

Building and Testing Fine-Grained Dataset of COVID-19 Tweets for Worry Prediction

Tahani Soud Alharbi, Fethi Fkih

Department of Computer Science, College of Computer
Qassim University, Buraydah 51452, Saudi Arabia

Abstract—The COVID-19 outbreak has resulted in the loss of human life worldwide and has increased worry concerning life, public health, the economy, and the future. With lockdown and social distancing measures in place, people turned to social media such as Twitter to share their feelings and concerns about the pandemic. Several studies have focused on analyzing Twitter users' sentiments and emotions. However, little work has focused on worry detection at a fine-grained level due to the lack of adequate datasets. Worry emotion is associated with notions such as anxiety, fear, and nervousness. In this study, we built a dataset for worry emotion classification called "WorryCov". It is a relatively large dataset derived from Twitter concerning worry about COVID-19. The data were annotated into three levels ("no-worry", "worry", and "high-worry"). Using the annotated dataset, we investigated the performance of different machine learning algorithms (ML), including multinomial Naïve Bayes (MNB), support vector machine (SVM), logistic regression (LR), and random forests (RF). The results show that LR was the optimal approach, with an accuracy of 75%. Furthermore, the results indicate that the proposed model could be used by psychologists and researchers to predict Twitter users' worry levels during COVID-19 or similar crises.

Keywords—COVID-19; sentiment analysis; emotion analysis; worry dataset; concern analysis

I. INTRODUCTION

At the end of the year 2019, China reported cases of pneumonia caused by an unknown virus in Wuhan City. Later, this pneumonia was defined by the World Health Organization (WHO) as the coronavirus disease 2019 (COVID-19)[1]. It was then declared a pandemic that has had multiple consequences, including the death and long-term effects of infected people. According to WHO, as of July 2022, the total number of reported COVID-19 cases was approximately 545 million, with a total of 6.3 million deaths¹. The uncertainty and low predictability of COVID-19 threaten people's both physical and mental health, especially in terms of emotions and cognition [2]. The most challenging effects of the pandemic, especially during lockdowns, are depression, anxiety, and worries due to unemployment, losing loved ones, or being personally affected by the disease [3]. While there are several programs that psychologists and therapists carry out to enable recovery from these issues, there is an immense need to study worry using other sources [4]. Traditional methods of public health monitoring, like questionnaires and clinical tests, have certain limitations; for example, they only cover a

limited number of participants and are restricted to the data collection period[5].

In contrast, social media are becoming a significant source of rich real-time information during crises, including disease outbreaks and natural disasters [6]. Twitter is a unique source of big data for public health researchers due to the real-time nature of the content and the ease of searching and accessing publicly available data [7]. In this vein, COVID-19-related behaviors and sentiments are available on social media. Twitter users continuously post about their feelings and worries regarding these unusual circumstances[8]. This situation drew the attention of computer scientists and researchers, leading to numerous studies on the understanding of the emotional states during current events, especially those related to the pandemic [9].

The research problem is related to the discrimination of the worry analysis studies. Most of the researchers have focused on discrete emotion theories, like Ekman's emotion classification schema [10], by annotating texts to the six basic emotions (i.e., anger, disgust, fear, happiness, sadness, and surprise) [11]. As the most dominant emotions during crises are worry and anxiety [12], [13], the existing methods for emotion detection are insufficient to capture the emotion of worry accurately [14].

Detecting worry is complex as people are either unwilling to disclose worries to medical personnel or prefer sharing their feelings on social media. Thus, there is a lack of datasets that could be used for worry analysis, as many studies depend on surveys and interviews [15], [16]. To the best of the authors' knowledge, this is the first study to build a to-date dataset about COVID-19-related worries that is to be applied to machine learning (ML) models. In the context of this paper, worry about COVID-19 is classified into three fine-grained levels: "no-worry", "worry", and "high-worry". The "no-worry" category includes people discussing the news and politics about the virus or content-containing statistics and figures. On the other hand, people expressing high levels of feelings such as panic or fear ("high worry" category) are distraught. Between these two categories ("worry" category), there are people expressing concern about the virus, who are considered stressed about the present and the future.

The contribution of this paper is two-fold. First, the WorryCov² dataset was built based on three classes: "no-worry", "worry", and "high-worry". It was built with experts

¹<https://covid19.who.int>

²Dataset is available from the authors upon reasonable request.

in linguistics and followed an annotation scheme under strict quality control. Then, several ML classification models were used to test the dataset.

The paper is outlined as follows. The related works are discussed in Section 2. Section 3 introduces the proposed approach. Section 4 provides the results and discussion, while Section 5 concludes the paper.

II. RELATED WORK

Worry analysis is considered one dimension of emotion analysis frequently studied in the literature. Therefore, this study focuses on concern, sentiment, and emotion analysis towards or during disasters or pandemics such as the COVID-19 pandemic.

