

“© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Detecting Community Depression Dynamics Due to COVID-19 Pandemic in Australia

Jianlong Zhou^{ID*}, Senior Member, IEEE, Hamad Zogan^{ID}, Shuiqiao Yang^{ID}, Shoaib Jameel^{ID},

Guandong Xu^{ID*}, Fang Chen^{ID}

Abstract—The recent COVID-19 pandemic has caused unprecedented impact across the globe. Many people are going through increased mental health issues, such as depression, stress, worry, fear, disgust, sadness, and anxiety, which have become one of the major public health concerns during this severe health crisis. Depression can cause serious emotional, behavioural and physical health problems with significant consequences, both personal and social costs included. This paper studies community depression dynamics due to COVID-19 pandemic through user-generated content on Twitter. A new approach based on multi-modal features from tweets and Term Frequency-Inverse Document Frequency (TF-IDF) is proposed to build depression classification models. Multi-modal features capture depression cues from emotion, topic and domain-specific perspectives. We study the problem using recently scraped tweets from Twitter users emanating from the state of New South Wales in Australia. Our novel classification model is capable of extracting depression polarities which may be affected by COVID-19 and related events during the COVID-19 period. The results found that people became more depressed after the outbreak of COVID-19. The measures implemented by the government such as the state lockdown also increased depression levels. Further analysis in the Local Government Area (LGA) level found that the community depression level was different across different LGAs.

Index Terms—Depression, Multi-modal features, COVID-19, Twitter, Australia.

I. INTRODUCTION

THE outbreak of the novel Coronavirus Infectious Disease 2019 (COVID-19) has caused an unprecedented impact on people’s daily lives around the world. People’s lives are at risk because the virus can easily spread from person to person either by coming in close contact with the infected person or sometimes may even spread through community transmission¹, which then becomes extremely challenging to contain. The infection has now rapidly spread across the world and there have been more than 10.3 million confirmed cases and more than 505,000 people have died because of the infection until 30 June 2020². Almost every country in the world is battling against COVID-19 to prevent it from

spreading as much as possible. While some countries such as New Zealand has been very successful in containing the spread, others such as Brazil and India have not. As a result, this outbreak has caused immense distress among individuals either through infection or through increased mental health issues, such as depression, stress, worry, fear, disgust, sadness, anxiety, and perceived stigmatisation [1], [2]. These mental health issues can even occur in people, not at high risk of getting infected. There could be even several people who are exposed to the virus may be unfamiliar with it as they may not follow the news, or are completely disconnected with the general population. [1].

Consider depression as an example, which is the most common mental health issue among other mental health issues according to the World Health Organisation (WHO), with more than 264 million people suffering from the depression worldwide [3]. Australia is one of the top countries where mental health disorders have high proportions over the total disease burden (see Fig. 1). A person can experience several complications as a result of depression. Complications linked to mental health especially depression include: unhappiness and decreased enjoyment of life, family conflicts, relationship difficulties, social isolation, problems with tobacco, alcohol and other drugs, self-harm and harm to others (such as suicide or homicide), weakened immune system. Furthermore, the consequence of depression goes beyond functioning and quality of life and extends to somatic health. Depression has been shown to subsequently increase the risk of, for example, cardiovascular, stroke, diabetes and obesity morbidity. However, the past epidemics can suggest some cues of what to look out for after COVID-19 in the next few months and years. For example, when patients with SARS and MERS were assessed a few months later, 14.9% had depression and 14.8% had an anxiety disorder [2].

