

MINDS: A Multi-label Emotion and Sentiment Classification Dataset Related to COVID-19

Anjali Bhardwaj¹, Muhammad Abulaish²

Department of Computer Science, South Asian University, New Delhi, India

Abstract

During times of crisis, such as the COVID-19 pandemic, there has been a sudden increase in information exchanges on social media. People gathered to share their feelings, experiences, knowledge, and ideas with one another. Twitter has emerged as the most authentic, admired, and widely used social media platform for users to express their sentiments, opinions or emotions. Understanding the emotions expressed in text facilitates the development of empathetic machines. In this paper, we intend to collect and annotate a large corpus of textual data that can be used to train classification models for detecting emotions and sentiments in user-generated content. We analyze user sentiments and emotions expressed in tweets during the third wave of the omicron sub-variant pandemic. Our curated dataset, MINDS, consists of 227,229 tweet instances that were annotated using a multi-model setup in order to quantify all aspects of model uncertainty. Each instance in the dataset is classified according to three sentiment classes (*positive*, *negative*, and *neutral*) and five emotion classes (*sadness*, *joy*, *fear*, *disgust*, and *anger*). We have used a well-performing benchmark dataset related to SemEval-2018's E-c subtask for determining the classification threshold. The MINDS datasets is publicly available at <http://www.abulaish.com/lds/dataset> for research purposes.

Keywords

Emotion dataset, Sentiment dataset, Multi-label classification, NLU, COVID-19, IBM Watson

1. Introduction

Human emotion is one of the fundamental components of cognitive processes. Without emotions, humans would be lifeless stones; emotions are what keep us alive and define who we are. Changes in the physiological aspect frequently convey emotions, as they correspond to psychological states that occur spontaneously and without conscious effort. It is a complex feeling caused by internal or external events concerning an object, such as a person, a topic, an event, or an item, resulting from thought processing, e.g., “my family thinks it’s a good idea for me to continue my education overseas, though they’ll miss me.” Emotions are the most significant aspect of human understanding and have a positive impact on our physical health, jobs, learning, economic, and social behaviors. In addition, whenever a choice must be made, humans seek the opinions of others.

During events such as pandemics, unrest, etc., there is a torrent of emotions. As people faced the unprecedented challenge of COVID-19, their emotional responses became overwhelming

WNLPe-Health 2022: The First Workshop on Context-aware NLP in eHealth, December 15-18, 2022, New Delhi, India

✉ bhardwaj.anjali200594@gmail.com (A. Bhardwaj); abulaish@sau.ac.in (M. Abulaish)

🆔 0009-0008-4063-4661 (A. Bhardwaj); 0000-0003-3387-4743 (M. Abulaish)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

and erratic. Since November 2019, the global community has been dealing with this issue, which has been disastrous for humanity. And due to social media, individuals were suddenly able to share their experiences. In order to be together safely in the future, one had to be apart in the interim.

The classification of emotions from textual data is more complex than sentiment classification. Although emotions and sentiments are synonymous, in sentiment analysis they represent two distinct concepts [1]. The AI hot topic of emotion and sentiment classification has numerous applications in the current technological era. Emotion analysis is used in a variety of AI-based applications, including human-machine interface, cognitive psychology, intelligent devices, automated identification, etc., which have become the gold standard. Thus, research on this topic enables the development of empathic machines [2].

Numerous social media platforms, including Instagram, Facebook, etc., have become indispensable for communicating with friends and family. In online social media, emotions are typically expressed through emoticons and texts that are unstructured, informal, and massive in size. Due to its immense size and unstructured nature, it is difficult to extract meaningful emotional information from such a repository of data. Moreover, different types of emotions, such as *fear*, *anger*, *sadness*, *disgust*, *joy*, and *surprise*, are not mutually exclusive; rather, they are interconnected, such that one emotion causes/triggers other emotion(s) [3]. Therefore, a document can be tagged with multiple emotions, making emotion detection a multi-label classification problem requiring deep learning. Multiple mutually non-exclusive classes or labels are predicted by deep learning algorithms [4]. However, it requires an enormous quantity of labeled data. In order to train and evaluate algorithms for multi-label classification, we intend to curate and label a large dataset.