Much previous research was carried out to determine the public health concerns toward disasters or epidemics based on sentiment analysis results. For example, the work in [17] aimed to analyze Twitter messages relating to Hurricane Irene and trained a dataset based on sentiment analysis classifiers to categorize tweets into levels of concern. They evaluated the impact of various tokenization strategies and feature choices like a bag of words (BOW) and lexicons on classification accuracy. With 84.27% accuracy, the best settings for the maximum entropy classifier were removing punctuation, converting the text to lowercase, removing stop words, and building a worry lexicon. The Epidemic Sentiment Monitoring System [5] provides visualization tools for Twitter posts responding to public concerns about different diseases. The degree of concern reported that multinomial Naïve Bayes (MNB) achieved the highest F1-score using term frequency-inverse document frequency (TF-IDF) features. To measure and monitor public health concerns about communicable diseases, a sentiment classification approach was applied to Twitter data by measuring different levels of concern [18]. The classifier was trained with a dataset automatically generated by a programming system using an emotion-oriented and clue-based method. Three ML classifiers were evaluated, with the NB classifier achieving the best accuracy for the epidemic-related dataset.

Regression is often used to detect public health concerns. For instance, in [19], a strategy to predict to what extent news about a public health issue can be disseminated was proposed using a data collection of microblog news posts. This ML method relies on the logistic regression (LR) algorithm that automatically categorized news posts into two classes: normal news or news posts that resonated with widespread public anxiety.

As for COVID-19, abundant works have already been published studying the effects of this pandemic on various aspects. For example, most research focused on analyzing Twitter data and finding the main critical topics that raise concerns for individuals regarding the COVID-19 pandemic. In [20], [21] used the topic modeling technique LDA (an unsupervised machine learning model) to identify the most common topics in the tweets and performed sentiment analysis. Furthermore, analyzing citizens' concerns during the COVID-19 epidemic has been studied in [22]. 30,000 COVID-19-related tweets were collected from March 14,

2020. Each tweet was labeled as very negative, negative, neutral, very positive, and positive by using the natural language processing (NLP) library. Then, the authors used sentiment analysis on pre-processed tweets to show the level of concern in various US states. They presented an approach for measuring citizens' concern levels through Twitter data by using the ratio of very negative and negative tweet counts over the total number of tweets in the dataset. As a result, school closing-related tweets cause the highest level of concern among citizens. Similarly, the study [23] presented a method to identify the COVID-19 topic's degree of concern through user conversations on Twitter based on two phases of the classification process. The first classification step is to separate tweets into two classes, namely COVID-19 and non-COVID-19. The second step is to classify the COVID-19 data into seven topics: donations, emotional support, warnings and suggestions, hoaxes, notification of information, seeking help, and criticism. Six pairs of combinations of word-level and character-level word embeddings, namely Word2Vec and fastText, with three deep learning models, CNN, RNN, and LSTM, were used to apply the text classification model. The best accuracy was achieved when fastText and LSTM were used together for both stages of classification, with 97.3% and 99.4%, respectively.

Significant research in public health has applied emotion analysis using social media-derived information to monitor public emotions during disease outbreaks. Emotions such as anxiety, anger, happiness, desire, disgust, fear, relaxation, and sadness have been widely studied. Emotions are often linked with topic modeling to identify the topics and their intensity level. For example, findings in [12] indicate that the longer texts gave insights into what people worry about during the pandemic: the economy and the family. In the SenWave system [8], seven fine-grained sentiment categories, namely, optimistic, thankful, empathetic, pessimistic, anxious, sad, annoyed, denial, official, and joking, are used to study the concern of Twitter users from different countries. The labeled tweets are used to train the deep learning language models such as XLNet, AraBert, and ERNIE, while over 105 million unlabeled tweets are used for the testing process. An XLNet pre-trained language model was used for English tweets. The classifier achieved an 80% accuracy, which proves the efficiency of the models. However, emotion analysis studies are minimal compared with sentiment research due to the lack of annotated data [24]. The EmoBERT model [24] was used to capture emotions related to emotional health (annoyed, anxious, empathetic, sad) to compare emotions expressed on social media before and during the COVID-19 epidemic. In comparison to BERT and XLNet, EmoBERT achieved better results.

