International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

Utilizing Advanced Artificial Intelligence for Early Detection of Epidemic Outbreaks through Global Data Analysis

*Kulbir Singh[1], Amit Bhanushali[2], Dr. Biswaranjan Senapati[3]

Submitted: 19/09/2023 **Revised**: 17/11/2023 **Accepted**: 29/11/2023

Abstract: Artificial Intelligence (AI) is increasingly becoming a pivotal tool in disease prediction, aiding in both medical diagnostics and outbreak containment. This study presents a novel approach to forecasting disease-prone areas using Text Analysis and Machine Learning, focusing on the power of social network data to anticipate epidemic outbreaks. We have developed an epidemic search model that utilizes a combination of data pre-processing techniques and diverse algorithms to analyse the likelihood and potential locations of outbreaks. Our methodology integrates Support Vector Machine (SVM), Naive Bayes, and Recurrent Neural Network-Long Short-Term Memory (RNN-LSTM) models, along with advanced text processing methods such as word-n-grams, word embedding, and Term Frequency-Inverse Document Frequency (TF-IDF). In our findings, the integration of Naive Bayes with TF-IDF emerged as the most effective technique, showcasing superior predictive capabilities. This research not only demonstrates the feasibility of using AI in epidemiological predictions but also underscores the potential of social network data in enhancing public health responses.

Keywords: *Artificial Intelligence, Epidemic Prediction, TF-IDF, SVM (Support Vector Machine), Naive Bayes, Social Network*

1. Introduction

Recent advancements in healthcare and medicine, coupled with the accumulation of extensive patient data, have catalysed the development of advanced medical diagnostic systems. The twentieth century witnessed significant contributions from researchers in creating systems that assist in interpreting diseases, selecting therapies, and generating hypotheses using AI. [1, 2] With the surge in information technology, public health monitoring has evolved, enabling the use of AI-based systems for comprehensive epidemic analysis and management. [6, 7, 8]

Despite these advancements, challenges persist in AI-driven solutions for managing infectious diseases, including issues related to data quality and availability. [9,10] Lessons from past epidemics like SARS have been instrumental in shaping responses to current public health threats like COVID-19. The use of diverse data streams, including credit card transactions, surveillance footage, and mobile phone records, has enhanced the tracking and management of infectious diseases. [14, 15, 16, 17]

The core of this study revolves around predictive modelling, a rapidly growing field in healthcare research. Utilizing analytics to discern patterns in data, especially in text analysis, has become crucial. Our focus extends to sentiment analysis using Twitter data, a challenging yet insightful domain that transcends social media to impact sociology, healthcare, and more. [18,19,20,21] Sentiment analysis, intersecting with natural language processing and computational linguistics, plays a pivotal role in understanding public sentiment and emotional states related to health crises. [22, 23, 24]

This research aims to harness the power of AI, particularly in predictive modelling and sentiment analysis, to anticipate and manage epidemic outbreaks. By analysing global data streams and public sentiment, we seek to contribute to a more proactive and informed approach in public health responses.

2. Review of Literature

Madurai Elavarasan and Pugazhendhi (2020) [25] in their 2020 study, Elavarasan and Pugazhendhi comprehensively explore the diverse roles played by technology in the containment of epidemics. They shed light on a variety of technological interventions and how they bolster healthcare systems, aid governmental efforts, and facilitate public response during epidemic emergencies. This research provides valuable insights into the integration of technology in managing public health crises.

R M et al. (2020) [26] The study by R M et al. (2020) employs machine learning techniques for sentiment analysis, particularly focusing on Twitter data. Their work aims to visualize and analyse the global repercussions of an epidemic, capturing a spectrum of public emotions and perceptions, both positive and negative.

[1]*Health Information Manager, IL, USA,*
ORCID ID : 0009-0007-0297-6647
[2]*Quality Assurance Manager, West Virginia University,WV, USA .*
ORCID ID : 0009-0005-3358-1299
[3]*Doctor in Computer and Data Science,Parker Hannifin Corp,USA*
ORCID ID : 0000-0002-0717-5888
**Corresponding Author Email: kulbir.klmna@gmail.com*

Shamman et al. (2021) [27] In their comprehensive 2021 research, Shamman et al. explore the deployment of innovative technologies in vital fields such as medical image processing, disease surveillance, and predictive modelling of outcomes in the context of pandemics. This study illuminates the pivotal role that technological advancements, bridging healthcare and computer science, play in the effective management of widespread health emergencies.

