Implementation of Novel Deep Learning Model for Covid Opinion Mining In Social Media

[1]Karthikgowda H S, [2] Dr. Sasidhar Babu Suvanam,[3] Dr. Vijayalakshmi Yellepeddi

School of Computer Science and Engineering, REVA University, Bangalore

Department of Computer Science, Bishop Cotton Academy of Professional Management, Bangalore

[1]karthikgowdahs27@gmail.com, [2]sasidharmails2@gmail.com, [3]vijayasasi11@gmail.com

Abstract.

Without a doubt, (COVID-19) has caused probably the greatest test, all things considered. The continuous COVID-19 pandemic has caused in excess of 150 million contaminated cases and 1,000,000 passing's internationally as of May 5, 2021. Understanding the opinion of individuals communicated in their virtual entertainment remarks can help in checking, controlling, and eventually annihilating the infection. This is a touchy matter as the danger of irresistible illness altogether influences the manner in which individuals think and act in different ways. In this review, we proposed a clever technique in view of the combination of profound learning and traditional managed AI model for computerized extraction of COVID-19-related conversations from online entertainment and a Natural language process (NLP) strategy in light of subject displaying to reveal different issues connected with COVID19 from general assessment with feeling examination of Covid related from web-based Entertainment. Additionally, we examined Covid related popular assessment to all the more likely comprehend the adjustment of the opinion design at various settings. Our discoveries uncover that the Covid pulled in the consideration of individuals from various nations at various times in fluctuating forces. Additionally, the opinion in their tweets is associated to the information and unwavering quality and occasions happened in their nations including the quantity of recently tainted cases, number of recuperations and passings. Also, normal opinion examples can be seen in different nations during the spread of the infection. We accept that different online entertainment stages significantly affect raising individuals' mindfulness about the significance of this illness as well as advancing preventive measures among individuals in the community. Our proposed model result is to carry the Deep learning model with high exactness to perform opinion and phony news location about Covid over web-based entertainment.

Keywords COVID-19; classification, Sentiment Analysis; Topic Modeling; Machine Learning; Deep Learning Natural Language Processing.

1. INTRODUCTION

The Covid illness (COVID-19) has spread quickly all through the world since it was first found in China. The World Health Organization (WHO) announced the COVID-19 flare-up a worldwide wellbeing crisis [1]. In light of the aftereffects of the COVID-19 circumstance report beginning from the authority WHO site on June 1, 2020, the COVID-19 flare-up has brought about in excess of 6,000,000 affirmed cases and more than 371,000 passings worldwide [2]. Research connected with general wellbeing investigation and public discussions on the spread of COVID-19 via virtual entertainment is additionally one of the features of exploration around the world. Online entertainment can spread disinformation about the infection. It was powering alarm and making the supposed infodemics [3].Moreover, online entertainment has for some time been perceived as a significant spreader of wellbeing falsehood [4]. Web-based entertainment use as a wellspring of data isn't managed. It can prompt wellbeing gambles through the spread of paranoid notion, which cause concerns with respect to spread of the COVID-19 paranoid fear on virtual entertainment [5]. Investigating public discussion will help the significant specialists get popular assessment and data holes among them and the public [6], assisting them with creating fitting crisis reaction systems to address existing issues locally during the pandemic [7]. Also, examining opinion examination can give data on the populace's feeling in various context [8].

This study expects to comprehend general wellbeing by investigating the opinion and point displaying of Indonesian public discussions on Twitter about the COVID-19 utilizing the NLP procedure. We applied a few modelling to examine the systemg, like Logistic regression,Random Forest and Deep Learning Model , to acquire the best classification model.

Past investigations of general wellbeing and local area discussions were completed utilizing managed text based information examination. In 2020 [9], Sear et al. dissected the rise and advancement of subjects around COVID-19 on Facebook Pages utilizing Latent Dirichlet Allocation (LDA). In that review, LDA was capable to recognize themes that appear to be legit in an assortment of posts from online social media around the immunization and COVID-19 discussions. It was likewise ready to deal with enormous information, and the brief outcomes were gotten utilizing measurable bunching methods, rather than having to depend on possibly one-sided, slow, and costly human classification class.