Our review shows that little research has addressed worry detection. However, many studies address anxiety as an issue of mental health, for instance, this study [25] utilized personal narratives from Reddit to detect anxiety disorders and classified anxiety-related posts into a binary level of anxiety. Using various linguistic features, including vector-space representations (Word2Vec and Doc2Vec), topic (LDA) models, Linguistic Inquiry and Word Count (LIWC) dictionary, and n-gram language models. Overall, all features

that have been used succeeded in classifying the level of anxiety, for single-source features, using Neural Network with N-gram probabilities achieved slightly better accuracy (92%) compare with using SVM with word-vector embeddings (word2vec), and for combined features, Neural Network has produced the highest accuracy of 98% by aggregating LIWC with word2vec embeddings and by aggregating N-gram features with LIWC. Moreover, this paper [26] developed its own binary classification dataset for detecting anxiety and depression users on social media who have not yet been diagnosed with mental illness. The authors have presented a comparative experimental evaluation using the traditional linear model and pre-trained LMs (language models). Their results showed that LMs (BERT and ALBERT) performed relatively well with balanced training data. However, in unbalanced training sets, Support Vector Machine (SVM) with word embeddings and TF-IDF features performed slightly better overall, with 0.750 F1-score, 0.747 for accuracy, and 0.740 for precision.

To our knowledge, Verma et al.'s study [14] is the most relevant to the prediction of worry using Twitter data. Using crowdsourcing, they re-annotated an existing dataset that contains four emotions (joy, anger, fear, and sadness) [27] for worry classification. A wide range of machine learning and deep learning models were evaluated. For traditional ML approaches, Multinomial Naive Bayes (MNB), and Support Vector Machine (SVM) are implemented by using feature-based models. For deep learning based on word embeddings, they used Hierarchical Attention Network (HAN) and CNN-static with combined Glove emoji2vec embeddings. While deep learning approaches based on contextual embeddings were also applied like RoBERT and XLNet. The results showed that deep learning methods outperform as compared to the traditional models for worry identification with 0.61 F1-score.

The gap in the current studies is related to the lack of a large new dataset for worry level detection related to COVID-19 tweets. Despite the many works, most previous results focus on the sentiment classification of tweets as positive, negative, and natural.

III. PROPOSED APPROACH

The proposed approach is shown in Fig. 1. Due to the limited dataset related to worry identification from text, we built a dataset and chose the classification task. We decided to select some machine learning models to validate the credibility of the collected dataset. The approach first described the data collection and the annotation process into three levels of worry using COVID-19-related tweets. The dataset was then used to extract features, run ML models, and evaluate the results.

A. Building the Benchmark Dataset

1) *Dataset collection and filtering:* To build the benchmark dataset, tweets were collected, filtered, and annotated. Twitter is one of the most popular social media and has a wide range of content including rich text, emojis, and hashtags [14]. The tweets related to COVID-19 were collected using Tweepy, the Python Twitter API library [28]. Initially, we used unified query keywords (i.e., coronavirus, covid-19, #coronavirus, and #covid-19), previously used in other studies [29], to identify the tweets related to COVID-19. The tweets were collected over three periods to ensure that they covered significant milestones during the pandemic. The three periods are consistent with [30] and are the following:

- First period: from January 30 to February 28, 2020. During this period, the first COVID-19-induced death was reported in China, and WHO announced a public health emergency.
- Second period: from March 29 to April 29, 2020. During this period, WHO declared COVID-19 a worldwide pandemic, leading many governments to impose restrictions on citizens in an attempt to reduce the spread of the virus.
- Third period: from May 10 to June 30, 2020. During this period, COVID-19 had spread globally, with an increased number of confirmed cases and deaths.

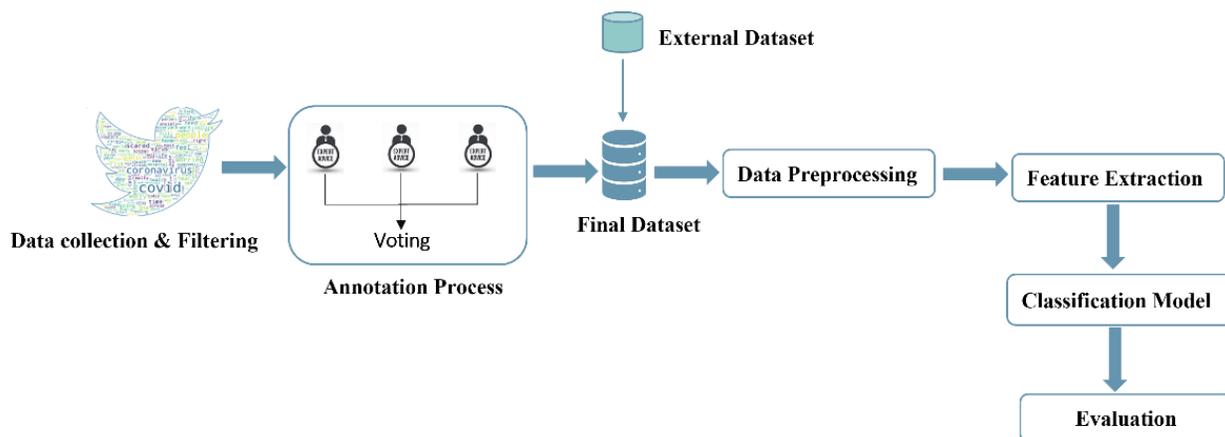


Fig. 1. Proposed Approach.