In each of these insightful studies, we gain a more nuanced understanding of the multifaceted ways in which AI and technology are being utilized to combat epidemics. Ranging from analysing public sentiment and advancing medical imaging techniques to their broader implications on enhancing healthcare systems and shaping government responses, these research efforts collectively paint a detailed picture of the critical role of technology in epidemic management.

3. Objectives

Our research focuses on developing sophisticated AI-based models that are adept at predicting epidemic outbreaks through the analysis of global data streams. This endeavour entails the application of robust machine learning and data mining techniques to effectively process and interpret extensive datasets derived primarily from social networks. The aim is to harness the predictive power of AI to analyse patterns and trends in social media data, which can offer early indicators of potential health crises.

4. Statement of the Problem

The introduction of predictive AI algorithms into public health research, especially in forecasting disease outbreaks through the analysis of global data streams, signifies a transformative moment in the discipline. However, this breakthrough comes with notable challenges. The primary issue lies in the dynamic and multifaceted nature of these data sources, which include social media feeds, electronic health records, and environmental monitoring systems. The heterogeneity and varying quality of data from these sources are significant obstacles in the path of developing reliable and precise predictive models. Moreover, the necessity to process this broad spectrum of data in real-time poses a formidable technical challenge.

In addition, employing personal and sensitive information for epidemic prediction raises serious ethical and privacy concerns. Balancing the effective use of this data for public health benefits with the imperative of safeguarding individual privacy and data security is a critical and complex task. These issues highlight the necessity for a sophisticated approach in applying AI to epidemic prediction—an approach that not only capitalizes on technological progress but also conscientiously navigates the complexities and ethical implications inherent in this field.

5. Research Methodology

Epidemic Search model that harnesses the information available from social network analysis to predict the likelihood of an outbreak and examine the regions of the world most likely to be affected by such an event. Twitter's publicly accessible social network data, or tweet-data, is a rich source of information on current events throughout the world.

a) DataSource

In our study, we capitalized on the rapid growth and widespread use of social media platforms, particularly Twitter. These platforms have become a goldmine of real-time, user-generated content, which presents a unique opportunity for monitoring and predicting epidemic trends.

To acquire the relevant data, we employed the Twitter API, which necessitated registration and authentication using our Twitter developer account. This approach enabled us to specifically target and retrieve tweets that were pertinent to our research objectives.

For the data retrieval process, we used Python in conjunction with the Tweedy library. This combination proved to be highly efficient for data collection, allowing us to seamlessly gather and store large volumes of tweet data in JSON format, which facilitated subsequent data processing and analysis.

One of the primary challenges we faced was managing the sheer volume of tweet data. Given its vastness and the continuous influx of new data, it was imperative to have a robust and scalable storage solution.

To address this, we utilized the Hadoop Distributed File System (HDFS) and the Apache Cassandra database management system. This distributed storage approach not only accommodated the large-scale nature of our data but also ensured efficient retrieval and management.

For processing and analyzing the stored data, we integrated the Spark cluster-computing framework. The use of Python's pyspark library allowed for smooth interaction with our data within Spark. Furthermore, we employed MongoDB, a versatile and robust document-oriented database, to organize and manage our data effectively. This setup was crucial in handling the complexities of our dataset and facilitated various analytical operations required for our study.

b) Pre-processing

In the preprocessing of Twitter data for the study, a comprehensive approach is employed to ensure the dataset's cleanliness and standardization for effective modeling. This involves using Python's nltk package to remove common stop words, thereby filtering out text that offers little analytical value. Regular expressions are applied to clean the tweets by removing URLs, mentions, slang, and non-standard language, ensuring relevance and uniformity. Emoji's and emoticons are either converted or removed to maintain textual consistency, especially for sentiment analysis. Text standardization includes converting all text to lowercase to prevent the misinterpretation of identical words in different cases and breaking down tweets into individual words or tokens, which is a fundamental step for natural language processing.

