaluation with all the classifiers used in the study.

With typical machine learning classifiers, the result obtained is extremely fascinating. With all of the testing assessment matrices, KNN yields the worst results. The KNN accuracy is 0.61197, with an accuracy of 0.5950, a recall of 0.5562, and an F1-score of 0.5647, the lowest of all the matrices evaluated. In several instances where the KNN has been employed with classification issues in machine learning, it has produced poor results [46]. When it comes to classification challenges, the best classifiers are random forest (RF) and support vector machine (SVM). Although random forest and precision produce the greatest results, the SVM produces best outcomes, as seen in Figures 2 and 3, respectively.

The random forest provides the greatest results (0.7443), followed by the support vector machine (0.7352) and the decision tree (0.6882), as shown in Figure 4. K-nearest neighbor, on the other hand, has the poorest results, with 0.61197 accuracies.

The precision findings in Figure 5 reveal that the support vector machine surpassed all other classifiers with a score of 0.7869, followed by random forest with a score of 0.7684 and the decision tree with a score of 0.65051. Because the recall and F1-score of any classifier evaluated did not reach 0.70, the accuracy and precision findings demonstrate that KNN is not a viable classifier for the proposed system.

As a result of the distinct classifiers, the outcome has a varied trend. However, as demonstrated in Figures 3–5, the overall findings suggest that KNN is not as successful as other classifiers. At the same time, the random forest surpassed all other classifiers in accuracy and precision, whereas recall and F1-score are shown in Table 1, where the results are similar in both the cases of precision and accuracy.

Table 1 shows that the F1-score and random forest recall fared best, followed by the SVM and decision tree. At the same time, the KNN results failed to meet any assessment metric's criterion.

*4.1.1. Evaluation with the Deep Learning Matrix.* The future impact of innovative COVID-19 was predicted using a deep learning assessment matrix. Table 2 gives a brief summary of each assessment matrix.

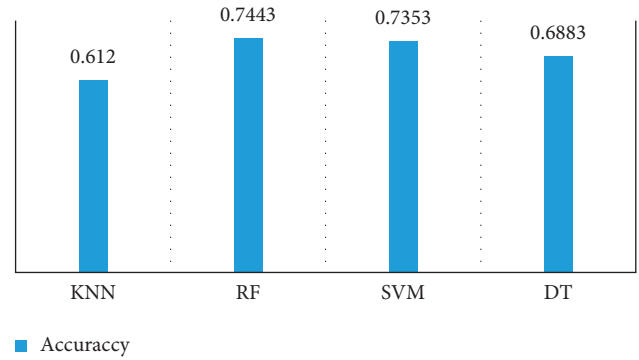


FIGURE 4: Accuracy with different classifiers.

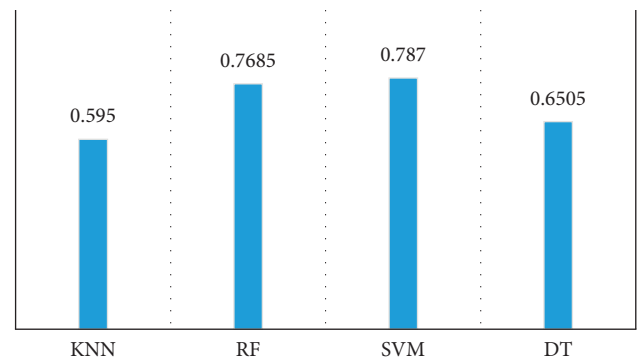


FIGURE 5: Precision with different classifiers.

TABLE 1: The F1-score and recall on different machine learning models used in the study.

Classifiers	F1-score	Recall
KNN	0.56472	0.566294
Random forest	0.690609	0.67716
SVM	0.670871	0.659631
Decision tree	0.640292	0.636801

TABLE 2: CNN and AdaBoost regressor results on MSE, RMSE, and MAE evaluation parameters.

	CNN	AdaBoost regressor
Mean square error	1.494219	0.5478916
Root mean square error	1.222382	0.740197
Mean absolute error	1.042075	0.60803

In the table above, the findings of CNN reveal that it performs better in all of the deep learning assessment parameters that were examined. The RMSE displays the best outcomes for the AdaBoost regressor while also providing the best results.