Following these periods and using the aforementioned keywords, 270,000 tweets were collected. Each tweet had 24 columns, including data and time, username, tweet text, and location. Since we wanted to detect feelings of worry at the tweet level, we removed the rest of the columns and only retained the text column. However, a large proportion of COVID-19-related tweets were probably not associated with one emotion; thus, annotating them would be costly and ineffective [31], [32]. To meet our objective, we focused solely on the worry emotion and used worry-related keywords to create a dataset of tweets representing this emotion. Following [33], we selected keywords (terms) to filter the collected data. The terms were extracted from Thesaurus.com by finding synonyms and terms related to worry; the dictionary is one of the trusted, free online dictionaries. The synonym keywords are shown in Table I.

Often, datasets contain noise and irrelevant text. Therefore, the following rules were applied to reduce the dataset to more concise and related tweets: (1) deleting duplicate tweets (i.e., retweeted by other users), (2) deleting non-English language tweets, and (3) deleting all tweets less than 40 characters (short tweet).

TABLE I. KEYWORDS USED FOR FILTERING TWEETS (EXTRACTED FROM THE ONLINE THESAURUS.COM DICTIONARY)

worry	anxiety	concern	apprehension	fear	afraid
worried	anxious	panic	stress	tension	terrify
worries	distress	nervous	uncertain	tense	terrified
scary	confusion	restless	doubt	horrible	terror
scared	confuse	pressure	uptight	horror	paranoid
discomfort	troubled	pain	upset	dread	alarm

2) *Annotation process*: Manual labeling of social media data is challenging and requires dedicated time from domain experts (time-consuming). However, it is a critical part of the data preparation process in supervised learning. We annotated the data for not just coarse classes (such as worry or no-worry) but also for fine-grained levels indicating the intensity or degrees of emotion. However, annotating instances for degrees of emotions is a more difficult task to ensure annotation consistency [33]. Therefore, this study followed a set of rules to overcome this challenge: (1) tweets were annotated to three classes only: “no-worry”, “worry”, and “high-worry”, (2) Three English speakers with more than three years of experience in linguistics were employed; (3) the majority vote was used to annotate an individual tweet, and when the three experts disagreed, the tweet was considered irrelevant and was removed from the dataset; moreover, a newly developed website application was used to help the annotators accomplish their work; and (4) each annotator got the same number of tweets (2,700) for each month of the three periods (8,100 tweets in total). This process was slow but ensured results in accordance with the following guidelines to classify each tweet:

- “No-worry” class:
 - News or politics (i.e., conspiracy theories, China-related discussion where the Chinese are blamed for the virus, US politics, critics of Donald Trump, etc.) and facts (e.g., numbers, statistics).
Example: “China's outbreak is serious. But flu killed *5000 Americans* in the 1st 2wks of 2020 coronavirus infected 6, killed 0. Not sure what info you have that CDC director doesn't "The immediate risk to the US public is low." Our US readers deserve to know they don't need to panic.”
 - Other diseases (i.e., tweets comparing COVID with other diseases, discussing symptoms and mortality rates).
Example: “Aids is a killer disease Cancer is a killer disease Ebola is a killer disease Swine flu is a killer disease The only thing that divides Coronavirus to this other diseases is the fact that it is just the latest, stop with the panic and take care of yourself! #coronavirus.”
 - Expressing some other emotions (i.e., tweets denying the existence of the virus or expressing any optimistic/positive attitude toward it).
Example: “markets are full of pads soap Dettol, etc. People are not freaking out as they know there's enough. They aren't crazy buying. Let's hope the panic ends soon all over the world and we live happily again #covid-19.”
- “Worry” class:
 - Expressing general concern (i.e., mentioning being worried/stressed/concerned about the present and future of COVID-19).
Example: “Also, I'm young and healthy and unlikely to die from covid-19, so no reason to be afraid at all for me. I'm nervous about infecting those who are less likely to survive though, so I will do my best to prevent that of at all possible if I get infected.”
- “High-worry” class:
 - Expressing concern (i.e. tweets expressing feelings of panic, fear, etc.).
Example: “i'm tired of crying. i'm tired of the anxiety, and panic attacks. i want to go outside again. please - STAY HOME. #COVID-19 #COVID19Ontario.”
 - Frequent use of intensifiers (e.g., extremely, so, very) and featuring content related to (fear of) death.
Example: “So much stress, so much anxiety, AND I'M PREGNANT. Headaches all day, puking many times a day, quarantined. People are dying, this is not cool. #coronavirus”