Presently the examination question/heading that comes out is to propose an effective profound brain organization model for the opinion order task. So this work centres around point demonstrating and intermittent brain network-based approaches for ABSA. We have picked LDA (Latent Dirichlet assignment), the most famous solo subject model, and LSTM intermittent brain organization. LDA is broadly utilized for solo theme mining, and LSTM ready to handle long haul conditions. Following are the contribution of this paper:

1. A hybrid deep learning model in view of point demonstrating and repetitive neural network is proposed for sentiment examination .

2. A productive multi-facet Bi-LSTM is proposed for opinion order with just two stacked layers for keeping the model less convoluted.

3. Incline climbing-based approach is proposed for tunning as far as possible model to get to the strong the proposed model's accuracy. Set up embeddings like Glove is utilized to encourage capacity furthermore.

4. A Comparative investigation utilizing different datasets is introduced, showing the exhibition improvement of the proposed approach.

The main objective of the proposed work is its proficiency with higher accuracy. Our discoveries shed light on the significance of utilizing public opinions and appropriate computational strategies to get issues encompassing COVID-19 and to direct related navigation. Generally speaking, the paper is organized as follows. To begin with, we give a short prologue to online media social discussions. Conversation of COVID-19-related issues and a few comparative works are given in area II. In area III, we depict the information pre-handling techniques embraced in our exploration, and the NLP and profound learning techniques applied to the COVID-19 remarks information base. Then, we present the outcomes and conversation. At long last, we close and examine future works in light of NLP approaches for investigating the online local area according to the subject of COVID-1.

2.   RELATED WORKS

Since the episode of the Covid pandemic, scientists have talked about its starting point, impacts, and patterns. This segment presents the tweet feeling examination utilizing different ML, DL, and NLP techniques. Separating significant data from boisterous information is a difficult errand. Steps involved in machine and deep learning models are utilized higher side to perform of the classification problem(17).Various social media propagates the news in that twitter plays major role as it used by most user around the world(18).As per the survey outcome from various county in India alone around 26000 tweets is spread about the Covid-19 in the twitter platform. Those opinions collected from the twitter is expression of people emotions towards covid 19 spread and regards to this pandemic, and they didn't perform tests utilizing ML methods.

One more examination zeroed in on the subjects and feelings of individuals communicated about the pandemic on covid is spread on twitter which is highly negative sentiment sometimes it also paper neutral opinion, at that point, investigated these tweets for feeling characterization utilizing different capabilities and classifiers. This work utilized just a single assessment metric, which is precise, and acquired the most elevated exactness utilizing BERT model ie Bidirectional Encoder Representations from Transformers which provides the accuracy of which is 93%. Some of the time, we use just characterization precision to survey our model's presentation; however this process lacking to proceed with proposed classification model utilizing accuracy, review, and F1-score alongside exactness (1).One more examination work zeroed in on the mental impact about the pandemic due to disaster of the real characteristic of nature in that aspect nature wins the human behaviour towards the nature way of behaving (20). It investigated that individuals are in emergency due to Covid and expanded nervousness levels due to COVID-19 news.

Different examinations show investigation about the modern emergency and monetary effect of the COVID-19 emergency across businesses and nations (21). Throughout the course of recent years, feeling examination in light of tweets has been used in various applications because of the enormous measure of information gathered from different virtual entertainment stages (22). It incorporates Twitter, Facebook, Reddit, and YouTube. The examination shows blemishes in the gathered data (23). Different ML and DL classifiers test the short and long text data. For assessment of a short text, calculated relapse, and Naive Bayes give normal aftereffects of 74 and 91%, individually, however on account of long text testing, both the models performed ineffectively (24). As of late, individuals have been intensely reliant upon web-based entertainment news, and they are conveying their perspectives, feeling, and sentiments about this original infection through virtual entertainment posting (25).