Additionally, non-alphanumeric characters such as punctuation are removed to focus on meaningful text content, while common abbreviations and slang are translated into their standard forms for clarity. Words are reduced to their base or dictionary form through stemming or lemmatization, standardizing similar words. The process also involves identifying and removing duplicate tweets to avoid skewing the data, along with the removal of irrelevant data (noise) to maintain the focus on pertinent information. Finally, handling missing data through imputation or removal ensures the integrity of the dataset. Each step in this preprocessing phase is crucial to prepare the Twitter data for the predictive models, ensuring clean, relevant, and structured data for optimal performance and insightful analysis outcomes.

c) Polarity Generation

The tweet's polarity, which is gleaned from the tweets' sentiments via the python text blob, serves as the dependent variable or predictor. The text is analyzed for sentiment, classified, and translated into several languages. It also aims to extract noun-phrase phrases from text. The sentiment analysis of tweets is the focus of the following simple function.

d) HashTagAnalysis

The # symbol designates all words that begin with that symbol as hash tags. These are useful for gaining insight into current topics of discussion. In a word cloud, the size of each word is related to how often it appears in the text. That's why the most popular things to tweet about tend to include loftier concepts. The majority of tweets address societal issues such as sickness, hunger, and famine, as well as the nations that are worst hit. "In our instance, most common hash tagged terms are Yemen, Cholera, Cholera Nairobi, zika, vaccinations, The Story of Yemen etc. The hash tag analysis function is shown below."

e) Top-Users/Citations

Analyzing @ symbols allows us to extract usernames, which are used similarly to hashtags. The Citations Word Cloud.

6. Results and Discussion

Understanding the distribution and total number of instances of each class in the dataset is crucial when the unstructured data is transformed into a supervised learning process. Now, let's take a look at the tweets' categorizations to see how they're spread out:

Positive Tweets: By filtering for tweets with a favourable sentiment and then extracting keywords, we can determine how often certain phrases are used. Health, water, vaccines, and sanitation all appear in the word cloud of upbeat tweets. According to the text blob, about 26.67 percent of all tweets fall into the "positive class," which consists of 771 examples. The graphic below depicts the class's word cloud.

Fig 1: Cloud of Words Expressing a Positive Attitude

Negative tweets: Similarly, a word cloud of unfavourable tweets reveals that the most often used words include "outbreak," "dengue," "worst," and "Yemen," among other words. In this particular instance, the overall number of instances belonging to the negative class comes to 692, making up 23.96% of the total cases. The word cloud representing the unfavourable classes appears as seen in the image below.

Fig 2: Cloud of Words Expressing a negative Attitude

Neutral tweets: The word cloud for neutral tweets reveals that the most often used terms include things like hotels, cholera, and Weston spread, among other things. “The overall number of instances that may be considered neutral comes to 1425, which is 49.34% of the total cases.” This group's word cloud looks like the one shown in the graphic below.

Fig 3: Word Cloud Regarding the Attitude of Neutrality

- **MachineLearning**

Since we already have the labels, after we have cleaned the data and have a general sense of what it contains, we will be able to utilize any supervised machine learning model for sentiment classification. N-gram characteristics are used by the majority of the algorithms that include text categorization. This falls under the "Bag of Words" concept since the model does not place any importance on the specific sequence in which the words are presented. In more recent sophisticated models that use RNN/LSTM models, word order is taken into account during the classification process.

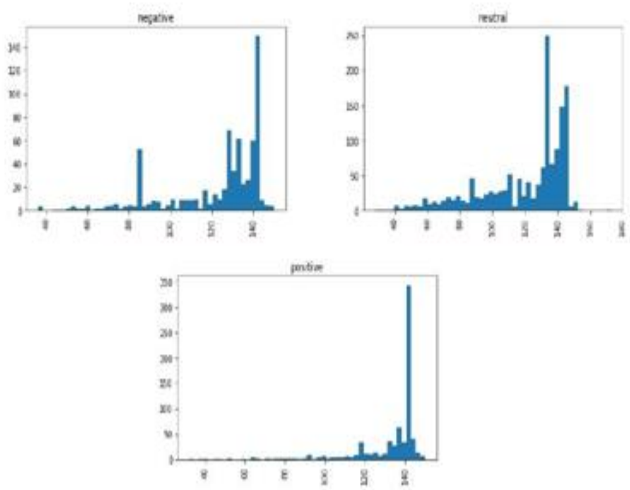

Fig 4: Graph showing the individual classes' sentiment scores as a histogram

- **SentimentClassifier:**

Now that we have a twitter dataset, we’ve used a number of different models and methods to classify the tweets and evaluate the performance of different classifiers. The dataset we just produced will serve this purpose and you can see how they compare against one another in the table below: Based on the results of the prior studies, the Decision Tree Classifier was shown to be the most effective method, followed by the K Nearest Neighbor Classifier.