The annotation resulted in 7,861 instances corresponding to the three classes. The “no-worry” class included 3,158 instances, the “worry” class had 3,127 instances, and the “high-worry” class included 1,576 instances. The remaining 239 tweets were eliminated as the annotators disagreed with their classification (not sure). However, we noticed that the WorryCov dataset is imbalanced. So, it should be solved to reduce skewness and increase the performance of ML models [34]. Therefore, we decided to expand the dataset using other external datasets. To our knowledge, no dataset focuses on only worry emotion. Therefore, we selected the intensity of anxiety based on [11] since it was considered a synonym for worry. Anxiety levels in [11] ranged from 1 to 9, where 1 was considered the lowest and 9 the highest. Considering this range, we chose the intensity levels 7, 8, and 9 as descriptive of the “high-worry” class, resulting in a total of 3,127 instances in the “high-worry” class.

B. Prediction of Worry Levels

The balanced benchmark dataset was used to evaluate the performance of different ML models. In this section, the data preprocessing, feature extraction, classification, and evaluation steps of this dataset are discussed.

1) *Data preprocessing*: Preprocessing generally improves the data quality by extracting meaningful fragments from a given text excluding the noise [35], [36]. Preprocessing steps include text cleaning such as URL, digit, punctuation removal, etc., and lemmatization.

In the cleaning step, we removed URLs, user mentions, and hashtags. Previous research on sample datasets shows that these items do not provide any evidence of the level of worry in tweets or useful information [37]. Next, each tweet was converted to lowercase to avoid considering the exact words as unique features, such as “HELP”, “Help”, or “help” will be converted to “help” [38]. Then, the contractions (i.e., “I’m” instead of “I am”) were replaced by the original phrase as described in [37]. Next, digits, punctuation marks, and extra spaces that do not provide any semantic information to the text were removed. NLP classification tasks often involve removing stop words to improve performance metrics [39]. However, in this dataset, worry feelings were frequently expressed as ideas about oneself, leading to the use of the “I” and “my” pronouns. Therefore, stop words were not removed to retain the linguistic characteristics of worried users. Finally, each word was lemmatized using Wordnet Lemmatizer available in the natural language toolkit (NLTK) library [40].

2) *Feature extraction*: Often called a features vector [37], this step refers to transforming raw data into numerical data that machines can understand. Term frequency-inverse document frequency (TF-IDF) is a popular text vectorization technique to generate vector representations of a text [41] and was employed in this experiment. The TF-IDF weighting scheme is based on two parts: term frequency (TF) and inverse document frequency (IDF). TF-IDF is mathematically formulated in the following (1) [42]:

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

where t denotes a term and d denotes a document.

TF is the frequency of any term within a given document and is calculated by dividing the number of mentions of a given word by the total number of words in the document [37], TF is defined by (2) [43]:

$$TF(t, d) = \frac{\text{Number of times the term } t \text{ appears in the document}}{\text{Total number of terms in the document}} \quad (2)$$

IDF represents the importance of a term in the corpus of the text. It is a technique that combined with TF reduces the impact of common words. There are some words, like “the”, “is”, “and”, etc., that occur frequently but are void of information. IDF is defined by Eq. (3) [44]:

$$IDF(t) = \log \left(\frac{\text{Number of documents}}{\text{Documents containing the term } t} \right) \quad (3)$$

C. ML-Based Classifiers

Four ML-based classifiers were used in the multi-classification task. These methods were multinomial NB (MNB), logistic regression (LR), Support Vector Machine (SVM), and Random Forest (RF). The default settings of these methods were taken from the scikit-learn library [45]. MNB is suitable for classifying discrete features or fractional counts such as TFIDF. LR calculates the likelihood of a target variable based on a collection of independent variables and a given dataset. SVM is a classification algorithm for two-group classification problems (in our case one-vs-rest scheme is used). Finally, the RF algorithm builds many random decision trees using bagging and feature randomness for each tree.

D. Evaluation Metrics

Each classifier was evaluated using the following performance measurements: accuracy, precision, recall, and F1-score. These standard metrics are defined as follows:

Accuracy is the ratio of the number of correct predictions to the overall number of predictions:

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FP + FN} \quad (4)$$

Precision is the ratio of the correctly predicted positive instances to the total positive instances:

$$\text{Precision (P)} = \frac{TP}{TP + FP} \quad (5)$$

Recall is the ratio of the correctly predicted positive instances to the total of all instances in the actual class:

$$\text{Recall (R)} = \frac{TP}{TP + FN} \quad (6)$$

F1-score is the harmonic average of precision and recall:

$$\text{F1-score} = \frac{2 \times P \times R}{P + R} \quad (7)$$

where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively.

