

Analyzing COVID-2019 Impact on Mental Health Through Social Media Forum

Huma¹, Muhammad Khalid Sohail², Nadeem Akhtar³, Dost Muhammad³, Humaira Afzal⁴, Muhammad Rafiq Mufti⁵, Shahid Hussain^{6,*} and Mansoor Ahmed¹

¹Department of Computer Science, COMSATS University, Islamabad, 45550, Pakistan

²Department of Management Sciences, Bahria Business School, Bahria University, Islamabad, 44000, Pakistan

³Department of Computer Science & IT, The Islamia University of Bahawalpur, Bahawalpur. 63100, Pakistan

⁴Department of Computer Science, Bahauddin Zakariya University, Multan, 60800, Pakistan

⁵Department of Computer Science, COMSATS University Islamabad, Vehari Campus, 61100, Pakistan

⁶Department of Computer and Information Science, University of Oregon, Eugene, 97401, Oregon, USA

*Corresponding Author: Shahid Hussain. Email: shussain@uoregon.edu

Received: 18 September 2020; Accepted: 05 December 2020

Abstract: This study aims to identify the potential association of mental health and social media forum during the outbreak of COVID-19 pandemic. COVID-19 brings a lot of challenges to government globally. Among the different strategies the most extensively adopted ones were lockdown, social distancing, and isolation among others. Most people with no mental illness history have been found with high risk of distress and psychological discomfort due to anxiety of being infected with the virus. Panic among people due to COVID-19 spread faster than the disease itself. The misinformation and excessive usage of social media in this pandemic era have adversely affected mental health across the world. Due to limited historical data, psychiatrists are finding it difficult to cure the mental illness of people resulting from the pandemic repercussion, fueled by social media forum. In this study the methodology used for data extraction is by considering the implications of social network platforms (such as Reddit) and leveraging the capabilities of a semi-supervised co-training technique-based use of Naïve Bayes (NB), Random Forest (RF), and Support Vector Machine (SVM) classifiers. The experimental results shows the efficacy of the proposed methodology to identify the mental illness level (such as anxiety, bipolar disorder, depression, PTSD, schizophrenia, and OCD) of those who are in anxious of being infected with this virus. We observed 1 to 5% improvement in the classification decision through the proposed method as compared to state-of-the-art classifiers.

Keywords: SARS-CoV-2; mental health; social media; Reddit; machine learning



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) was firstly reported in 2003 during epidemic in China and soon it spread in 26 countries. SARS-CoV was supposed to be an animal virus probably spread from bats to other animals. Subsequently, it infected those human beings who were living in the Guangdong province of China [1]. As soon as the virus spread in Wuhan, its news spread around the world due to the lethal nature of the virus. The virus on one side killed the people and on the other side it affected the financial state of the city, distressing the economy of the country. ICTV (International Committee on Taxonomy of Viruses) disseminated the name of the new virus “Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2).” Subsequently, the WHO (World Health Organization) entitled it as “Coronavirus disease 2019 (COVID-19).” The name was chosen based on the fact that the virus is genetically similar to the coronavirus, which was responsible for the spread of SARS in 2003 [2]. The progress of COVID-19 is shown in Fig. 1.

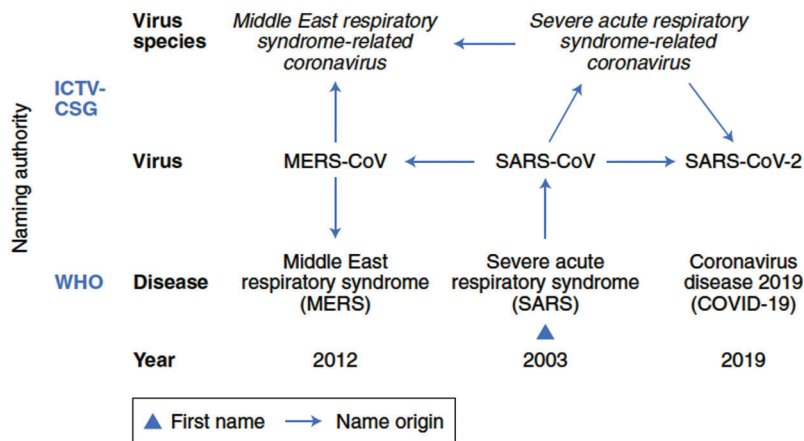


Figure 1: Growth of COVID-19 (Source ICTV)