The new COVID-19 examinations depend on general assessment, feeling, and opinion investigation in English web-based entertainment opinion over the media with social communication. Our proposed model facilitate with deep learning model with pre trained data using LSTM model to bring the best F1-score and accuracy when compare with the other machine learning algorithms like random forest, logistic regression for opinion feeling recognition (26). Online entertainment stages, for example, Reddit permits medical care specialist organizations to gather information connected with general assessments, which can be utilized for human conduct examination and information revelation. This study presents a conventional methodology in view with respect to natural language about the covid-19 opinion to achieve the classification by Zhang et al. (28) at a post level over the media. This system relates all the sentiment about the opinion with wordnet strategies to better assessment about the classification model.so this proposed strategies mainly focused on evaluating the sentences Eextricated highlights is been measured by the recurrence during the audits. Other creators give synopsis in light of highlights. They recognized the main elements such as F1 score 83.4% over the dataset using LSTM modelfrom the dataset and accomplished the best F1-score of 83.6% utilizing the BILSTM model. Recurrent NN model implemented with natural language system to get the better investigation in full duplex way according to Mukherjee et al. (29).Using Deep learning with NLP has lots of work has achieved to extract the opinion of the tweets and evaluate the performance analysis on the classification results (30).Deep learning mode such us LSTM shows higher accuracy on the covid dataset of around 140 observation about the user opinion in social media so far almost 1 million observation from individual tweets is been collected and extracted the safeguard unit in covid They additionally proposed system of natural language to break down with development of positive social media opinion. Proposed investigation observes well-known subjects about the online post from all the earlier work. Similarly detailed standard discoveries the online opinions utilizing NLP draws near. Our research main focus to propose the integration of machine and deep learning mode to perform the sentiment analysis upon tweet propagated in the social media. Our implementation shows significant list of capabilities fully intent on further developing precision. The recommended framework inspects the feelings of gathered tweets for opinion arrangement and concentrates the main capabilities, which helps with further developing grouping results when contrasted with the pattern method.

3.   PROPOSED METHODOLOGY

This segment explains the strategies used to examine the fundamental commitments to this review, which proposes the utilization of an unsupervised subject model, with a cooperative profound learning model in view of LSTN to examine COVID-19-related remarks from sub-reddits. The created system, shown in Fig. 1, involves feeling and semantic investigation for mining and assessment examination of COVID-19-related remark.

Figure 1: The flow of the proposed algorithm.
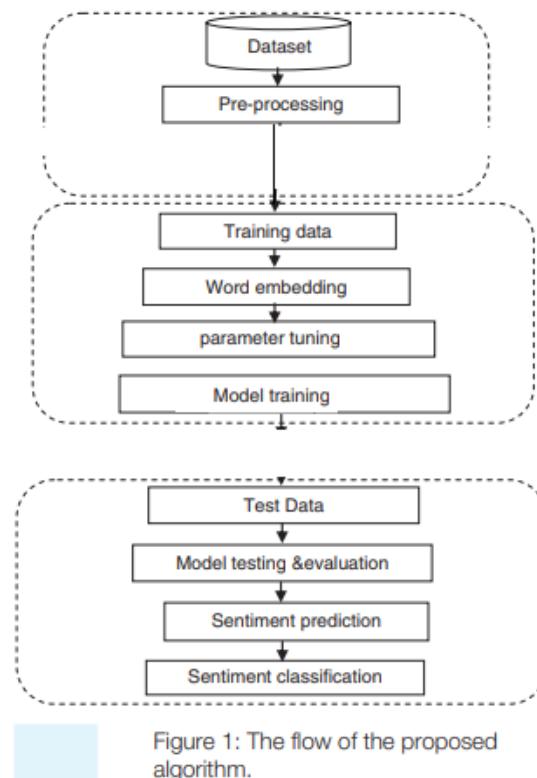
A.   *Preparing the data and Text Pre-Processing*

Information accumulated from social it is all the more frequently uproarious and heterogeneous to organize media stages. We make the Twitter stream prepared for exploratory examination; the pre-handling step first changes the capitalized letters to bring down case, then, at that point, we retrieved all the tweets from social media and extract the standalone hashtags for further processing eliminates every one of the extraordinary characters, URLs, stop words, notices into minuscule pieces as separated hashtags that decidedly influence the information groups. Camel Cases has is been utilized over couple hashtags with respect of hashtags like "#StayHome," those are not difficult to change over into the portion. In any case, then again, some hashtags that include no camel case, e.g., "#stayhome," a tremendous jargon is expected to extract the hashtag and retrieve the longest

string objects as a text pre-processing added to that it utilizes a bunch of jargon of just about 70,000 English words to deal with these difficulties. Recognizing enlightening substance from an enormous and loud dataset, for example, tweets is a difficult errand. To accomplish this, the accompanying strategies are done inside the provided request to upgrade the text.