IV. RESULTS AND DISCUSSION

After filtering (see Section 3.1), we obtain 15,000 tweets. Fig. 2 presents the word cloud of the most commonly used words in the WorryCov dataset. The most frequent keywords are related to the COVID-19 pandemic, such as “Covid”,

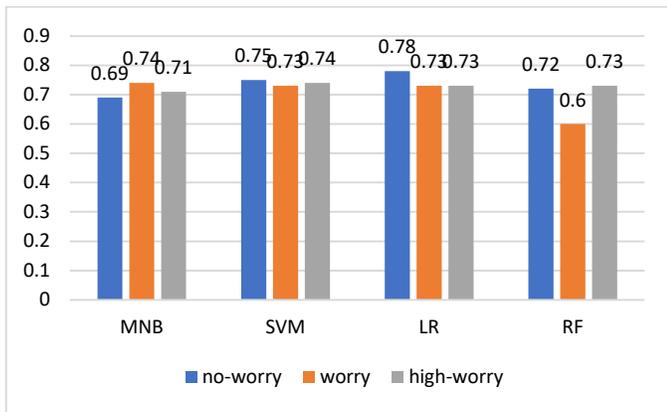


Fig. 6. Recall Results for the four ML Models per Worry Class.

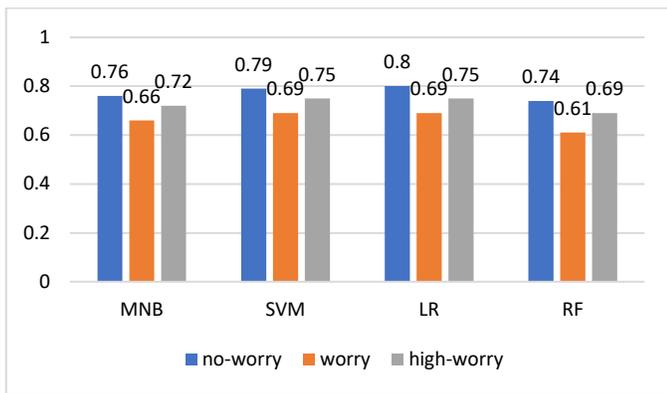


Fig. 7. F1-score Results for the Four ML Models per Worry Class.

In general, the current method, compared to Verma et al.'s study [14] is based on a new dataset. The new approach is also much more focused on the worry levels compared to 9 anxiety levels in [11], ranging from 1 to 9, where one was considered the lowest and nine the highest.

V. CONCLUSION

In this paper, we compiled a fine-grained benchmark dataset for the classification of worry levels concerning the COVID-19 pandemic. The dataset was collected from Twitter and was annotated using a majority vote among three experts. The WorryCov dataset was used to classify and predict the level of worry among Twitter users during the pandemic. Several experiments were conducted using the following ML algorithms: NB, LR, RF, and SVM. The optimal performance was achieved by LR, with an accuracy of 75%. It is recommended that the proposed approach be used for decision-making in healthcare entities to plan programs for the affected people. However, the current work has a few limitations. For example, the dataset is relatively small and was collected based on a short period from 2020–2021. Moreover, human behavior changes over time due to interaction with infected people, vaccination initiatives, and governments' health policies. Therefore, a new set of keywords that represent the new set of tweets might be needed to uncover the new trends in human worry levels. In the future, several deep learning models could be used to enhance the performance of the current approaches.