## 5. Conclusions

COVID-19's total impact has proven to be a difficult topic for social media platforms to navigate. Social media platforms may be the finest way for people to quickly communicate and share their sentiments and views throughout

the world [47]. Because of the possibilities of a full product being so bleak, much work has been done to remedy the various flaws of social networking sites [48]. However, the lack of categorization using various machine learning algorithms prompted the authors to enhance the performance and identify the influence of social media sites, and this study offered a system to do so. On a tweets dataset retrieved from Kaggle, several state-of-the-art machines and deep learning approaches were trained. The impact has been classified using accuracy, precision, recall, and F1-score. The best results are shown by RF and SVM, respectively, whereas KNN's classification performance was not up to the mark in any situation. When it comes to regression difficulties, however, CNN outperforms the AdaBoost. According to this study, the random forest classifier performs well across all testing assessment criteria, with above 0.72 percent outcomes when using an average strategy, as well as SVM, which may be directly matched to the CNN classifier. However, the findings of this study are intriguing and can be improved with hybrid machine learning approaches. Simultaneously, putting the suggested method into a real-time context would be a fantastic contribution to social media sites.

### Data Availability

The dataset used in this study can be provided on request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### Acknowledgments

The authors would like to thank the Qatar National Library (QNL) for its support to conduct this research work.

### References

- [1] D. L. Blustein, R. Duffy, J. A. Ferreira, V. Cohen-Scali, R. G. Cinamon, and B. A. Allan, "Unemployment in the time of COVID-19: a research agenda, journal of vocational behaviour," *Journal of Vocational Behavior*, vol. 119, 2020.
- [2] B. Pfefferbaum and C. S. North, "MD, MPE mental health and the covid-19 pandemic, the new England journal of medicine," *The New England Journal of Medicine*, vol. 383, 2020.
- [3] J. Bedford, D. Enria, J. Giesecke et al., "COVID-19: towards controlling of a pandemic," *Lancet*, vol. 395, pp. 1015–1018, 2020.
- [4] I. Mandal and S. Pal, "COVID-19 pandemic persuaded lockdown effects on environment over stone quarrying and crushing areas," *Science of the Total Environment*, vol. 732, 2020.
- [5] E. Mahase, "Covid-19: how does local lockdown work, and is it effective?" *BMJ*, vol. 370, 2020.
- [6] M. A. Fullana, D. Hidalgo-Mazzei, E. Vieta, and J. Radua, "Coping behaviors associated with decreased anxiety and depressive symptoms during the COVID-19 pandemic and lockdown," *Journal of Affective Disorders*, vol. 275, pp. 80–81, 2020.
- [7] Y. A. Helmy, M. Fawzy, A. Elswad, A. Sobieh, S. P. Kenney, and A. A. Shehata, "The COVID-19 pandemic: a comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control," *Journal of Clinical Medicine*, vol. 9, no. 4, 2020.