1. One well known method for breaking down COVID Sentiment information must be figure out word frequencies to see the value in the way routinely words are utilized in online opinion. Along these lines, the initial step is lemmatization that cycles with the utilization of a jargon and morphological examination of expressions and returns root words. We utilized lemmatization with the nltk technique that changes an expression over to its base structure, for instance, "passings" to "death" or "mindful" to "care").

2. The subsequent advance is to eliminate the stop words. It is the most appropriate strategy to beat the commotion from the text based tweets, (for example, "the," "a," "an," "in"). Stop words can be separated from the message to be handled, and it really does never again influence comprehension opinion strength over online media valence. Overall, as a pre-processing we have removed all the unnecessary words Which is called as stop words in NLP from the collected tweets like "the", "off"," and "

3. To eliminate the complexity our model perform the conversion of the case of the message and also comprehends the certified word into extraordinary word such as covid consider as coronavirus.our mechanism never alter the semantics of the word or sentences.

4. Constriction with the aid of other techniques to remove all the special symbols like punctuation question mark to enhance the quality of phrases in the tweets. These days, individuals spend their time on various social media like Fb,Twitter,whats app to express their opinion about any context like covid ,politics. Many individuals speak with one another; individuals for the most part utilize abbreviated structures and contractions of words in their text. We utilized the withdrawal planning technique that drops the vowels from the words. Expulsion of compression planning is connected with message normalization, and it is useful while working with Twitter information in opinion examination.

5. Twitter information is loud, which influences classification model enable to pre-process all the URL and dispose @client_mentions.Adittinally we also elimate all the alphanumeric characters and ASCII from our data repository for better opinion analysis on the grounds that they don't assist us with distinguishing feeling. We likewise supplant emoticons with their comparing response in text.

6. For hashtags, we kill the "#" image from the beginning of the expression. We utilized tokenizer to part hashtags into proper words, for instance, "#stayhomestaysafe," tokenizer changed over it into "remain," "home," "remain," "safe." Many words are linked at the end of the day, and we likewise performed word-division to accomplish this.

B. Feature Extraction

In our proposal we perform various operation like vectorization on the words, word embedding and also we deals with term frequency -inverse term frequency (TF-IDF) methods are utilized for include extraction. The count vectorizer include extraction procedure is utilized to change the retrieved opinion from the social media into multidimensional vector and enables those tweets in view of the most usually utilized words (count) that impacted over all the received social media opinion in the tweets. Word grid is form with the aid of vectorization of words where each remarkable word addresses the section of the network, and the chose message from the archive addresses the column of the

framework. Along these lines, we include the word in that specific text test. We additionally utilized the term frequency-inverse document frequency (TF-IDF) include extraction procedure alongside vectorization and its count. In our proposal ,term frequency and inverse term frequency is utilized for interrogate the social media opinion with weighted based implementation support to achieve the vectorization task. Followed by the entire process term frequency on the document is been compared with inverse term frequency in order to extract individual items in the record and prepared the total term frequency with the following characterized equation(i)

$$\text{tf}\,(t,d) = \frac{Frequency\ of\ term\ t, in\ document\ d}{Total\ number\ of\ terms\ in\ document\ d} \quad \text{(i)}$$

Where as in the above equation (i), TF t referred as number in report d and addressed with countt,d while the general number of terms in that archive is addressed by total countd. Inverse term frequency imagines that the document expansion over the keyword element t ie term will be highly utilized parameter educational for model preparation. Our inverse term frequency analysis may be characterized as in condition 2.