Table 1: Analysing the relative efficacy of several algorithms

Model	**Accuracy**
Logistic regression	0.5
Random Forest Classifier	0.8
Ada Boost Classifier	0.8
Bernoulli NB	0.9
Multinomial NB	0.9
K neighbours Classifier	0.8
Decision Tree Classifier	0.9
Extra Tree Classifier	0.9
Gradient Boosting Classifier	0.5

- **Wordsn-Grams:**

As with sentences and documents, a tweet is represented as the bag (multi set) of its words, with syntax and word order ignored but the tweet's multiplicity maintained.

Advantages: The feature representation is conventional and quite reliable.

Weaknesses: Forget about correct language and sentence structure.

Hyper-values: The program's algorithm definition

See below for an n-grams-based Feature Evaluation of several methods:

- **Unigram and bigram Naive Bayes**

In this case, we used a pair of Naive Bayes algorithms. Bernoulli Naive Bayes and Multinomial Naive Bayes are two examples. “In contrast to the frequency-based Multinomial Naive Bayes model, the Bernoulli Naive Bayes model follows the Bernoulli distribution. We have implemented the 5-fold cross-validation method, whereby 4 of the 5 folds (or samples) are utilized for model training, while the last fold is used for model validation. The results of the 5-fold cross validation technique show that the multinomial NB has an average accuracy of 78.11%, while the Bernoulli NB has an average accuracy of 78.42%, the Bernoulli NB has a minimum accuracy of 61.53%, and the multinomial NB has a maximum accuracy of 83.36%." Bernoulli Naive Bayes outperformed Multinomial Naive Bayes in terms of average and lowest accuracy. However, we discover that Multinomial Naive Bayes has the highest accuracy. Accuracy of multinomial NB for hold-out Validation is 84.43 percent, and the confusion matrix looks like this:

Table 2: Multinomial NB Confusion Matrix

Label	**Actual Negative**	**Actual Neutral**	**Actual Positive**
Predicted Negative	175	32	22
Neutral	20	373	32
Positive	6	23	184
Support	201	428	238	867

It can be seen from this analysis that 175 of the 213 samples from the negative class were accurately predicted, whereas 20 were incorrectly assigned to the neutral class, and 6 were assigned to the positive class. In a sample of 423 tweets categorized as neutral, only 32 were allocated to the negative class and 23 were projected to the positive class, demonstrating a bias towards negative tweets in our neutral labelling. There were a total of 54 incorrect classifications for tweets with positive sentiments (22 were assigned to the negative category, 32 to the neutral category), whereas 184 were assigned properly. This also demonstrates that the likelihood of a positive tweet being forecasted as negative is far lower than the opposite. Even though it reported an accuracy of 0.84, the misclassification rate was higher than indicated by the confusion measure. Precision, recall, and f1-score values all reflect the same pattern of analysis:

Table 3: Analysis

Class-name	**Negative**	**Neutral**	**Positive**	**Avg. / Total**
Precision	0.76	0.88	0.86	0.85
Recall	0.87	0.87	0.77	0.84
F1-Score	0.81	0.87	0.82	0.84

Unigram and bigram use in linear SVM

When employing the cross-validation approach with a 5-fold split, our accuracy increases to 83.50% from 71.06%. The average precision is 80.09 percent. As a result, we will be basing our final verdict on the mean accuracy for more accurate comparison and portrayal. Measure of model confusion is:

Table 4: The SVM Error Matrix

Label	**Actual Negative**	**Actual Neutral**	**Actual Positive**
Predicted Negative	145	1	0
Neutral	68	421	73
Positive	0	1	158

Out of 213 negative samples, 145 were properly predicted, whereas all incorrect predictions were neutral. This demonstrates that tweets with positive and negative attitudes can be differentiated from one other with considerably more ease than tweets with negative and neutral sentiments. Only 2 of 423 anticipated neutral tweets were incorrect, leaving 421 right predictions. This suggests that the model did better when dealing with neutral tweets. In addition, it misclassified 73 out of 231 tweets with a favorable mood. It follows the same trend as the tweets people are complaining about.