Our contributions are summarized as follows:

- Curation of a large-scale multi-labeled dataset, MINDS (**M**ulti-label emotIo**N** and senti-ment classification **D**ata**S**et), of textual data containing 227, 229 instances. The tweets were collected using the top-six trending hashtags – #Omicron, #OmicronVariant, #OmicronVariant, #OmicronInIndia, #OmicronVirus, and #Omicronvirus india from December 17, 2021 to February 4, 2022 using Twitter’s Tweepy API.
- Classification of each tweet into the most appropriate emotion categories (annotation labels), which are amongst the universal emotions of Ekman [5], namely *joy*, *sadness*, *fear*, *anger*, *surprise*, and *disgust*, except *surprise*. The *surprise* emotion is dropped due to its ambiguous appearance, indicating both *positive* and *negative* sentiments. The sentiment categories comprise *positive*, *negative* or *neutral*, with numeric scores.
- Using a multi-model setup, namely IBM Watson NLU API, Komprehend Text Analysis API, and Text2emotion Python package to annotate the dataset to quantify all aspects of model uncertainty.
- Based on the SemEval-2018 E-c subtask dataset, determining a classification threshold. After determining the threshold value, the micro-average and macro-average f1-score for each emotion were analyzed. We obtained a Jaccard index of 0.52 and micro-average and macro-average f1-score values of 63% and 61%, respectively.

The remaining sections are organized as follows. The section 2 provides a brief overview of the related datasets for emotion and sentiment classification on textual data based on multi-label

classification. Section 3 provides background information on our work. Section 4 provides the functional steps involved in the creation of our dataset. Finally, section 5 concludes the paper.

2. Related Datasets

Multi-label emotion and sentiment classification have increasingly become an active research topics. Many researchers investigated multi-label emotion classification in textual data [3, 6, 7, 4]. The authors used multiple *Convolution Neural Network (CNN)* along with self-attention and performed multi-label emotion classification on *Twitter* data [3]. Similarly, the authors [7] used transfer learning to enhance the multi-label emotion classification performance on *Twitter* data.

Mostly datasets are hand-annotated, some of them are SemEval-2018 Task 1: Affect in Tweets (AIT) [8], GoEmotions [9], EMOBANK [10], and Emotion Intensities (EmoInt) [11]. EmoInt dataset [11] was created for detecting the emotional intensities of four emotions in tweets.

SemEval-2018 [8] is a multilingual dataset used to train and test supervised machine learning algorithms. This dataset contains 10,690 instances, annotated based on 11 emotions categories. The shared task evaluates automatic systems for E-c (multi-label Emotion-classification), EI-oc (Emotion Intensity-ordinal classification), EI-reg (Emotion Intensity-regression), V-oc (Valence-ordinal classification), and V-reg (Valence-regression) detection in three languages - English, Spanish, and Arabic. Moreover, the authors detected valence, an ordinal class of intensity of emotion (slightly *sad*, very *angry*, etc.), and detected an ordinal class of valence (or sentiment).

In GoEmotions [9] consists of 58k instances collected from English Reddit comments and categorized into 27 emotion categories and neutral classes. The authors used principal preserved component analysis and conducted transfer learning experiments with existing benchmarks to show that their dataset generalizes well to other domains. EMOBANK [10] is another dataset containing above 10,000 sentences labeled according to the emotion representation model of VAD. The authors collected data from various sources, including essays, blogs, fiction, travel guides, letters, newspapers, and news headlines of readers and writers. In addition, a subset of the concerned dataset was categorically labeled using Ekman's emotion model, making it desirable for dual representational designs.

3. Background

This section briefly discusses social media platforms, mainly *Twitter* and their content. Moreover, we discuss well-known APIs, which integrate the processing of textual data to deliver the evoked emotions and sentiments.