$$\text{idf}\,(t) = log\,\frac{Total\ number\ of\ documents}{Number\ of\ documents\ with\ term\ t\ in\ it} \quad \text{(ii)}$$

Whereas I" denotes the total number of documents retrived for the processing and DF" represent the amount expression t term is incorporated in the document. Whenever a term t regularly shows up in many archives, IDF processes the loads of an expression t low. At the moments stop words will have low in inverse term frequency ration amount the documents with respect to that inverse term frequency can be framed as following equation (iii)

$$tf\ idf\,(t,\,d,\,D) = tf\,(t,\,d)\,.\,idf\,(t,\,D) \quad \text{(iii)}$$

B. Machine learning and Deep Learning Model

Feeling examination or assessment mining is a field of study that breaks down individuals' perspectives, feelings, assessments, decisions, mentalities, and feelings towards item substances, authoritative administrations, people, issues, occasions, subjects, and qualities [15]. Feeling investigation is frequently alluded to as subjectivity examination, assessment mining, and evaluation extraction with a few connections connected with full of feeling registering, to be specific PC acknowledgment and enthusiastic articulation [16]. Different AI techniques can be utilized to order tweets in view of opinion extremity. The procedures utilized in this study are as per the following:

Logistic Regression

In this segment, we'll utilize the removed elements to foresee the opinion of a tweet. Calculated relapse is valuable for this as it utilizes a sigmoid capacity to yield a likelihood somewhere in the range of nothing and one. Review that in managed AI we have input

highlights X and a bunch of marks Y.In request to make forecasts, we want a capacity with boundaries θ to plan elements to yield names ^Y.To improve the planning of highlights to names, we limit the expense work by contrasting how close the result ^Y is to the genuine marks Y from the information. After this, the boundaries are refreshed and the cycle is rehashed until the expense work is limited to a palatable level. With strategic relapse, the capacity Fis equivalent to the sigmoid capacity. Specifically, the capacity use make forecasts in calculated relapse his the sigmoid capacity that relies upon boundaries θ and the elements vector xi, where I signifies the ith perception:

$$h(x^{(i)}, \theta) = \frac{1}{1 + e^{-\theta^T x^{(i)}}}$$

Finally the logistic regression notation utilized with the training the model alongside of weight space of vector θ.

Random Forest

One of the most used algorithms is random forest which has bunch of decision tree to perform the classification of the opinions. It is very like the choice tree however contains a huge number of choices in the branches and provide various classification results on the final target variable to provide the class and the class name is the mode worth of the classes anticipated by individual choice trees. This calculation is productive in dealing with huge datasets and large number of info factors without their erasure. Random forest purpose's greater part vote and returns the class mark with greatest votes by the individual choice trees. Social media document is handled as per the topic heading and enable the user model to perform the classification model, that document has two major element one is part of speech and the heads on the document.

what's more, document head with text [18]. The choice tree technique can be connected with a standard based framework. A few principles appear in the choice tree calculation while the preparation information document with final class based on the document terms is been. Performance analysis based on accuracy, precision are utilized to extract the best outcome of the model for the given sentiment analysis problem.

LSTM

Various solutions will be given using the latest techniques like deep learning, artificial neural network is implemented to achieved the sentiment classification from the baog of words to formulate the public opinion over the media. For that, it has used LSTM (Long-Short Term Memory) techniques as a combination of deep learning and neural networks. To address the problems in web mining challenges such as geo location and time of generated data can be maintained in a separate memory using LSTM techniques. The following figure 2 denotes the architecture of the LSTM
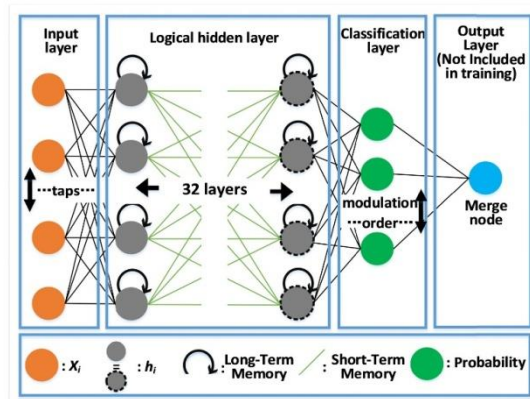
Structure of LSTM