Within a few months, the disease outbreak spread into 213 countries and territories like a storm. On 12th of March 2020, the World Health Organization (WHO) declared COVID-19 as an epidemic [3]. Scientists/pharmacists have either failed or are trying to develop a vaccine to cure the virus victims. After a few hits and trials, the most effective precaution adopted worldwide was lockdown as a safety measure. China also imposed a strict lockdown in Wuhan for 76 days in response to this pandemic. However, the lockdown strategy could not break the chain and affected billions of people around the globe. Therefore, to avoid the spread of the virus, the government of each country banned international travel to impose a strict lockdown. However, the duration of the lockdown varies from country to country and ranges from a few weeks to months. There are a few countries which are still under the lockdown. The anxiety and fear of infected people has spread and become a vital source to compel people to keep social distancing and follow lockdown procedures. However, no one anticipated that the lockdown and social distancing would last for such a long time, hence leading to several challenges for humans such as financial loss, isolation, and concern for the health of loved ones.

Severe depression, anxiety, and insomnia had been reported in the health care workers of China, during the early days of the pandemic. A bit of anxiety led to the implementation

of SOPs for prevention of COVID-19 spread like washing hands oftenly and working on the hygiene. However redundant anxiety leads to severe health issues including weak immune system and brain damage. Consequently, the COVID-19 has not only affected the physical health or economy but have adversely affected the mental health of the human beings. Psychologist believed that the pandemic not only hit the world but also affect the mental health and routine of people. Moreover, it is believed by few researchers and doctors that COVID-19 vaccination would be only used to control the transmission and not the disease itself. As it is still too early to say that how long the virus will persist, therefore it is permissible to say that only social distancing is effective to interrupt the spread of virus. The only mode to interact with others is social media. So, people are now spending more time on the social media for their communication which could be propitious for healthcare stakeholders to identify the impact of prevailing situation on the mental illness. However, due to high volume of data, it is difficult for a healthcare stakeholder to extract meaningful information manually about the mental illness of a patient through their communication on social media platforms. Machine Learning (ML) has played better role in several domains such as text classification, software development, opinion mining, healthcare system, and sentiment analysis. Consequently, the implications of ML could be realized to identify the mental illness of COVID-19 affected people through their communication on social media platforms such as Reddit. Such information could be more beneficial for healthcare stakeholders to manage the effective treatment of COVID-19 affected humans. In this study, we target the issue and propose a methodology by leveraging the capabilities of semi-supervised co-training learning method in which ensemble algorithm (such as NB, RF, and SVM) is trained with a large number of unlabeled data and a small number of labeled data due to lack of domain experts in terms of assigning labels. Firstly, we target the Reddit social media platform for extracting (using PRAW, a Python API for Reddit) the post and comments which are published in the context of mental health issues such as; anxiety, bipolar disorder, depression, PTSD, schizophrenia, and Reddit's OCD subreddits, and COVID-19 pandemic. Secondly, we applied a set of preprocessing activities to construct a Vector Space Model (VSM) for the training purposes of classifiers. Third, a label is assigned to some instances through the expert opinion to start the classifier's initial learning and getting label of each new instance which later become the part of the classifier's training.

The rest of the paper is organized into seven sections. Section 2 represents the overview of the work which is performed in the context of social media approaches for mental issues during COVID-19. Section 3 represents the brief overview of a semi-supervised co-training technique. Section 4 represents the proposed methodology. Section 5 represent the results. Section 6 presents the threats to validity. Finally, Section 7 describes the conclusion.

2 Related Work

In their study, Manguri et al. [4] performed sentiment analysis from social media application twitter by utilizing the TextBlob library of python. They extracted tweets by utilizing python library tweepy, with the hashtag of #coronavirus and #COVID-19. The time span was considered from one high peak week of the COVID-19 pandemic. The results show the people's reactions towards coronavirus vary day by day. Similarly, Chen et al. [5] revealed the difference in the utilization of neutral and controversial terms (such as "Chinese virus" and "COVID-19") on twitter for coronavirus by doing sentiment analysis and topic modeling. The author concluded that the "Chinese virus" term is frequently used either as a neutral or a racist.

In their study, Kleinberg et al. [6] performed topic modeling by accumulating Real World Worry data from participants during COVID-19. By using the predictive modeling approach,

the authors successfully predicted the emotional response of participants within 14% of their actual value.