3.1. Social Media

Social media platforms have become modern communication tool and have a large user base worldwide. Among these platforms, *Twitter* is the most authentic, admired, and extensively used site by users to share their information, thoughts, ideas, and opinions/emotions simultaneously regarding social events, products, services, and political and marketing campaigns. Due to

constant updates on the repository of opinions, banter, facts, and other minutiae, Twitter has garnered much interest from decision-makers, business leaders, and politicians. This is because of the inherent desires of knowing people's perspectives and opinions regarding specific topics. Therefore, we choose Twitter data to analyze the emotions expressed in tweets. Collecting data using API is the most popular and recommended practice. Almost every social media service provider has an API that assists users with several libraries or packages in various data-extracting activities.

3.2. Well-known APIs

Over the last few years, APIs (Application Programming Interfaces) have increased, breaking the barriers of using a third-party text analytic functionality rather than building their model. APIs play a significant role in today's digital age as they encourage innovation, offer flexible experiences, save cost, and make easy availability. There are dozens of APIs which are essential for the digital transformation, creation, and development of innovative models. The majority of datasets are annotated manually and tend to be small. Moreover, there is no API-based multi-label emotion and sentiment classification dataset is available. So, we annotated our dataset MINDS using a multi-model setup (i.e., APIs and python package) which are discussed below:

3.2.1. IBM Watson NLU

IBM developed Watson NLU (Natural Language Understanding)¹ API that analyzes data with the help of text analytics for extracting categories, classification, concepts, keywords, relations, entities, emotion, sentiment, semantic roles, and syntax. It uses deep learning to extract the meanings and metadata from textual data that is unstructured. Moreover, NLU supports 25 languages depending on whose features one analyzes. Sentiment features analyze the sentiments towards specific phrases in target and also the overall document's sentiment. While, emotion features analyze emotion presented by specific target phrases or the document itself.

3.2.2. Komprehend Text Analysis

This is the most comprehensive document classification and NLP API² for software developers. This API trained their NLP models on more than a billion documents and provided state-of-the-art performance on most common use cases of NLP, such as sentiment analysis and emotion detection. Moreover, this API supports 15 languages. The main advantages of this API is that it works on diverse data, is accurate, supports flexible deployment, and maintains privacy. This API used *Long Short-Term Memory (LSTM)* algorithms which divide text blob sentiment's into *positive* and *negative*. *LSTMs* represents sentences as a series of context-based forget-remember decisions. It was trained differently to handle informal and formal language based on social media and news data. Moreover, this algorithm is trained using various specific datasets for

¹<https://cloud.ibm.com/apidocs/natural-language-understanding>

²<https://komprehend.io/>

different clients. Sometimes the sentiment classes-*positive*, *negative*, and *neutral* are not enough to understand the aspects associated with the underlying tone of any sentence. As a result, the emotion analysis classifier of this API is trained on the proprietary dataset. It can determine whether the emotion conveyed through the text is *sadness*, *happiness*, *fear*, *anger*, *bored*, or *excitement*. Deep learning-based algorithms were used to capture features from the text data. These features were employed to categorize the emotions conveyed by the data. The classifier was trained using *CNNs* on a tagged dataset.

3.2.3. Text2emotion

This is the python package³ that extracts the emotions from the text, and it is compatible with five emotion categories: *sad*, *happy*, *angry*, *fear*, and *surprise*. Details of the package are as follows:

- Text pre-processing: The primary goal is to promote data cleaning while making the content more suitable for emotion analysis. The unwanted textual part is removed from the text; NLP techniques were used to identify the well-pre-processed text.
- Emotion investigation: Pre-processed text are analyzed, and appropriate words are found that express emotions or feelings and category of emotions. The count of emotions relevant to the words are stored.
- Emotion analysis: The output of the text is in the form of a dictionary where the emotion categories and scores are represented as keys and values, respectively. A larger score of a particular emotion category, indicates the text belonging to that category.

4. Dataset Creation

This section gives the functional steps of our dataset creation and analysis, including *data extraction*, *data preprocessing*, *dataset annotation*, *evaluations*, and *observations* for emotion and sentiment classification. Fig. 1 shows the framework of our dataset annotation approach.