- [8] Y. Asim, A. K. Malik, B. Raza, A. R. Qamar, and N. Qamar, "Predicting influential blogger's by a novel, hybrid and optimized case based reasoning approach with balanced random forest using imbalanced data," *IEEE Access*, vol. 9, pp. 6836–6854, 2021.
- [9] E. Lenihan, "A classification of Antifa Twitter accounts based on social network mapping and linguistic analysis," *Social Network Analysis and Mining*, vol. 12, no. 1, p. 12, 2022.
- [10] W. Zhang, M. Zhu, and J. Zhu, "An efficient hierarchical identification method with kernel-based SVM for equivalent systems of aircrafts," *IEEE Access*, vol. 7, pp. 83243–83250, 2019.
- [11] T. B. Lopez-Garcia, J. A. Coronado-Mendoza, and A. Domínguez-Navarro, "Artificial neural networks in microgrids: a review," *Engineering Applications of Artificial Intelligence*, vol. 95, 2020.
- [12] V. Chamola, V. Hassija, V. Guizani, and M. Guizani, "A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact," *IEEE Access*, vol. 8, pp. 90225–90265, Article ID 2992341, 2020.
- [13] R. S. Patil and S. R. Kolhe, "Inflectional and derivational hybrid stemmer for sentiment analysis: a case study with Marathi tweets," *Recent Trends in Image Processing and Pattern Recognition. RTIP2R 2021*, vol. 1576, 2022.
- [14] S. P. Swayanshu, "Implementation of logistic regression & SVM for titanic survival prediction using python (IJSRD)," *International Journal for Scientific Research & Development*, vol. 6, 2018.
- [15] R. Kalaiselvi, K. Meena, and V. Vanitha, "Liver disease prediction using machine learning algorithms, 2021," in *Proceedings of the International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, pp. 1–6, Coimbatore, India, 2021.
- [16] M.-W. Huang, C.-W. Chen, W.-C. Lin, S.-W. Ke, and C.-F. Tsai, "SVM and SVM ensembles in breast cancer prediction," *PLoS One*, vol. 12, Article ID e0161501, 2017.
- [17] X. Gao, J. Wen, and C. Zhang, "An improved random forest algorithm for predicting employee turnover," *Mathematical Problems in Engineering*, vol. 2019, 12 pages, Article ID 4140707, 2019.
- [18] H. A. G. Galal Elsayed, S. Chaffar, S. B. Brahim Belhaouari, and H. Raissouli, "A two-level deep learning approach for emotion recognition in Arabic news headlines," *International Journal of Computers and Applications*, pp. 1–10, 2020.
- [19] Z. Long, R. El Saddik, and A. El Saddik, "NeedFull-a tweet analysis platform to study Human needs during the COVID-19 pandemic in New York state," *IEEE Access*, vol. 8, pp. 136046–136055, 2020.
- [20] S. B. Belhaouari and H. Raissouli, "Madl: a multilevel architecture of deep learning," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, p. 693, 2021.
- [21] S. Bhatia and J. Malhotra, "Naïve Bayes classifier for predicting the novel coronavirus," in *Proceedings of the Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, Jaipur, India, 2021.
- [22] H. Krishnan, P. G. A. Poosari, A. Jayaraj, C. Thomas, and G. M. Joy, "Machine learning based sentiment analysis of coronavirus disease related twitter data," in *Proceedings of the*