In study [7], the author proposed the automatic detection of COVID-19 positive reports from social media by using machine learning and linguistic tools. The goal of the proposed work was to better understand the outbreak in China. The preliminary result of this work was to detect the positive reports by using machine learning state-of-the-art models. Similarly, in study [8], the author compared the higher education student's reaction and sentiment analysis towards COVID-19 by the general public. The aim of study was to discover the social issues of the general public and students during COVID-19 with the help of social media data. The author claimed the novel attempt in the proposed study that signified the difference between student's sentiments compared to general people's sentiment towards any major crises.

In their study, Li et al. [9] gathered situational informational data from Weibo and used natural language processing techniques to classify the COVID-19 related information by dividing the situational information into seven types. The author found certain feature in reposting the amount of each type of situational information. Subsequently, in consecutive studies [10,11], the authors predicted the future posts of an individual by using a clustering technique on the Reddit platform. They extracted posts from clinical subreddits and classify them to their resultant mental illness. The authors concluded that Reddit posts could be used to identify mental illness appear in the near future of an individual.

In study [12], the author has introduced an EAR technology which captured the ongoing behavioral data without relying on the self-reports. The provided data from the participants could be beneficial for researchers to improve their study related to human behavior and other psychological aspects. In study [13], Home Health Care (HHC) planning problem was introduced to integrate the patients and group-based care services assignment aspects and resource dimensioning issues. Author employed fuzzy c-means and performed sensitivity analysis. The results showed the effectiveness of the proposed method with respect to cost. In study [14], the author proposed a methodology by leveraging the supervised learner to rank the patients with respect to their data relevancy. Author concluded that LambdaMart perform well in terms of Normalized Discounted Cumulative Gain than Adaboost And Coordinate Ascent. Moreover, Mean Average Precision (MAP) was used to predict patients with chronic disease.

Finally, in study [15] the author extracted the data from Reddit and classified the data of chronic mental disease patients by adopting the C-training algorithm. The experimental results exhibited the performance of Co-training technique as compared to the state-of-the-art technique with a margin of 3%. From this discussion, we concluded that Reddit data is a better choice to train the classifiers in the context of mental illness diseases.

3 Semi-Supervised Co-Training Technique

We concluded from the research community that the number of structured labeled data is much lower than the number of unstructured labeled data especially in the context of social network computing.

In 1998, Blum and Mitchell [16] proposed a new learning technique that uses a smaller number of labeled data for first training by using a weak classifier. The working procedure of the co-training algorithm is shown in Fig. 2. Subsequently, the classifier label the new instances which could be considered as unlabeled data. After that, we can get samples of high confidence from weak classifiers. As a result, we placed these samples in a dataset and labeled dataset to

train the weak classifiers with the newly labeled data to increase the re-accuracy. Similarly, by taking advantage of co-training training algorithms [16,17], we construct a model with base level classifiers. Subsequently, we compared base level models with the constructed model. The features from different subreddit posts are sufficient for better classification accuracy.

Algorithm Co-Training()
Begin
 1. Repeat
 2. $m_i \leftarrow \text{train.classifier}(L_i)$
 3. $m_i \leftarrow \text{train.classifier}(L_i)$
 4. **for** $x \in U$ **do**
 5. if $\max m_i(x) > \tau$ **and** $\max m_i(x) < \tau$ **then**
 6. $L_i \leftarrow L_i \cup \{(x, p_i(x))\}$
 7. if $\max m_i(x) > \tau$ **and** $\max m_i(x) < \tau$ **then**
 8. $L_i \leftarrow L_i \cup \{(x, p_i(x))\}$
 9. **Until** no more predictions are confident based on one classifier
End

Figure 2: Co-training algorithm [16]

4 Proposed Methodology

We scrapped posts and top comments from the subreddit of Reddit. We choose posts which [18] were mentioned in the proposed work with the addition of OCD post.

From the studies [19–21], we concluded that NB, RF, and SVM are the best choice for text classification and multi-class classification. Their accuracy can be maintained through feature selection and extraction techniques. Therefore, we are building NB, RF, and SVM models for our problem domain [22,23]. We split the dataset in 50–30–20 ratio for the train, cross-validate, and test the classifiers. The training data summarized by the never seen data. Fig. 3 gives an overview of the proposed methodology which consists of several sections as follow.