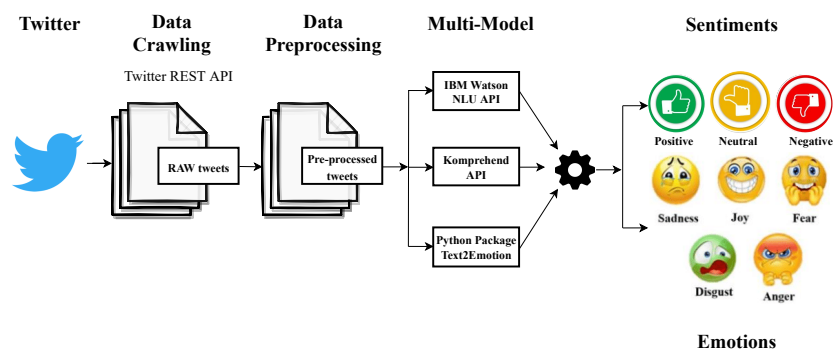


Figure 1: A framework of our dataset annotation approach.

³<https://pypi.org/project/Text2emotion/>

Table 1

Details of the curated corpus of tweets

Hashtag	# of instances
#Omicron	194,942
#OmicronVariant	28,315
#OmicronVariant	8,079
#OmicronInIndia	5,562
#OmicronVirus	4,161
#Omicronvirus india	360
Total tweets	241,419

4.1. Data Extraction

We have created a data crawler utilizing Python 3.6, and Twitter Tweepy-4.6.0 API to collect COVID-19 related posts, mainly those of Omicron (sub-variant of the COVID-19 virus responsible for the third wave of infections) in the English language and stored them in a local repository. To find valuable insights from public reactions and shared posts on Twitter and model public emotions, we collected tweets from December 17, 2021 to February 4, 2022. The tweets are collected using six top trending hashtags such as #Omicron, #OmicronVariant, #OmicronVariant, #OmicronInIndia, #OmicronVirus, and #Omicronvirus india. Table 1 shows the details of the Twitter Omicron corpus containing 2,41,419 instances of raw tweets when the Omicron wave had just begun. Through Twitter API, we obtained various tweet-related information, such as *tweet id*, *text*, *author screen name*, *author id*, *created at*, *source*, *user verified*, *count*, *language*, *favorite count*, *username*, *user id*, *location*, etc.

4.2. Data Pre-processing

Pre-processing is an essential step before analyzing any dataset to deal with the problem of outliers and noises for a proper representation of data. Corpora curated from Twitter as the one presented in Table 1 are often noisy and contain unwanted or irrelevant constituents; eliminating such kind of undesirable information is vital for better performance and efficiency of the system. Since tweets are generated in an uncontrolled manner which is generally unstructured and informal, basic pre-processing steps are applied to transform input data into the ready-to-analyze format. The following steps were performed as part of the pre-processing stage:

1. The uppercase letters were converted into lower-case letters to normalize the input text data.
2. HTML tags, URLs, extra white spaces, hexa-characters (UTF-8), double quotes, and duplicate tweets were removed to promote ease of further processing.
3. The stop-words, punctuation, and emoticons are not removed because they play a significant role in understanding the context of tweets.

Table 2

Sample tweets from our proposed dataset

Tweets	
t_1	not enough people through the doors to cover wages, much less rent and food costs and our #omicron wave hasnt even hit yet.
t_2	#omicron is not mildno one against the businessmen/women how bravely the government stands against them? or they reciprocally the same entity to another? “a” is not “all the time, sometime” could be “a” in one point, and “a” acted on behalf of.
t_3	so proud of the scientists and ex-colleagues who continue to keep norwich safe i miss working with you all - you’re so fantastic at what you do #norwich #omicron

4.3. Dataset Annotation

As presented in Table 1, 241, 419 tweets were scraped from Twitter, after which pre-processing resulted in 227, 229 instances of filtered tweets. Due to highly emotion-rich data from Twitter, a multi-model setup is employed to label the pre-processed tweets and analyze the different user emotions within them. The main advantage of such a setup is to diminish the effect of model errors along with bias that results from each individual model. Additionally, a multi-model setup promotes reliability, consistency, and better labeling by performing automatic text annotation to annotate large quantities of data in less time.