Figure 2: LSTM architecture

The above figure explain about the LSTM structure which provides four layers for data processing namely input, logical hidden, classification and the output layerThe input layer accepts all the series of data generated from various sources on a large network and it will be accessed by a logical hidden layer. This layer using separate logic to categorize the input data and it will be sending for preprocessing to remove the unwanted things. There are a total of 32 layers are working for this operation and the trained data set will be given to the classification layer. Again the classical classification techniques are used to find out the time of generated data with its location by deep learning concepts then the output will be given as output to the next layer. Finally, these trained network data has to be given as an input to the RNN system for data accessing purpose repeatedly.

Next level the classified data is taken as a trained input for the second step through a recurrent neural network. Deep learning concepts are used here for classification and it is running in the hidden layer. The time of generated input data has stored as a variable named t and these values will be given as an input for the next level procedures. So different data time intervals can be calculated from the log files and their validity also monitored. Finally, the location of the data which was generated from various sources around the world on the larger network than predicted output will be generated.

4. RESULTS

|	UserName	ScreenName	Location	TweetAt	OriginalTweet	Sentiment
0	1	44953	NYC	02-03-2020	TRENDING: New Yorkers encounter empty supermar...	Extremely Negative
1	2	44954	Seattle, WA	02-03-2020	When I couldn't find hand sanitizer at Fred Me...	Positive
2	3	44955	NaN	02-03-2020	Find out how you can protect yourself and love...	Extremely Positive
3	4	44956	Chicagoland	02-03-2020	#Panic buying hits #NewYork City as anxious sh...	Negative
4	5	44957	Melbourne, Victoria	03-03-2020	#toiletpaper #dunnypaper #coronavirus #coronav...	Neutral

Figure 3 Covid-19 Dataset
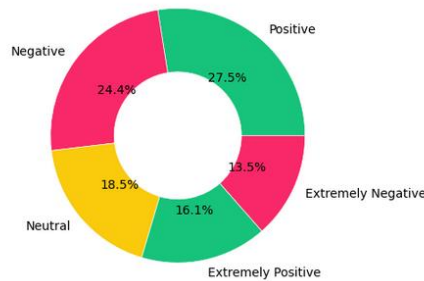
Figure 4 Proportion of tweets associated with each topic

```
Accuracy score: 0.6300684570826751
Macro-f1 score:0.6416159377897189
Micro-f1 score:0.6300684570826751
```
Figure 5 Accuracy of Logistic Regression model

5. CONCLUSION

Due to the pandemic situation though covid made everyone stay back in their home, working from home, and "isolation time," casual correspondence media has been broadly broadcast the news sentiments, sentiments, urging; regardless, an enormous part of the data by means of online amusement are unessential and don't have a spot with the certified circumstance. our research proposed a method for managing deal with the social media assessment using the COVID Sentiment opinion. We survey deep and machine learning with novel approach removing techniques that subsequently learn features in absence of human deterrent. We saw that people comply with government courses of action and Guideline Operating Procedures and began to lean toward shut down and implement social isolating in the year 2020 of march, but the solicitation by the public authority is in February 2020. There is a ton of trickiness by means of online diversion; in this manner, prosperity affiliations need to encourage a consistent system for perceiving Covid precisely to obstruct the spread of fake information. Our model gave the higher accuracy and precision among various machine learning techniques when diverged from similar

state of the art assessments. As further implementation we plan to take apart general feelings toward the importance or remarkable word fundamental points, like government reaction to the pandemic circumstance, medical care offices by government, disconnected assessment, and psychological well-being by utilizing DL calculations to expand their exhibition on the dataset. One constraint of this work is that it is explicit and doesn't check out at the disposition and feelings of individuals. Further work should be possible on the discovery of mind-set-based opinion examination.