4.1 Phase 1: Data Acquisition

The first phase of the proposed methodology is data acquisition. Reddit has limitation of extracting only 1000 posts from subreddits of Reddit. We used PRAW (Python API for Reddit) to extract 1000 top posts from the subreddit r/Anxiety, r/Depression, r/Bipolar, r/PTSD, r/Schizophrenia, and r/OCD. Few posts were deleted by the user from the subreddit therefore the total post count is 5887. Moreover, we extracted top comments from each of the subreddits which are used as unlabeled data.

Table 1: Selected subreddit and post

Subreddit	Post#	Subreddit	Post#
r/Anxiety	1000	r/PTSD	1000
r/Depression	1000	r/Schizophrenia	1000
r/Bipolar	1000	r/OCD	1000

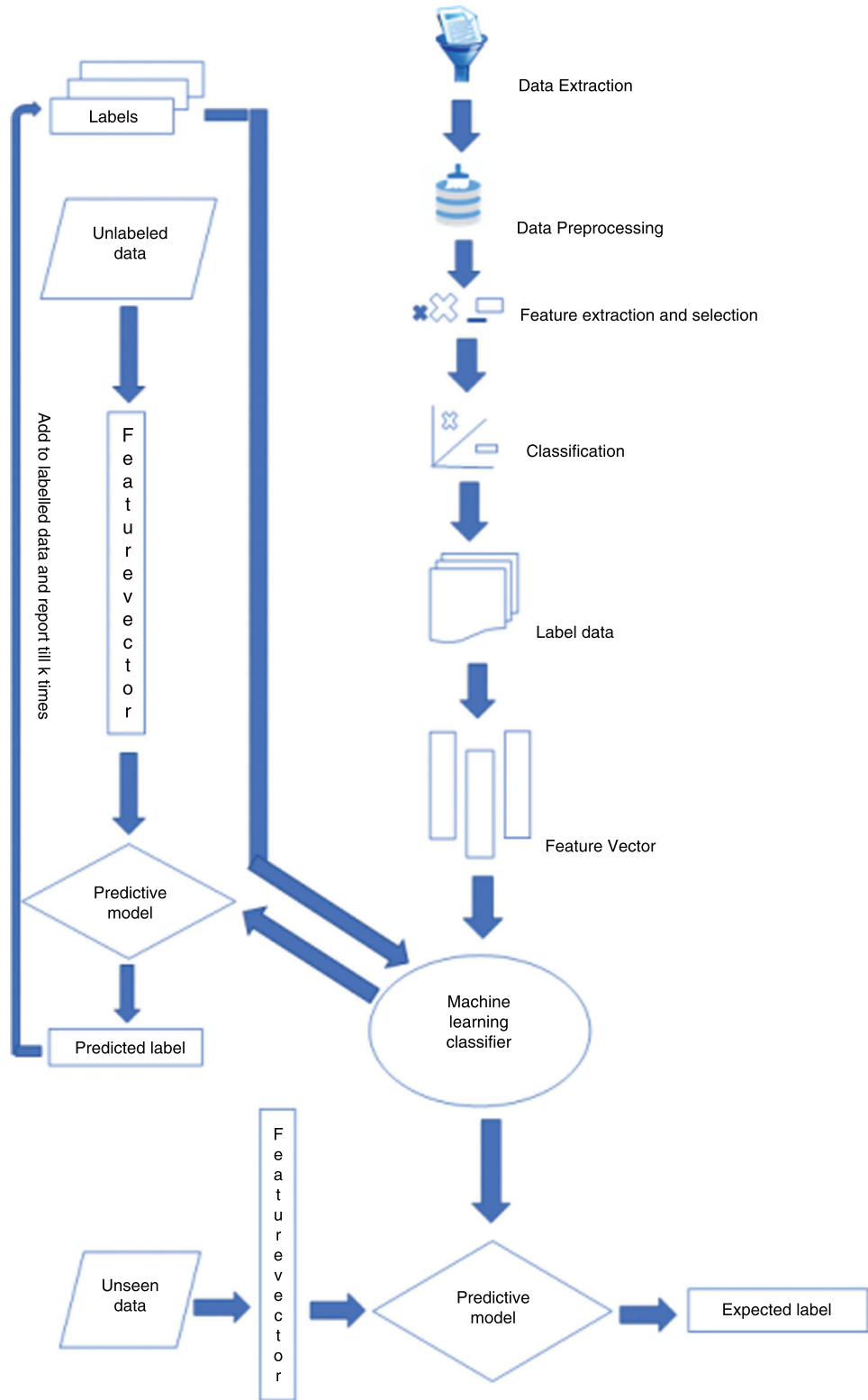


Figure 3: Overview of proposed methodology

For feature extraction, though several weighting could be performed such as Entropy, Term Frequency (TF), Binary and Length Term Collection (LTC). However, we applied the widely used Term Frequency-Inverse Document Frequency (TF-IDF) weighting method to preprocessed the data. TFIDF is processed into two parts. Firstly, the number of times a term used in the document is calculated (i.e., calculating Term Frequency). While the second part is calculating the inverse of TF which is known as IDF. The term IDF reduces the word weight that is frequently occurred. If we did not apply IDF after TF, then these most frequently occurred words will affect the accuracy of the classifier. From the given output, we select the features and split our data set into a 50–30–20 ratio. For training, cross-validation, and testing we used 50%, 30%, and 20% data respectively.

4.4 Phase 4: Classification

The fourth and the last phase of our proposed methodology is entitled as classification, which is accomplished by leveraging the capabilities of machine learning algorithms. Firstly, we look at the pre-built model NB, RF, and SVM by tuning their parameters and classifying the data. Subsequently, we leverage the capabilities of semi-supervised co-training techniques and construct NB, RF, and SVM that are based on the Co-training technique. For the semi-supervised Co-training technique, we need two views of a dataset. In this regard, we merged our feature set into labeled and unlabeled data from the dataset. Label data refer to the post of subreddits while unlabeled data refer to the comments from each post. Hence two models of NB, RF, and SVM were created and trained on the label data (i.e., post) for both views. Moreover, since unlabeled data is greater than label data so we consider them as weak classifiers. After training of weak classifier, we used some unlabeled data for prediction. We used those instances of each class in which these weak classifiers were most confident. Similarly, these most confident weak classifiers examples are added to labeled data. Subsequently, we trained these weak classifiers again on this new merged data that is labeled dataset. This process is recurring k epochs. For comparison, we took prediction after each iteration of the classifier.