Each independent API predicts scores respectively to identify the evoked emotions and sentiments. Based on these scores, each tweet is classified into the most appropriate emotion categories, which are amongst the universal emotions of Ekman [5], namely *joy*, *sadness*, *anger*, *fear*, *disgust*, and *surprise*, except *surprise*. The *surprise* emotion is dropped due to its ambiguous appearance, indicating both *positive* and *negative* sentiments. At the same time, sentiment categories comprise *positive*, *negative* or *neutral*, with scores. Watson API, Komprehend API, and the Text2emotion python package are employed in the multi-model setup for annotation. The required scores are obtained in the following manner:

- Watson predicts the sentiment label viz. *positive*, *negative*, or *neutral*, and those of emotions, namely *anger*, *fear*, *joy*, *disgust*, and *sadness*, with their corresponding scores for both, the sentiment (ranging from -1 to 1) and the emotion (ranging from 0 to 1) as shown in Table 3 (based on sample tweets given in Table 2).
- The Komprehend labels data amongst the same three sentiments, i.e., *positive*, *negative*, or *neutral* and their emotions amongst *happy*, *sad*, *angry*, *fearful*, *excited*, or *bored*, scores for both of which range from 0 to 1.
- Text2emotion, on the other hand, scores the input only for their emotions amongst the categories *happy*, *anger*, *sad*, *surprise*, and *fear*, scores of which range from 0 to 1.

One difference between the score assignment of Watson and Komprehend is that the former presents a single score labeling the sentiment of each tweet, whereas Komprehend returns individual scores corresponding to each sentiment for the input data. However, the ultimate labels being the same for both, those from Watson are considered as the final value for sentiment classification. Note that the four emotions *sadness*, *joy*, *fear*, and *anger* are considered by each of the above APIs, due to which the scores for the emotion *disgust* are considered directly from

Table 3

Sample annotations obtained using the IBM Watson NLU

Tweets	Sentiment		Emotion				
	Label	Score	Sadness	Joy	Fear	Disgust	Anger
t_1	negative	-0.97301	0.24589	0.093889	0.210093	0.100609	0.186584
t_2	neutral	0	0.309616	0.080319	0.047491	0.094984	0.0778
t_3	positive	0.962837	0.083365	0.7824	0.013969	0.007002	0.007283

Watson. Post obtaining the real-valued scores for each tweet using the aforementioned models, these are converted into binary using the method discussed in section 4.3.2. The method of arriving at the threshold for final annotation is discussed below.

4.3.1. Classification Threshold

In order to obtain the classification threshold, SemEval-2018 [8] E-c subtask, a benchmark dataset, is employed. SemEval-2018 contains 10,983 instances split into three subsets: training set, validation set, and testing set, with 6838, 886, and 3259 instances, respectively. SemEval-2018 is labeled using the three APIs mentioned above, which provide real-valued scores for each tweet. These real-valued scores are transformed into binary using a threshold value. This threshold value, which is considered to be the *InterQuartile Range (IQR)*, helps in making the classification for proper inference of tweets while being a statistics robust to outliers with a better representation of the amount of spread in the data than the range.

Thus, the real-valued scores of the aforementioned multi-model setup corresponding to the ground-truth are separated for both the classes (i.e., 0 and 1), for which the *IQR* is calculated separately. The *IQR* corresponding to the ground-truth value labeled 1 opted for threshold. Finally, these values for the emotions *sadness*, *joy*, *fear*, *disgust*, and *anger* are as follows: 0.26, 0.20, 0.25, 0.06, and 0.12, respectively.

4.3.2. Binary Annotation

The SemEval-2018 dataset is then classified based on the *IQR* for the presence ($>$ threshold) and absence ($<$ threshold) of the particular emotion considered. Post this, majority rating is employed for the final classification of emotions in tweets of the prepared corpus, i.e., two or more models resulting with the same classification (presence/absence) for a tweet classifies the latter as the same (presence/absence of the considered emotion). E.g., if tweet t_1 's score of Watson is 0, Komprehend is 1, and Text2emotion is 0 for emotion *sadness* as shown in Table 4, then the final label is 0, i.e., emotion is not present shown in Table 5. In the vast majority of cases, a tweet is labeled as 0 among five emotions. The curated Twitter-based corpus is then fully labeled using these classification rules.