Meanwhile, to reduce the risk of the virus spreading among people and communities, different countries have taken strict measures such as locking down the whole city and practising rigorous social-distancing among people. For example, countries such as China, Italy, Spain, and Australia are fighting the COVID-19 pandemic through nation-wide lockdown or by cordoning off the areas that were suspected of having risks of community spread throughout the pandemic, expecting to “flatten the curve”. However, the long-term social activity restriction policies adopted during the pandemic period may further amplify the mental health issues of people. Therefore,

J. Zhou, S. Yang, and F. Chen are with Data Science Institute, University of Technology Sydney, Australia, e-mails: Jianlong.Zhou@uts.edu.au, Shuiqiao.Yang@uts.edu.au, Fang.Chen@uts.edu.au

H. Zogan and G. Xu are with Advanced Analytics Institute, University of Technology Sydney, Australia, e-mails: Hamad.A.Zogan@student.uts.edu.au, Guandong.Xu@uts.edu.au

S. Jameel is with University of Essex, UK, email: Shoaib.Jameel@essex.ac.uk

*Corresponding authors

¹<https://www.who.int/publications/i/item/preparing-for-large-scale-community-transmission-of-covid-19>

²<https://coronavirus.jhu.edu/>

³<https://ourworldindata.org/>

Mental health disorders as a share of total disease burden, 2016

Mental health and neurodevelopment disorders (not including alcohol and drug use disorders) as a share of total disease burden. Disease burden is measured in DALYs (Disability-Adjusted Life Years). DALYs measure total burden of disease - both from years of life lost and years lived with a disability. One DALY equals one lost year of healthy life.

Our World
in Data

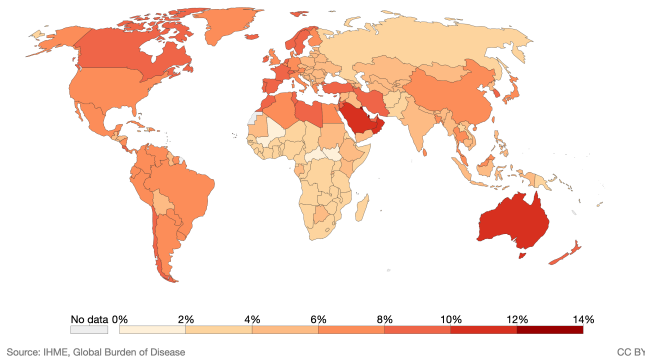


Fig. 1: The world mental health disorders in 2016³.

it is important to examine people's mental health states as a result of COVID-19 and related policies, which can help governments and related agencies to take appropriate measures more objectively if necessary.

On the other hand, we have witnessed increased usage of online social media such as Twitter during the lockdown⁴. For instance, 40% consumers have spent longer than usual on messaging services and social media during the lockdown. It is mainly because people are eager to publicly express their feelings online given an unprecedented time that they are going through both physically and emotionally. The social media platforms represent a relatively real-time large-scale snapshot of activities, thoughts, and feelings of people's daily lives and thereby reflect their emotional well-being. Every tweet represents a signal of the users' state of mind and state of being at that moment. Aggregation of such digital traces may make it possible to monitor health behaviours at a large-scale, which has become a new growing area of interest in public health and health care research [4].

Since social media is social by its nature, and social patterns can be consequently found in Twitter feeds, for instance, thereby revealing key aspects of mental and emotional disorders. As a result, Twitter recently has been increasingly used as a viable approach to detect mental disorders of depression in different regions of the world [5]. For example, the research found that the depressed users were less active in posting tweets, doing it more frequently between 23:00 and 6:00. The use of vocabularies could also be an indicator of depression in Twitter, for example, it was found that the use of verbs was more common by depressed users, and the first-person singular pronoun was by far the most used by depressed users [5]. Hence, many research work has been done to extract features like user's social activity behaviours, user profiles, texts from their social media posts for depression detection using machine learning techniques [6], [7], [8], [9], [10]. For example, De Choudhury [6] et al., proposed to predict depression for social media users based on Twitter using support vector machine

(SVM) for prediction based on manually labelled training data.