5 Results

To evaluate the effectiveness of our proposed methodology, we performed several experiments by applying k-fold cross-validation. We employed the performance measures which are widely used in classification such as Precision (P), Recall (R), and F-measure. We need to construct a confusion matrix to derive the values of these metrics. The positive case (i.e., Yes for mental illness due to communication on COVID-19) of our model is the detection of mental illness of a patient through its communication as a COVID-19. While the negative case (i.e., No for a mental illness through communication on COVID-19) leads to a lack of mental illness. The same case is applied for each mental illness which is considered in the proposed study. The class-wise performance of classifiers when ensemble with the proposed co-training technique is shown in [Tab. 2](#). The second column of [Tab. 2](#) indicates the existence of a mental illness which is identified through the communication about COVID-19 on the social media platform. The last three columns indicate the value of the proposed measure used to assess the effectiveness of the proposed method. The average performance of co-training-based classifiers for the identification of all mental illness is shown in [Tab. 3](#). The main consequences of the experimental results of [Tabs. 2](#) and [3](#) are as follow.

- Since the positive case of our proposed model is the identification of mental illness, so in this case we perceived the highest F-measure value ([Tab. 2](#)) of SVM (F-measure = 0.98),

RF (F-measure = 0.99), and NB(F-measure = 0.94). The highest value of the F-measure for each classifier indicates the performance of the proposed method for the identification of mental illness of patients through their communication in terms of their posts and comments.

- We observed RF (F-measure = 0.98 in [Tab. 3](#)) as an outperformed co-training-based classifier as compared to SVM and NB, which indicate its capability to assign labels effectively.
- Though the performance of NB (F-measure = 0.95 in [Tab. 3](#)) is good to identify the mental illness when ensemble with co-training but it is not better than the rest of the classifiers.
- Though we observed good precision of all co-training-based classifiers but their recall values were not so good ([Tab. 2](#)) that could be considered as a threat to our co-training-based models.