4.4. Evaluation Metrics

In multi-label tasks, the results can be partially correct or wrong. To capture the notion of partial correctness, one can use metrics. A *Jaccard index*, equivalent to multi-label accuracy,

Table 4

Sample binary annotations obtained using the threshold in a multi-model setup

Tweets	IBM Watson NLU					Komprehend Text Analysis				Text2emotion			
	Sadness	Joy	Fear	Disgust	Anger	Sadness	Joy	Fear	Anger	Sadness	Joy	Fear	Anger
t_1	0	0	0	1	1	1	0	0	1	0	0	1	1
t_2	0	0	0	1	0	0	0	0	1	0	0	1	0
t_3	0	1	0	0	0	0	1	0	0	0	1	1	0

Table 5

Annotations obtained for the tweets given in Table 2

Tweets	Sentiment		Emotion				
	Label	Score	Sadness	Joy	Fear	Disgust	Anger
t_1	negative	-0.97301	0	0	0	1	1
t_2	neutral	0	0	0	0	1	0
t_3	positive	0.962837	0	1	0	0	0

which was denoted as JI . It can be defined as the intersection size divided by the size of the union of the ground-truth and predicted labels, is formulated as:

$$JI = \frac{1}{|T|} * \sum_{i \in T} \frac{GT_i \cap P_i}{GT_i \cup P_i} \quad (1)$$

where GT_i is the set of the ground-truth labels, for instance, i , P_i is the set of the predicted labels for instances i , and T is the set of tweets. Additionally, label-based metrics have also been utilized for performance evaluation. Mathematically, *micro-average* is ascertained by aggregating *micro-average precision* (p_{micro}) and *micro-average recall* (r_{micro}), which are defined using equations 2 and 3, respectively.

$$p_{micro} = \frac{\sum_{l_i \in d} TP(l_i)}{\sum_{l_i \in d} TP(l_i) + FP(l_i)} \quad (2)$$

$$r_{micro} = \frac{\sum_{l_i \in d} TP(l_i)}{\sum_{l_i \in d} TP(l_i) + FN(l_i)} \quad (3)$$

Similarly, *macro-average* is ascertained by aggregating *macro-average precision* (p_{macro}) and *macro-average recall* (r_{macro}), which are defined using equations 4 and 5, respectively.

$$p_{macro} = \frac{1}{d} * \sum_{l_i} \frac{TP(l_i)}{TP(l_i) + FP(l_i)} \quad (4)$$

$$r_{macro} = \frac{1}{d} * \sum_{l_i} \frac{TP(l_i)}{TP(l_i) + FN(l_i)} \quad (5)$$

where l_i is the predicted label and d is the number of labels.

Table 6

Statistical analysis of the proposed dataset

Hashtags	# sentiment-classified tweets			# emotion-classified tweets				
	Positive	Negative	Neutral	Sadness	Joy	Fear	Disgust	Anger
#Omicron	40,403	78,716	62,686	22,249	9,796	30,156	40,741	28,320
#OmicronVariant	5,479	13,393	8,433	3,000	1,638	4,028	7,309	4,544
#OmicronVariet	1,604	3,861	2,578	904	503	1,186	2,277	1,341
#OmicronInIndia	1,074	2,168	2,314	499	244	672	1203	719
#OmicronVirus	771	1,820	1,570	454	234	651	1,121	697
#Omicronvirus india	44	124	191	43	6	44	16	52

Formally, $f1$ -score is defined as the harmonic mean of *precision* and *recall*, and its value spans between 0 to 1. Similarly, *micro* and *macro* $f1$ -scores are computed using equation 6. The *micro* $f1$ -score gives equal weight to each testing instance, whereas *macro* $f1$ -score gives equal weight to each emotion. A higher value of $f1$ -score indicates better multi-label classification results.