6. REFERENCES

[1] Naseem U, Razzak I, Khushi M, Eklund PW, Kim J. Covidsenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis. *IEEE Trans Comput Soc Syst*. (2021) 8:1003–15. doi: 10.1109/TCSS.2021.3051189

[2] Iwendi C, Mahboob K, Khalid Z, Javed AR, Rizwan M, Ghosh U. Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system. *Multimedia Syst*. (2021) 1–15. doi: 10.1007/s00530-021-00774-w

[3] Bhattacharya S, Maddikunta PKR, Pham QV, Gadekallu TR, Chowdhary CL, Alazab M, et al. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: a survey. *Sustain Cities Soc*. (2021) 65:102589. doi: 10.1016/j.scs.2020.102589

[4] Mansoor M, Gurumurthy K, Prasad V. Global Sentiment Analysis Of COVID-19 Tweets Over Time. *arXiv preprint* arXiv:201014234. (2020).

[5] Mittal M, Siriaraya P, Lee C, Kawai Y, Yoshikawa T, Shimojo S. Accurate spatial mapping of social media data with physical locations. In: *2019 IEEE International Conference on Big Data (Big Data)*. Los Angeles, CA: IEEE (2019). p. 4113–6.

[6] Chauhan S, Banerjee R, Chakraborty C, Mittal M, Shiva A, Ravi V. A self-congruence and impulse buying effect on user's shopping behaviour over social networking sites: an empirical study. *Int J Pervasive Comput Commun*. (2021) doi: 10.1108/IJPCC-01-2021-0013

[7] Chauhan S, Mittal M, Woźniak M, Gupta S, Pérez de Prado R. A technology acceptance model-based analytics for online mobile games using machine learning techniques. *Symmetry*. (2021) 13:1545. doi: 10.3390/sym13081545

[8] Dhama K, Khan S, Tiwari R, Sircar S, Bhat S, Malik YS, et al. Coronavirus disease 2019-COVID-19. *Clin Microbiol Rev*. (2020) 33:e00028–20. doi: 10.1128/CMR.00028-20

[9] Ayoub A, Mahboob K, Javed AR, Rizwan M, Gadekallu TR, Abidi MH, et al. Classification and categorization of covid-19 outbreak in pakistan. *Comput Mater Continua*. (2021) 1253–69. doi: 10.32604/cmc.2021.015655

[10] Bhar A. Is it Possible to Ensure COVID19 Vaccine Supply by Using Plants?. Springer (2021).

[11] Aslam B, Javed AR, Chakraborty C, Nebhen J, Raqib S, Rizwan M. Blockchain and ANFIS empowered IoMT application for privacy preserved contact tracing in COVID-19 pandemic. *Pers Ubiquitous Comput*. (2021) 1–17. doi: 10.1007/s00779-021-01596-3

[12] Jelodar H, Wang Y, Orji R, Huang S. Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach. *IEEE J Biomed Health Inform*. (2020) 24:2733–42. doi: 10.1109/JBHI.2020.3001216

[13] Ahmed, S. T., Singh, D. K., Basha, S. M., Nasr, E. A., Kamrani, A. K., & Aboudaif, M. K. (2021). Neural Network Based Mental Depression Identification and Sentiments Classification Technique From Speech Signals: A COVID-19 Focused Pandemic Study. *Frontiers in public health*, 9.

[14] Ashokkumar P, Shankar GS, Gautham S, Reddy MPK, Reddy GT. A Two-Stage Text Feature Selection Algorithm for Improving Text Classification. *ACM Trans Asian Low-Resour Lang Inf Process*. New York, NY: Association for Computing Machinery (2021). p.19. doi: 10.1145/3425781

[15] Hakak S, Alazab M, Khan S, Gadekallu TR, Maddikunta PKR, Khan WZ. An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generat Comput Syst*. (2021) 117:47–58. doi: 10.1016/j.future.2020.11.022

[16] Lopez CE, Vasu M, Gallemore C. Understanding the perception of COVID-19 policies by mining a multilanguage Twitter dataset. *arXiv preprint* arXiv:200310359. (2020).

[17] Rustam F, Khalid M, Aslam W, Rupapara V, Mehmood A, Choi GS. A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis. *PLoS ONE*. (2021) 16:e0245909. doi: 10.1371/journal.pone.0245909

[18] Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. *J Med Internet Res*. (2020) 22:e19016. doi: 10.2196/19016