The F1 score of neutral tweets is greater than that of negative and positive tweets, lending credence to the aforementioned argument.

Table 5: Analysis

Class-name	**Negative**	**Neutral**	**Positive**	**Avg. / Total**
Precision	0.99	0.75	0.99	0.87
Recall	0.68	1	0.68	0.84
F1-Score	0.81	0.85	0.81	0.83

Support	213	423	231	867

In light of the above comparison:

Table 5: Evaluations of Various Algorithms with multiple Iterations

Multinomial NB	**Bernoulli NB**	**SVC**
78.11%	78.42%	80.09%
61.53%	65.16%	71.06%
83.36%	82.70%	84.40%

In this evaluation, SVC with a "Linear" kernel fared very well.

- **TF-IDF Victimizer**

To generate the features that will be fed into the various classifiers, we have made use of scikit-learn's TFidf Transformer. Here are the classifiers we've employed and how well they've performed:

NaiveBayes: 94.145%

SVC: 49.34%

SVM (TFIDF):87.9%

Naive Bayes (TFIDF):83.21%

After adding n-gram analysis to our model, we produced the following:

Unigram classifier (with and without mark-negation)
A bigram classifier (with and without mark-negation)
Unigram/Bigram classifier (with and without mark-negation)

Following is the outcome analysis:

UnigramClassifier: 88.75%, 89.44%

BigramClassifier: 88.58%, 88.58%

UnigramandBigramClassifier: 89.10%, 88.75%

- ***LSTM Networks***

LSTM is a recurrent neural network that uses past time steps. Corporate gates in LSTM solve the Basic RNN cell's long sequence learning problem. We tokenized input tweets after minimal cleaning for this model. After tokenization, phrase is mapped to vocabulary word index and padding index is preserved as 0, ensuring input record length is the same.

Few things have been explored for sentence-model gearbox.

- The vocabulary comprises 5000 most often used terms, preventing the acquisition of less common ones.
- Sentences are set to 500 words, with shorter sentences padded and larger ones trimmed.

Self-embedding is used to learn word embedding, and the word size is 32. The model has 5 layers.

Layer (type)	Output Shape	Param #
embedding_1 (Embedding)	(None, 500, 32)	160000
dropout_1 (Dropout)	(None, 500, 32)	0
lstm_1 (LSTM)	(None, 100)	53200
dropout_2 (Dropout)	(None, 100)	0
dense_1 (Dense)	(None, 1)	101

Fig 5: Architecture of LSTM

Below is LSTM model training: The validation accuracy of LSTM is 67.04%.

```
Train on 1732 samples, validate on 1156 samples
Epoch 1/9
1732/1732 [==============================] - 158s - loss: -0.3619 -
acc: 0.5456 - val_loss: -0.8101 - val_acc: 0.6159
Epoch 2/9
1732/1732 [==============================] - 146s - loss: -2.0605 -
acc: 0.6443 - val_loss: -2.3857 - val_acc: 0.6237
Epoch 3/9
1732/1732 [==============================] - 152s - loss: -3.0746 -
acc: 0.6796 - val_loss: -2.7091 - val_acc: 0.6427
Epoch 4/9
1732/1732 [==============================] - 167s - loss: -3.3368 -
acc: 0.7021 - val_loss: -2.8678 - val_acc: 0.6514
Epoch 5/9
1732/1732 [==============================] - 152s - loss: -3.4819 -
acc: 0.7252 - val_loss: -2.9131 - val_acc: 0.6557
Epoch 6/9
1732/1732 [==============================] - 151s - loss: -3.5609 -
acc: 0.7396 - val_loss: -2.9819 - val_acc: 0.6644
Epoch 8/9
1732/1732 [==============================] - 160s - loss: -3.6095 -
acc: 0.7436 - val_loss: -2.8963 - val_acc: 0.6678
Epoch 9/9
1732/1732 [==============================] - 141s - loss: -3.6211 -
acc: 0.7494 - val_loss: -3.0327 - val_acc: 0.6704

Out[ ]:
<keras.callbacks.History at 0x2102682afd0>
```