The most recent work using Twitter to analyse mental health issues due to COVID-19 are [4]. These work focus more on public sentiment analysis. Furthermore, little work such as Li et al., [11] classify each tweet into the emotions of anger, anticipation, disgust, fear, joy, sadness, surprise and trust. The two emotions of sadness and fear are more related to severe negative sentiments like depression due to COVID-19. However, little work is done to detect depression dynamics at the state level or even more granular level such as suburb level. Such granular level analysis of depression dynamics not only can help authorities such as governmental departments to take corresponding actions more objectively in specific regions if necessary but also allows users to perceive the dynamics of depression over the time to learn the effectiveness of policies implemented by the government or negative effects of any big events. The answers to the questions that we wish to find are: 1) How people's depression is affected by COVID-19 in the time dimension in the state level?, 2) Can we detect the effects of policies/measures implemented by the government during the pandemic on depression?, 3) Can we detect the effects of big events on depression during the pandemic?, 4) How effective is the model in detecting people's depression dynamics?

This paper aims to examine community depression dynamics due to COVID-19 pandemic in Australia. A new approach based on multi-modal features from tweets and term frequency-inverse document frequency (TF-IDF) is proposed to build a depression classification model. Multi-modal features aim to capture depression cues from emotion, topic and domain-specific perspectives. TF-IDF is simple, scalable, and has proven to be very effective to model a wide-range of lexical datasets including large and small datasets. In contrast, recent computationally demanding frameworks such as deep learning, which are usually generated rely upon large datasets because that helps give them faithful co-occurrence statistics might not be obtained from sparse and short texts from user-generated content such as tweets. Our approach uses the TF-IDF model can generalise well in various situations where both small and large datasets can be used leading to a reliable and scalable model. After building the model for depression classification, Twitter data in the state of New South Wales in Australia are then collected and input to our depression classification model to extract depression polarities which may be affected by COVID-19 and related events during the COVID-19 period. The contributions of this paper primarily include: 1) Novel multi-modal features including emotion, topic and domain-specific features are extracted to describe depression in a comprehensive approach, 2) A faithful depression classification model based on TF-IDF, which is simple and scalable in generalisation, is proposed to detect depressions in short texts such as tweets, 3) Instead of the depression examination of the whole country, a fine-grained analysis of depression in local government areas of a state in Australia is investigated, 4) The links between the community depression and measures implemented by the government during the COVID-19 pandemic are examined.

To the best of our knowledge, this study is the first work

⁴<https://www.statista.com/statistics/1106498/home-media-consumption-coronavirus-worldwide-by-country/>

to investigate the community depression dynamics across the COVID-19 pandemic, as well as to conduct the fine-grained analysis of depression and demonstrate the links of depression dynamics with measures implemented by the government and the COVID-19 itself during the pandemic.

II. RELATED WORK

In this section, we review the related work for depression detection. We also highlight how our work differs from these existing approaches.

A. Machine learning based depression detection

Social media has long been used as the data source for depression detection due to the largely available user-generated text data [12], [13]. The shared text data and the social behaviour of the social network users are assumed to contain clues for identifying depressed users. To find the depression pattern for social media users, many works have been done to adopt traditional machine learning models such as Support Vector Machine (SVM) and J48 for depression classification and detection based on different feature engineering techniques. For example, Wang et al. [14] have proposed a binary depression detection model for Chinese Sina micro-blogs to classify the posts as depressed or non-depressed. Based on the features extracted from the content of the micro-blog, the users' online interactions with others and the user behaviours, they have trained J48 tree, Bayes network, and rule-based decision table as classifiers for depression detection.

De Choudhury [6] et al. have investigated to predict depression for social media users based on Twitter and found that social media contain meaningful indicators for predicting the onset of depressions among individual users. To get the ground truth of users' suffered depression history, De Choudhury et al. adopted the crowdsourcing to collect Twitter users who have been diagnosed with clinical (Major Depressive Disorder) MDD based on the CES-D2 (Center for Epidemiologic Studies Depression Scale) screening test. Then, to link the depression symptoms with the social media data, they extracted several measures such as user engagement and emotion, egocentric social graph, linguistic style, depressive language user, and the mentions of antidepressant medications from users' social media history for one year. Finally, they trained SVM as the depression classifier base on the ground truth and the extracted features for the tested Twitter users with prediction accuracy around 70%. Similarly, Tsugawa et al. [7] have investigated in recognizing depression from the Twitter activities. To get the ground truth data of users' depression degree record, they chose the web-based questionnaire for facts collection. Then, similar features such as topic features extracted using topic modelling like Latent Dirichlet Allocation (LD), polarities of words and tweet frequency are extracted from the Twitter users' activity histories for training an SVM classifier.