Table 2: Class wise effectiveness of co-training based SVM, RF, NB for all mental illness classification

Classifier	Mental illness with COVID-19	Precision	Recall	F-measure
SVM	Yes	0.97	0.99	0.98
	No	0.99	0.62	0.80
RF	Yes	0.96	1.00	0.99
	No	0.90	0.72	0.81
NB	Yes	0.94	0.95	0.94
	No	0.80	0.72	0.76

Table 3: Average effectiveness of co-training based SVM, RF, NB for all mental illness classification

Classifier with co-training	F-measure
SVM	0.97
RF	0.98
NB	0.95

Moreover, to benchmark the effectiveness of our proposed approach i.e., Co-training with base classifiers, we evaluate the performance of state-of-the-artbase level classifiers with Co-training models. The results of [Tab. 4](#) exhibits the performance assessment of our Co-training models (SVM, NB, RF) in terms of F-measure. We observe the significant persistence in the effectiveness of the classification of mental illness. The main consequences of [Tab. 4](#) results are:

- We observe that co-training-based classifier outperformed the state-of-the-art classifiers especially co-training based NB well performed as compared to its individual use.
- We observe a 1 to 5% improvement in the classification decision of classifiers when they are incorporated with co-training based proposed method.
- We could not observe better improvement in the classification decision of RF (1%) as compared to other classifiers.

We observed better improvement in the classification decision of NB (5%) as compared to the rest of the classifiers.

Table 4: Comparison of classifier performance with and without co-training

Classifier	Performance with co-training	Performance without co-training
SVM	97%	93%
RF	95%	94%
NB	98%	93%

6 Threats to Validity

We observe some threats in our work. The first threat is related to the generalization of the result. We used a limited data set and few number of classifiers to present the efficacy of the proposed method. The result can be modified by changing the number of data sets and classifiers. The second threat is the use of classifiers with preset parameters. The efficacy of our proposed method may be altered by tuning the default parameters.

7 Conclusion

From the experimental results, we conclude the performance of the co-training based proposed methodology for the classification as compared to the state-of-the-art classifiers. The proposed co-training-based classifiers are evaluated in the context of COVID-19 by considering posts and comments of human mental illness on the Reddit social media platform. We used PRAW (Python API for Reddit) for the extraction of the top 1000 posts and first comments from each post. We have performed several experiments to classify the top post from subreddit Anxiety, Depression, Bipolar, PTSD, Schizophrenia and OCD by using Naïve Bayes (NB), Support Vector Machine (SVM), and Random Forest (RF) classifiers. The main consequences are; (1) We observed the effectiveness of the proposed method for classifying the human mental illness on the basis of their posts and comments in terms of accuracy, recall, and F-measure. Such as, we observed the performance of co-training based RF as the F-measure = 0.99, (2) We observed the improvement in the classification decision through co-training based classifiers as compared to their state-of-the-art usage, as the performance of NB significantly improved (5%) when it works through the proposed method, (3) We observed a 1% to 5% improvement in the mental illness classification that may varied in the proposed study when other classifications are considered. In future, we are planing to consider more human posts and comments using several social media platforms.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declares that they have no conflicts of interest to report regarding the present study.