Fig 6: LSTM training outcomes

A revolutionary step forward in public health would be the introduction of predictive AI algorithms for predicting disease breakouts using global data streams. Improving the models' forecasting powers requires incorporating data from a wide variety of sources, such as social media, health reporting, and environmental monitoring. [28] It becomes clear that real-time data processing is a critical factor, and further study is needed to determine how to maximise speed and efficiency in this context. Ethical constraints and privacy concerns add a degree of complexity, demanding research of responsible data usage frameworks. [29] The precision and dependability of prediction models may also be increased by the constant refining of algorithmic methodologies. To increase the worldwide capability for early epidemic detection and response, the debate focuses on optimising the synergy between data streams, boosting real-time processing, resolving ethical problems, and refining algorithms. [30]

7. Conclusion

The study of unstructured data like tweets might anticipate epidemic-hit areas, which could greatly affect global public health. Count Vectorization, TF-IDF, and Topic Modelling feature extraction methods are used to shape unstructured data. Accuracy is vital for determining prediction dependability, and the confusion matrix helps select the best method.

Comparative investigation shows that Naive Bayes with TF-IDF predicts epidemic-hit areas better than other methods. This emphasises the relevance of sentiment analysis, especially tweet polarity, and the usefulness of sophisticated machine learning algorithms for such jobs. The results show that integrating these strategies might improve early diagnosis and response, saving lives globally. As we enhance and extend these prediction models, data science and public health might mitigate global epidemics.

7.1 Findings of the study

Epidemic strike region forecast may save lives worldwide. This involves pre-processing unstructured data to make it organized and feeding it to machine learning algorithms using features extraction methods like Count Vectorization, TF-IDF, Topic Modelling, etc. To determine polarity, tweet emotion is very significant. We utilized the confusion matrix to choose the best method based on prediction "accuracy". The results of Naive Bayes employing TF-IDF were superior to other methods.

7.2 Scope for further research

The field of predictive AI models for disease outbreaks utilising global data streams has great potential. Textual, visual, and aural data might be used to improve prediction models in future research. Additionally, research on real-time data processing and analysis might speed up epidemic detection and response. Algorithmic advancements and new data sources may help us construct more robust and timely prediction models, enabling us to proactively address global health issues.

References

[1] Thomas,DavidR."Ageneralinductiveapproachforanalyzingqualitativeevaluationdata."American journal ofevaluation27.2237-246,2016.

[2] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." Foundations and Trends R inInformationRetrieval2.1–2:1-135, 2015.

[3] Adhikari,NimaiChandDas."PREVENTIONOFHEARTPROBLEMUSINGARTIFICIALINTELLIGENCE." 2021.

[4] Waaijenborg, Sandra, et al. "Waning of maternal antibodies against measles, mumps, rubella, andvaricella in communities with contrasting vaccination coverage." The Journal of infectious diseases208.1:10-16, 2018.

[5] Miner, Gary, John Elder IV, and Thomas Hill. Practical text mining and statistical analysis for non-structuredtext dataapplications.AcademicPress,2019.

[6] Barbosa, Luciano, and Junlan Feng. "Robust sentiment detection on twitter from biased and noisydata."Proceedingsofthe23rdinternationalconferenceoncomputationallinguistics:posters.Association for Computational Linguistics,2020.

[7] Han, Eui-Hong Sam, George Karypis, and Vipin Kumar. "Text categorization using weight adjustedk-nearest neighbor classification." Pacific-asia conference on knowledge discovery and data mining.Springer,Berlin,Heidelberg,2021.

[8] Pereira, Fernando C., Yoram Singer, and Naftali Tishby. "Beyond word n-grams." Natural LanguageProcessingUsingVeryLargeCorpora.Springer,Dordrecht,2017.121-136.

[9] Niesler, Thomas R., and Philip C. Woodland. "A variable-length category-based n-gram languagemodel."Acoustics,Speech,andSignalProcessing,2018.

[10] Adhikari, Nimai Chand Das, Arpana Alka, and Raju K. George. "TFFN: Two Hidden Layer FeedForwardNetworkusingtherandomnessof Extreme Learning Machine."2022.