REFERENCES

- [1] C. Sohrabi et al., "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *International Journal of Surgery*, vol. 76. Elsevier Ltd, pp. 71–76, Apr. 01, 2020. doi: 10.1016/j.ijisu.2020.02.034.
- [2] S. Li, Y. Wang, J. Xue, N. Zhao, and T. Zhu, "The impact of covid-19 epidemic declaration on psychological consequences: A study on active weibo users," *Int J Environ Res Public Health*, vol. 17, no. 6, Mar. 2020, doi: 10.3390/ijerph17062032.
- [3] S. Zhang et al., "The COVID-19 Pandemic and Mental Health Concerns on Twitter in the United States," *Health Data Science*, vol. 2022, pp. 1–9, Feb. 2022, doi: 10.34133/2022/9758408.
- [4] V. Sideropoulos, H. Kye, D. Dukes, A. C. Samson, O. Palikara, and J. van Herwegen, "Anxiety and Worries of Individuals with Down Syndrome During the COVID-19 Pandemic: A Comparative Study in the UK," *J Autism Dev Disord*, 2022, doi: 10.1007/s10803-022-05450-0.
- [5] X. Ji, S. A. Chun, and J. Geller, "Monitoring public health concerns using twitter sentiment classifications," *Proceedings - 2013 IEEE International Conference on Healthcare Informatics, ICHI 2013*, no. September, pp. 335–344, 2013, doi: 10.1109/ICHI.2013.47.
- [6] U. Qazi, M. Imran, and F. Ofli, "GeoCoV19: A Dataset of Hundreds of Millions of Multilingual COVID-19 Tweets with Location Information," *SIGSPATIAL Special*, vol. 12, no. 1, pp. 6–15, Jun. 2020, doi: 10.1145/3404820.3404823.
- [7] L. Sinnenberg, A. M. Buttenheim, K. Padrez, C. Mancheno, L. Ungar, and R. M. Merchant, "Twitter as a tool for health research: A systematic review," *American Journal of Public Health*, vol. 107, no. 1. American Public Health Association Inc., pp. e1–e8, Jan. 01, 2017. doi: 10.2105/AJPH.2016.303512.
- [8] Q. Yang et al., "SenWave: Monitoring the Global Sentiments under the COVID-19 Pandemic," pp. 1–14, 2020.
- [9] A. H. Alamoodi et al., "Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review," *Expert Systems with Applications*, vol. 167. Elsevier Ltd, Apr. 01, 2021. doi: 10.1016/j.eswa.2020.114155.
- [10] "Paul Ekman. 1999. Basic emotions. In".
- [11] A. S. Imran, S. M. Daudpota, Z. Kastrati, and R. Batra, "Cross-cultural polarity and emotion detection using sentiment analysis and deep learning on covid-19 related tweets," *IEEE Access*, vol. 8, pp. 181074–181090, 2020, doi: 10.1109/ACCESS.2020.3027350.
- [12] B. Kleinberg, I. van der Vegt, and M. Mozes, "Measuring Emotions in the COVID-19 Real World Worry Dataset," vol. 1, 2020.
- [13] X. Li, M. Zhou, J. Wu, A. Yuan, F. Wu, and J. Li, "Analyzing COVID-19 on Online Social Media: Trends, Sentiments and Emotions," May 2020.
- [14] R. Verma, C. von der Weth, J. Vachery, and M. Kankanhalli, "Identifying Worry in Twitter: Beyond Emotion Analysis," pp. 72–82, 2020, doi: 10.18653/v1/2020.nlpccs-1.9.
- [15] R. A. Faisal, M. C. Jobe, O. Ahmed, and T. Sharker, "Replication analysis of the COVID-19 Worry Scale," *Death Stud*, vol. 46, no. 3, pp. 574–580, 2022.
- [16] P. J. Schulz, E. M. Andersson, N. Bizzotto, and M. Norberg, "Using Ecological momentary assessment to study the development of COVID-19 worries in Sweden: Longitudinal study," *J Med Internet Res*, vol. 23, no. 11, p. e26743, 2021.
- [17] B. Mandel, A. Culotta, J. Boulahanis, D. Stark, B. Lewis, and J. Rodrigue, "A demographic analysis of online sentiment during Hurricane Irene," *Proceedings of the 2012 Workshop on Language in Social Media*, no. Lsm, pp. 27–36, 2012.
- [18] X. Ji, S. A. Chun, Z. Wei, and J. Geller, "Twitter sentiment classification for measuring public health concerns," *Soc Netw Anal Min*, vol. 5, no. 1, pp. 1–25, 2015, doi: 10.1007/s13278-015-0253-5.
- [19] J. Pei, G. Yu, X. Tian, and M. R. Donnelley, "A new method for early detection of mass concern about public health issues," *J Risk Res*, vol. 20, no. 4, pp. 516–532, Apr. 2017, doi: 10.1080/13669877.2015.1100655.