- 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), Jalandhar, India, 2021.
- [23] S. T. H. Kieu, A. Bade, M. H. A. Hijazi, and H. Kolivand, "COVID-19 detection using integration of deep learning classifiers and contrast-enhanced canny edge detected X-ray images," *IT Professional*, vol. 23, no. 4, pp. 51–56, 2021.
- [24] N. Iksan, D. A. Widodo, B. Sunarko, E. D. Udayanti, and E. Kartikadharma, "Sentiment analysis of public reaction to COVID19 in twitter media using naïve Bayes classifier," in *Proceedings of the IEEE International Conference on Health, Instrumentation & Measurement, and Natural Sciences (InHeNce)*, pp. 1–4, Medan, Indonesia, 2021.
- [25] "COVID19 Tweets," 2021, <https://www.kaggle.com/gpreda/covid19-tweets>.
- [26] M. Ramzan, S. Mehta, and E. Annapoorna, "Are tweets the real estimators of election results?" in *Proceedings of the Tenth International Conference on Contemporary Computing (IC3)*, pp. 1–4, Noida, India, 2017.
- [27] A. Ali, S. Qadri, W. K. Mashwani et al., "Machine learning approach for the classification of corn seed using hybrid features," *International Journal of Food Properties*, vol. 23, no. 1, pp. 1110–1124, 2020.
- [28] H. R. H. Al-Absi, M. A. Refaee, A. U. Rehman, M. T. Islam, S. B. Alam, and T. Alam, "Risk factors and comorbidities associated to cardiovascular disease in Qatar: a machine learning based case-vontrol study," *IEEE Access*, vol. 9, pp. 29929–29941, 2021.
- [29] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "KNN model-based approach in classification," *Lecture Notes in Computer Science*, vol. 2888, 2003.
- [30] D. O. McInerney McInerney and M. Nieuwenhuis, "A comparative analysis of kNN and decision tree methods for the Irish national forest inventory," *International Journal of Remote Sensing*, vol. 30, no. 19, pp. 4937–4955, 2009.
- [31] M. E. Ozer, P. O. sarica, and K. Y. Arga, "OMICS," *A Journal of Integrative Biology*, vol. 24, no. 7, 2020.
- [32] P. Chandrasekar, K. Qian, H. Shahriar, and P. Bhattacharya, "Improving the prediction accuracy of decision tree mining with data preprocessing," *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, vol. 1, pp. 481–484, 2017.
- [33] L. Demidova and M. Ivkina, "Approach to determining the boundaries of the search range for the number of trees in the random forest algorithm," in *Proceedings of the 9th Mediterranean Conference on Embedded Computing (MECO)*, pp. 1–4, Budva, Montenegro, 2020.
- [34] S. Mishra, D. Mishra, and G. H. Santra, "Adaptive boosting of weak regressors for forecasting of crop production considering climatic variability: an empirical assessment," *Journal of King Saud University—Computer and Information Sciences*, vol. 32, no. 8, pp. 949–964, 2020.
- [35] L. F. O. Chamon, G. J. Pappas, and A. Ribeiro, "The mean square error in Kalman filtering sensor selection is approximately supermodular," in *Proceedings of the 2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 343–350, Melbourne, Australia, 2017.
- [36] D. Elavarasan, D. R. Vincent, V. Sharma, A. Y. Zomaya, and K. Srinivasan, "Forecasting yield by integrating agrarian factors and machine learning models: a survey," *Computers and Electronics in Agriculture*, vol. 155, pp. 257–282, 2018.
- [37] A. Polyvyanyy, A. Solti, M. Weidlich, C. D. Ciccio, and J. Mendling, "Monotone precision and recall measures for comparing executions and specifications of dynamic systems," *ACM Transactions on Software Engineering and Methodology*, vol. 29, no. 3, p. 41, 2020.
- [38] W. M. Bramer, D. Giustini, and B. M. R. Kramer, "Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and google scholar: a prospective study," *Systematic Reviews*, vol. 5, p. 39, 2016.
- [39] L. Thompson, J. Khuc, M. S. Saccani, N. Zokaei, and M. Cappelletti, "Gamma oscillations modulate working memory recall precision," *Experimental Brain Research*, vol. 239, no. 9, pp. 2711–2724, 2021.
- [40] M. Gjoreski, M. Z. Gams, M. Lustrek, P. Genc, J. U. Garbas, and T. Hassan, "Machine learning and end-to-end deep learning for monitoring driver distractions from physiological and visual signals," *IEEE Access*, vol. 8, pp. 70590–70603, 2020, 2020.
- [41] K. Takahashi and T. Fujii, "Least MSE regression for view synthesis," in *Proceedings of the 2014 2nd International Conference on 3D Vision*, pp. 385–392, Tokyo, Japan, 2014.
- [42] S. A. Marhon, C. J. F. Cameron, and S. C. Kremer, "Handbook on neural information processing," *Intelligent Systems Reference Library*, vol. 49, 2013.
- [43] N. M. Rezk, M. Purnaprajna, T. Ul-Abdin, and Z. Ul-Abdin, "Recurrent neural networks: an embedded computing perspective," *IEEE Access*, vol. 8, pp. 57967–57996, 2020.
- [44] S. A. J. Zaidi, S. Tariq, and S. B. Belhaouari, "Future prediction of COVID-19 vaccine trends using a voting classifier," *MDPI*, vol. 6, no. 11, 2021.
- [45] T. O. Hodson, "Root mean square error (RMSE) or mean absolute error (MAE): when to use them or not," *Model Development Discussions*, 2022.
- [46] N. Ali, D. Neagu, and P. Trundle, "Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets," *SN Applied Sciences*, vol. 1, no. 12, p. 1559, 2019.
- [47] K. Nimala, R. Jebakumar, and M. Saravanan, "Retraction note to: sentiment topic sarcasm mixture model to distinguish sarcasm prevalent topics based on the sentiment bearing words in the tweets," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, 2022.
- [48] J. F. de Oliveira, H. T. Marques-Neto, and M. Karsai, "Measuring the effects of repeated and diversified influence mechanism for information adoption on Twitter," *Social Network Analysis and Mining*, vol. 12, no. 1, p. 16, 2022.