[11] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification usingmachine learning techniques." Proceedings ofthe ACL-02 conferenceon Empirical methods innaturallanguageprocessing-Volume10.AssociationforComputationalLinguistics,2022.

[12] Dave,Kushal, Steve Lawrence, and David M. Pennock. "Mining the peanut gallery: Opinion extraction andsemantic classification of product reviews." Proceedings of the 12th international conference onWorldWideWeb.ACM,2021.

[13] Joachims,Thorsten.AProbabilisticAnalysisoftheRocchioAlgorithmwithTFIDFforTextCategorization. No. CMU-CS-96-118. Carnegie-mellonunivpittsburgh pa dept of computer science,2018.

[14] Tang, Duyu, Bing Qin, and Ting Liu. "Document

modeling with gated recurrent neural network forsentimentclassification."Proceedingsofthe2015conferenceonempiricalmethodsinnaturallanguageprocessing.2019.

[15] Adhikari,NimaiChandDas,ArpanaAlka,andRajatGarg."HPPS:HEARTPROBLEMPREDICTIONSYSTEMUSINGMACHINELEARNING."2021.

[16] Metsky HC, Freije CA, Kosoko-Thoroddsen T-SF, Sabeti PC, Myhrvold C. CRISPR-based COVID-19 surveillance using a genomically-comprehensive machine learning approach. *bioRxiv*. 2020.

[17] Husnayain A, Fuad A, Su EC-Y. Applications of Google Search Trends for risk communication in infectious disease management: a case study of the COVID-19 outbreak in Taiwan. *Int J Infect Dis*. 2019.

[18] Kolozsvari LR, Berczes T, Hajdu A, et al. Predicting the epidemic curve of the coronavirus (SARS-CoV-2) disease (COVID-19) using artificial intelligence. *medRxiv*. 2020.

[19] Yan L, Zhang H-T, Goncalves J, et al. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. *medRxiv*. 2018.

[20] Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J Trav Med*. 2022.

[21] Rada G, Verdugo-Paiva F, Ávila C, et al. Evidence synthesis relevant to COVID-19: a protocol for multiple systematic reviews and overviews of systematic reviews. *Medwave*. 2021

[22] Gilbert M, Pullano G, Pinotti F, et al. Preparedness and vulnerability of African countries against importations of COVID-19: a modelling study. *The Lancet*. 2020;395(10227):871–7.

[23] Sadilek A, Kautz H, Silenzio V. Predicting disease transmission from geo-tagged micro-blog data. In: *AAAI'12: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*. (12th), AI Access Foundation; 2017:136–142.

[24] Obeid JS, Davis M, Turner M, Meystre SM, Heider PM, Lenert LA. An AI approach to COVID-19 infection risk assessment in virtual visits: a case report. *J Am Med Inf Assoc*. 2022;27(8):1321–1325. 10.1093/jamia/ocaa105.

[25] Madurai Elavarasan R., Pugazhendhi R. Restructured society and environment: A review on potential technological strategies to control the COVID-19 pandemic Sci. Total Environ. 2020.

[26] R M., A B., K S. COVID-19 outbreak: Tweet based analysis and visualization towards the influence of coronavirus in the world GEDRAG Organ. Rev., 33 2020.

[27] Shamman A.H., Hadi A.A., Ramul A.R., Zahra M.M. A., Gheni H.M. The artificial intelligence (AI) role for tackling against epidemic Mater. Today: Proc. 2021.

[28] Bogoch II, Watts A, Thomas-Bachli A, Huber C, Kraemer MU, Khan K. Pneumonia of unknown aetiology in Wuhan, China: potential for international spread via commercial air travel. *J Trav Med*. 2021;27(2):taaa008.

[29] Xu X, Jiang X, Ma C, Du P, Li X, Lv S, Yu L, Ni Q, Chen Y, Su J, Lang G. A. *A deep learning system to screen novel coronavirus disease 2019 pneumonia. Engineering*. 2017;6(10):1122–1129.

[30] Chung M, Bernheim A, Mei X, et al. CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology*. 2018;295(1):202-207.

[31] M. Dagar, A. Kajal and P. Bhatia, "Twitter Sentiment Analysis using Supervised Machine Learning Techniques," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021, pp. 1-7, doi: 10.1109/ISCON52037.2021.9702333.