- [20] A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hai, and Z. Shah, "Top concerns of tweeters during the COVID-19 pandemic: A surveillance study," *J Med Internet Res*, vol. 22, no. 4, Apr. 2020, doi: 10.2196/19016.
- [21] Et al. Song M, Emilsson L, Bozorg SR, Nguyen LH, Joshi AD, Staller K, "Understanding Weekly COVID-19 Concerns through Dynamic Content-Specific LDA Topic Modeling," *Lancet Gastroenterol Hepatol*, vol. 5, no. 6, pp. 537–547, 2020, doi: 10.18653/v1/2020.nlpcss-1.21.Understanding.
- [22] S. A. Chun, A. C. Y. Li, A. Toliyat, and J. Geller, "Tracking citizen's concerns during COVID-19 pandemic," in *ACM International Conference Proceeding Series*, Jun. 2020, pp. 322–323. doi: 10.1145/3396956.3397000.
- [23] N. A. Hasanah, N. Suciati, and D. Purwitasari, "Identifying degree-of-concern on covid-19 topics with text classification of twitters," *Register: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 7, no. 1, pp. 50–62, 2021, doi: 10.26594/register.v7i1.2234.
- [24] O. T. Aduragba, J. Yu, A. I. Cristea, and L. Shi, "Detecting Fine-Grained Emotions on Social Media during Major Disease Outbreaks: Health and Well-being before and during the COVID-19 Pandemic," *AMIA Annu Symp Proc*, vol. 2021, pp. 187–196, 2021.
- [25] J. H. Shen and F. Rudzicz, "Detecting Anxiety through Reddit," *Association for Computational Linguistics*, 2017.
- [26] D. Owen, J. C. Collados, and L. Espinosa-Anke, "Towards Preemptive Detection of Depression and Anxiety in Twitter," Nov. 2020.
- [27] S. M. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, "SemEval-2018 Task 1: Affect in Tweets," *NAACL HLT 2018 - International Workshop on Semantic Evaluation, SemEval 2018 - Proceedings of the 12th Workshop*, pp. 1–17, 2018, doi: 10.18653/v1/s18-1001.
- [28] J. Roeslein, "Tweepy: Twitter for Python!," URL: <https://github.com/tweepy/tweepy>, 2020.
- [29] U. Naseem, I. Razzak, M. Khushi, P. W. Eklund, and J. Kim, "COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis," *IEEE Trans Comput Soc Syst*, vol. 8, no. 4, pp. 976–988, Aug. 2021, doi: 10.1109/TCSS.2021.3051189.
- [30] Y. S. Malik et al., "Coronavirus Disease Pandemic (COVID-19): Challenges and a Global Perspective," *Pathogens*, vol. 9, no. 7, 2020, doi: 10.3390/pathogens9070519.
- [31] M. Saif and S. Kiritchenko, "Understanding emotions: A dataset of tweets to study interactions between affect categories," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, Miyazaki, Japan, 2018, pp. 7–12.
- [32] T. Sosea, C. Pham, A. Tekle, C. Caragea, and J. J. Li, "Emotion analysis and detection during COVID-19," *ArXiv*, 2021.
- [33] S. M. Mohammad and F. Bravo-Marquez, "Emotion intensities in tweets," *arXiv preprint arXiv:1708.03696*, 2017.
- [34] D. Ramyachitra and P. Manikandan, "Imbalanced dataset classification and solutions: a review," *International Journal of Computing and Business Research (IJCBR)*, vol. 5, no. 4, pp. 1–29, 2014.
- [35] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu, and V. Varadarajan, "Ai based emotion detection for textual big data: Techniques and contribution," *Big Data and Cognitive Computing*, vol. 5, no. 3, 2021, doi: 10.3390/bdcc5030043.
- [36] N. Babanejad, A. Agrawal, A. An, and M. Papagelis, "A Comprehensive Analysis of Preprocessing for Word Representation Learning in Affective Tasks," pp. 5799–5810, 2020, doi: 10.18653/v1/2020.acl-main.514.
- [37] U. Naseem, I. Razzak, S. K. Khan, and M. Prasad, "A Comprehensive Survey on Word Representation Models: From Classical to State-of-the-Art Word Representation Language Models," *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, vol. 20, no. 5, pp. 1–46, Jun. 2021, doi: 10.1145/3434237.
- [38] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Information (Switzerland)*, vol. 10, no. 4, pp. 1–68, 2019, doi: 10.3390/info10040150.
- [39] S. Adhikari et al., "Exploiting linguistic information from Nepali transcripts for early detection of Alzheimer's disease using natural language processing and machine learning techniques," *International Journal of Human Computer Studies*, vol. 160, Apr. 2022, doi: 10.1016/j.ijhcs.2021.102761.
- [40] E. Loper and S. Bird, "NLTK: The Natural Language Toolkit," 2002. [Online]. Available: <http://nltk.sf.net/>.
- [41] H. Du et al., "Twitter vs News: Concern analysis of the 2018 California wildfire event," *Proceedings - International Computer Software and Applications Conference*, vol. 2, pp. 207–212, 2019, doi: 10.1109/COMPSAC.2019.10208.
- [42] Y. Didi, A. Walha, and A. Wali, "COVID-19 Tweets Classification Based on a Hybrid Word Embedding Method," *Big Data and Cognitive Computing*, vol. 6, no. 2, p. 58, 2022, doi: 10.3390/bdcc6020058.
- [43] K. S. Kalaivani, S. Uma, and C. S. Kanimozhiselvi, "A Review on Feature Extraction Techniques for Sentiment Classification," *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020*, no. Iccmc, pp. 679–683, 2020, doi: 10.1109/ICCMC48092.2020.ICCMC-000126.
- [44] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The impact of features extraction on the sentiment analysis," *Procedia Comput Sci*, vol. 152, pp. 341–348, 2019, doi: 10.1016/j.procs.2019.05.008.
- [45] O. Kramer, "Scikit-learn," in *Machine learning for evolution strategies*, Springer, 2016, pp. 45–53.