Later, Yang et al. [8] have proposed to analyse the spatial patterns of depressed Twitter users based on Geographic Information Systems (GIS) technologies. They firstly adopted Non-negative Matrix Factorization (NMF) to identify the depressed tweets from online users. Then, they exploited the geo-tagged

information as the indicator of users' geographical location and analyzed the spatial patterns of depressed users. Shen et al. [10] have made efforts to explore comprehensive multi-modal features for depression detection. They adopted six groups of discriminant depression-oriented features extracted from users' social network, profile, visual content, tweets' emotions, tweets' topics and domain-specific knowledge are used as representations for Twitter users. Then, a dictionary learning model is used to learn from the features and capture the jointly cross-modality relatedness for a fused sparse representation and train a classifier to predict the depressed users.

Different from these work in feature extraction which may be either impractical like using crowdsourcing to manually label each user or too customized like extracting different kinds of social behaviours of users, we propose to combine the robust and simple text feature representation method: Term Frequency-Inverse Document Frequency (TF-IDF) with other multi-modal features such as topics and emotions to represent the text data.

B. Deep learning based depression detection

More recently, the rapidly developed deep learning techniques have also been used for depression detection. For example, Shen et al. [15] have proposed a cross-domain deep neural network with feature transformation and combination strategy for transfer learning of depressive features from different domains. They have argued that the two major challenges regarding cross-domain learning are defined as isomerism and divergence and proposed DNN-FATC which includes Feature Normalization & Alignment (FNA) and Divergent Feature Conversion (DFC) to better transfer learning. Orabi et al. [9] have proposed to explore the word embedding techniques where they used pre-trained Skip-Gram (SG) and Continuous Bag-of-Words (CBOW) models currently implemented in the word2vec [16] package for better textual feature extraction to train a neural network-based classifier. We have also seen researchers using deep reinforcement learning to depression detection on Twitter. For instance, Gui et al. [17] have proposed a cooperative multi-agent model to jointly learn the textual and visual information for accurate depression inference. In their proposed method, the text feature is extracted using a Gated Recurrent Unit (GRU) and the visual feature is extracted using Convolutional Neural Networks (CNN). The selection for useful features from GRU and CNN is designed as policy gradient agents and trained by a centralized critic that implements difference rewards.

Even though deep learning has become the dominated method for many classifications or prediction tasks, it cannot guarantee that deep learning is feasible on any tasks. For example, [18] compared a bunch of methods including different conventional methods and deep learning methods for information extraction tasks from a text corpus. We are not surprised to find that deep learning did not outperform the conventional methods as expected. Furthermore, deep learning techniques are usually dependent on large datasets with faithful co-occurrence statistics which might not be obtained from sparse and short texts especially online user-generated content

such as tweets which is characterised by relatively limited labelled text in the depression study. While TF-IDF is simple and scalable, it has proven to model a wide-range of datasets including large and small datasets. Therefore, in our study, we propose to adopt the traditional classification methods with TF-IDF for depression detection to obtain a faithful model with a strong generalisation ability in various situations of small and large datasets. Most importantly, even non-experts such as those from the government and NGOs could easily understand the intricacies of the model and apply our model on their data to detect community dynamics in their region and make further decisions accordingly.

C. Depression detection due to COVID-19

The impact of COVID-19 on people’s mental health has been recently reported in various research. For instance, Galea et al. [19] have pointed out that the mental health consequences due to this pandemic are applicable in both the short and long term. They have appealed to develop ways to intervene with the inevitability of loneliness and its consequence as people are physically and socially isolated. Huang et al. [20] have exploited a web-based cross-sectional survey based on the National Internet Survey on Emotional and Mental Health (NISEMH) to investigate people’s mental health status in China by spreading the questionnaire on Wechat (a popular social media platform in China). To infer the depression symptoms for the anonymous participants, they have adopted a predefined Epidemiology Scale to identify whether participants had depressive symptoms. Similarly, Ni et al. [21] have conducted a survey on the online platform to investigate the mental health of 1577 community-based adults and 214 health professionals during the epidemic and lockdown. The results show that around one-fifth of respondents reported probable anxiety and probable depression. These works are mainly based on questionnaires and pre-defined mental health scale models for inference. In contrast, our proposed work relies on detecting depression from social media data automatically which shows advantages in monitoring a large number of people’s mental health states.