$$f1_{macro/micro} = 2 * \frac{p_{macro/micro} * r_{macro/micro}}{p_{macro/micro} + r_{macro/micro}} \quad (6)$$

4.5. Observations

After calculating the threshold using a benchmark dataset, each emotion’s *micro-average* and *macro-average recall*, *precision*, and *f1-score* were evaluated to validate the thresholds. A *Jaccard index*, equivalent to multi-label accuracy of 0.52 and *micro-average* and *macro-average f1-score* as 63% and 61% were obtained respectively.

Table 6 shows the statistical details of our proposed dataset MINDS. The table indicates the number of sentiment and emotion-classified tweets of six hashtags. Among those, #Omicron contains more instances compared to other hashtags. Most of the hashtags have a large number of *negative* tweets, except #OmicronInIndia and #Omicronvirus india, which have large *neutral* tweets. At the same time, *disgust* emotion has large instances, except #Omicronvirus india hashtag, which has 52 instances in emotion *anger*.

5. Conclusion

In this paper, we have presented the curation of a large-scale, multi-labeled emotion and sentiment classification dataset, MINDS, that contains COVID-19-related textual data. We used a multi-model setup to classify the data into the most relevant emotion categories (annotation labels), which are among the universal emotions of Ekman and sentiment categories. The multi-model configuration included the models, such as the IBM Watson NLU API, the Komprehend API, and the Text2emotion Python package for automatic data annotation. This resolved the critical issue of time-consuming and labor-intensive dataset labeling. A benchmark dataset (SemEval-2018, subtask E-c) was used to determine the classification threshold. We obtained

the Jaccard index as 0.52 and the micro-average and macro-average f1-score as 63% and 61%, respectively. The MINDS dataset is publicly available at <http://www.abulaish.com/lds/dataset> for research purposes. We are currently expanding the dataset to include additional emotions, such as *love*, *amusement*, *desire*, *admiration*, and *grief*.

References

- [1] A. Yadollahi, A. G. Shahraki, O. R. Zaiane, Current state of text sentiment analysis from opinion to emotion mining, *ACM Computing Surveys (CSUR)* 50 (2017) 1–33.
- [2] F. A. Acheampong, C. Wenyu, H. Nunoo-Mensah, Text-based emotion detection: Advances, challenges, and opportunities, *Engineering Reports* 2 (2020) e12189.
- [3] Y. Kim, H. Lee, K. Jung, Attnconvnet at semeval-2018 task 1: Attention-based convolutional neural networks for multi-label emotion classification, *arXiv preprint arXiv:1804.00831* (2018).
- [4] M. Jabreel, A. Moreno, A deep learning-based approach for multi-label emotion classification in tweets, *Applied Sciences* 9 (2019) 1123.
- [5] P. Ekman, Basic emotions, *Handbook of cognition and emotion* 98 (1999) 16.
- [6] W. Ying, R. Xiang, Q. Lu, Improving multi-label emotion classification by integrating both general and domain knowledge, in: *Proceedings of the 5th Workshop on Noisy User-Generated Text (W-NUT 2019)*, Association for Computational Linguistics, Hong Kong, China, 2019, pp. 316–321.
- [7] J. Yu, L. Marujo, J. Jiang, P. Karuturi, W. Brendel, Improving multi-label emotion classification via sentiment classification with dual attention transfer network, in: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 1097–1102.
- [8] S. Mohammad, F. Bravo-Marquez, M. Salameh, S. Kiritchenko, SemEval-2018 task 1: Affect in tweets, in: *Proceedings of the 12th international workshop on semantic evaluation*, Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 1–17.
- [9] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, S. Ravi, GoEmotions: A dataset of fine-grained emotions, *arXiv preprint arXiv:2005.00547* (2020).
- [10] S. Buechel, U. Hahn, Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis, 2022.
- [11] S. M. Mohammad, F. Bravo-Marquez, WASSA-2017 shared task on emotion intensity, *arXiv preprint arXiv:1708.03700* (2017).