References

- [1] T. A. Ghebreyesus, “SARS (Severe Acute Respiratory Syndrome),” WHO, 2020. [Online]. Available: <https://www.who.int/diseases/sars/en/?fbclid=IwAR3MjttUuAfr8KnJ5cbv2fjYpqe9kfzUnBOvMZ71d8kWNqk2zKEaEj5WbOQ>.
- [2] M. Serbini, “Virus taxonolomy: The ICTV report on virus classification and taxon nomenclature,” *ICTV*, 2020. [Online]. Available: <https://talk.ictvonline.org/>.
- [3] T. A. Ghebreyesus, “Cronavirus disease2019 outbreak,” WHO, 2020. [Online]. Available: <https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic>.
- [4] K. H. Manguri, R. N. Ramadhan and P. R. Mohammed Amin, “Twitter sentiment analysis on worldwide COVID-19 outbreaks,” *Kurdistan Journal of Applied Research*, vol. 2020, pp. 54–65, 2020.
- [5] L. Chen, H. Lyu, T. Yang, Y. Wang and J. Luo, “In the eyes of the beholder: Sentiment and topic analyses on social media use of neutral and controversial terms for COVID-19,” Cornell University, pp. 3–6, 2020, [Online]. Available: <http://arxiv.org/abs/2004.10225>.
- [6] B. Kleinberg, I. van der Vegt and M. Mozes, “Measuring emotions in the COVID-19 real world worry dataset,” vol. 1, Cornell University, 2020. [Online]. Available: <http://arxiv.org/abs/2004.04225>.
- [7] N. Karisani and P. Karisani, “Mining coronavirus (COVID-19) posts in social media,” Cornell University, pp. 2–5, 2020. [Online]. Available: <http://arxiv.org/abs/2004.06778>.
- [8] V. Duong, P. Pham, T. Yang, Y. Wang and J. Luo, “The Ivory tower lost: How college students respond differently than the general public to the COVID-19 pandemic,” Cornell University, 2020. [Online]. Available: <http://arxiv.org/abs/2004.09968>.
- [9] L. Li, Q. Zhang, X. Wang, J. Zhang, T. Wang *et al.*, “Characterizing the propagation of situational information in social media during COVID-19 epidemic: A case study on Weibo,” *IEEE Transactions Computational Social Systems*, vol. 7, no. 2, pp. 556–562, 2020.
- [10] R. Thorstad and P. Wolff, “Predicting future mental illness from social media: A big-data approach,” *Behavioral Research Methods*, vol. 51, no. 4, pp. 1586–1600, 2019.
- [11] G. Mujtaba, L. Shuib, R. G. Raj, R. Rajandram and K. Shaikh, “Prediction of cause of death from forensic autopsy reports using text classification techniques: A comparative study,” *Journal of Forensic and Legal Medicine*, vol. 57, no. 8, pp. 41–50, 2018.
- [12] M. R. Mehl, J. W. Pennebaker, T. G. A. Technologies and J. H. Price, “The electronically activated recorder (EAR): A device for sampling naturalistic daily activities and conversations,” *Behavioral Research Methods, Instruments & Computers*, vol. 33, no. 4, pp. 517–523, 2001.
- [13] A. J. Nasir, S. Hussain and H. Dang, “Integrated planning approach towards home health care, telehealth and patients group-based care,” *Journal of Network and Computer Applications*, vol. 117, pp. 30–41, 2018.
- [14] S. Hussain, H. Afzal, M. R. Mufti, A. Khan, A. A. Khan *et al.*, “A methodology to rank the patients prone to chronic diseases in telehealth,” *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 3, pp. 418–425, 2019.
- [15] S. Tariq, N. Akhtar, H. Afzal, S. Khalid, M. R. Mufti *et al.*, “A novel co-training-based approach for the classification of mental illnesses using social media posts,” *IEEE Access*, vol. 7, pp. 166165–166172, 2019.
- [16] A. Blum and T. Mitchell, “Combining labeled and unlabeled data with co-training,” in *Proc. of the Eleventh Annual Conf. on Computational Learning Theory (COLT’98)*, pp. 92–100, 1998.
- [17] S. Ruder, “An overview of proxy-label approaches for semi-supervised learning,” 2018. [Online]. Available: <https://ruder.io/semi-supervised/>.
- [18] D. M. Low, L. Rumker, T. Talkar, J. Torous, G. Cecchi *et al.*, “Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during COVID-19,” *Journal of Medical Internet Research*, vol. 22, no. 10, pp. 1–41, 2020.
- [19] S. Hussain, J. Keung, M. K. Sohail, M. Ilahi and A. A. Khan, “Automated framework for classification and selection of software design patterns,” *Applied Soft Computing*, vol. 75, pp. 1–20, 2019.

- [20] W. Hadi, Q. A. Al-Radaideh and S. Alhawari, "Integrating associative rule-based classification with naïve bayes for text classification," *Applied Soft Computing*, vol. 69, pp. 344–356, 2018.
- [21] J. Hartmann, J. Huppertz, C. Schamp and M. Heitmann, "Comparing automated text classification methods," *International Journal of Research in Marketing*, vol. 36, no. 1, pp. 20–38, 2019.
- [22] S. Hussain, H. Afzal, M. R. Mufti, M. Imran, A. Ali *et al.*, "Mining version history to predict the class instability," *PLoS One*, vol. 14, no. 9, pp. 1–21, 2019.
- [23] S. Hussain, J. Keung and A. A. Khan, "Software design patterns classification and selection using text categorization approach," *Applied Soft Computing*, vol. 58, pp. 225–244, 2017.