III. DATA

A. Study location

In this study, a case study for analysing depression dynamics during the COVID-19 pandemic in the state of New South Wales (NSW) in Australia is conducted. NSW has a large population of around 8.1 million people according to the census in September 2019 from Australian Bureau of Statistics⁵. The NSW’s capital city Sydney is Australia’s most populated city with a population of over 5.3 million people. The Local Government Areas (LGAs) are the third tier of government in the Australian state (the three tiers are federal, state, and local government).

⁵<https://www.abs.gov.au/>

TABLE I: Summary of the collected Twitter dataset.

Description	Size
Total Twitter users	183,104
Average Twitter user per LGA	1,430.5
Average tweets per LGA	739,900.5
Total tweets	94,707,264

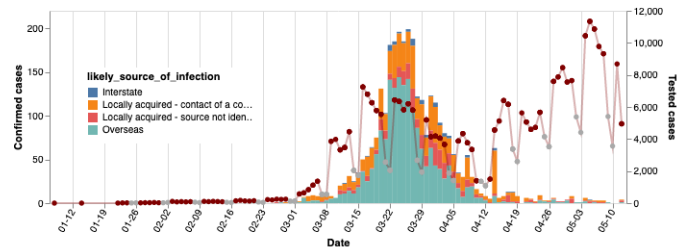


Fig. 2: The tests and confirmed cases of COVID-19 in NSW until 22 May 2020.

B. Data collection

To analyse the dynamics of depression during the COVID-19 pandemic period at a fine-grained level, we collected tweets from Twitter users who live in different LGAs of NSW in Australia. The time span of the collected tweets is from 1 January 2020 to 22 May 2020 which covers dates that the first confirmed case of coronavirus was reported in NSW (25 January 2020) and the first time that the NSW premier announced the relaxing for the lockdown policy (10 March 2020). There are 128 LGAs in NSW. In this study, Twitter data were collected for each LGA separately so that the depression dynamics can be analysed and compared for LGAs. Twitter data were collected through the user timeline crawling API *user_timeline*⁶ in Tweepy which is a python wrapper for the official Twitter API⁷. Table I shows the summary of the collected tweet dataset. In summary, 94,707,264 tweets were collected with averagely 739,901 tweets for each LGA during the study period. Datasets of COVID-19 tests and confirmed cases in NSW during the study period were collected from DATA.NSW⁸.

Fig. 2 shows the overview of the number of tests (polylines with dots) and confirmed cases (bars) of COVID-19 in NSW over the study period. It demonstrates that there usually had test peaks at the beginning of each week and had fewer numbers at the weekend, which well aligns with the people’s living habits in Australia. It shows that the outbreak peak of COVID-19 in NSW was on 26 March 2020 and tests were significantly increased after 13 April 2020. It also shows that most of the confirmed cases were originally related to overseas.

C. Dataset for depression model training

In order to detect depression at the tweet level, we created two labelled datasets for both depressed and non-depressed tweets:

⁶<http://docs.tweepy.org/en/latest/api.html#api-reference>

⁷<https://developer.twitter.com/en/docs/api-reference-index>

⁸<https://data.nsw.gov.au/>

- Positive tweets (depressed): we used a previous work dataset from [10]. Shen et al. [10] published around 300K tweets with 1400 depressed users. In order to have a better performance, we increased the number of positive tweets by crawling additional 600K tweets from the Twitter keyword streaming API. We adopted the same keywords search that selected by [10] where users identified themselves as depressed, and we also used a regular expression to find positive tweets (e.g. I'm depressed AND suicide).
- Negative tweets (non-depressed): In order to balance the negative tweets with the positive tweets, we randomly selected 900K tweets which were not labelled as depressed from the collected tweets.

After the collection of experimental data, features need to be extracted from social media text. However, because of the free-style nature of social media text, it is hard to extract useful features directly from raw text data and apply them to a classifier. The raw text data also affect the efficiency of extracting reliable features, and it makes it difficult to find word similarities and apply semantic analysis. Therefore, raw data must be pre-processed before feature extraction, in order to clean and filter data and ensure the data quality for feature extraction. Pre-processing may also involve other procedures such as text normalization.

Natural Language Processing (NLP) toolkit has been widely used for text pre-processing due to its high-quality processing capabilities such as processing sentimental analysis datasets. Natural Language Processing Toolkit (NLTK) library is considered as one of the most powerful NLP libraries in Python programming. NLTK contains packages that make data processing with human language easily and is used widely by various researchers for text data pre-processing. Therefore, before feeding our data to the model, we used NLTK to remove user mentions, URL links and punctuation from tweets. Furthermore, we removed common words from each tweet such as “the”, “an”, etc.). There are various reasons which have been mentioned in the literature where removing the stop words has had a positive impact on the model’s quantitative performance, for instance, sometimes stop words deteriorate the classifications performance, sometimes they also have a huge impact on the model efficiency because these stop words increase the parameter space of the model, among various other reasons. NLTK has a set of stop words which enable removing them from any text data easily. Finally, we stem tweets using NLTK using the Porter Stemmer.

IV. OUR MODEL

A. Proposed method

In this study, two sets of features are extracted from raw text and used to represent tweet. Since extracting features would be challenging due to the short length of the tweets and single tweet does not provide sufficient word occurrences, we, therefore, combine multi-modal feature with Term Frequency-Inverse Document Frequency (TF-IDF) feature to analyze depressed tweets. Our proposed framework is shown in Fig. 3.

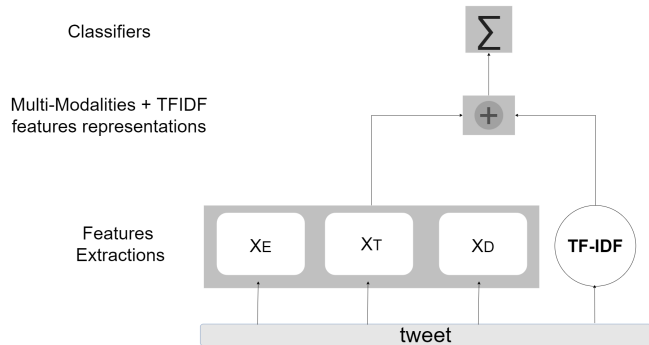


Fig. 3: The proposed framework to detect depressed tweets during the COVID-19

B. Multi-modal features

User behaviours at the tweet level could be modelled by the linguistic features of tweets posted by the user. Inspired by [10], we defined a set of features consisting of three modalities. These three modalities of features are as follows:

- **Emotional features:** The emotion of depressed people is usually different from non-depressed people, which influences their posts on social media. In this work, we studied user positive and negative emoji in each tweet to represent emotional features. Furthermore, users in social media often use a lot of slang and short words, which also convey positive and negative emotions [22]. In this study, positive and negative emotion features are also extracted based on those slang and short words.
- **Topic-level features:** We adopted Latent Dirichlet Allocation (LDA) [23] to extract topic-level features since LDA is the most popular method used in the topic modelling to extract topics from text, which could be considered as a high-level latent structure of content. The idea of LDA is based on the assumption of a mixture of topics forms documents, each of which generates words based on their Dirichlet distribution of probability. Given the scope of the tweet content, we defined 25 latent topics in this study, which topic number is often adopted in other studies. We have also found that this number of topics gives satisfactory results in our experiments. We implemented LDA in Python with Scikit-Learn.
- **Domain specific features:** Diagnostic and Statistical Manual of Mental Disorders 4th Edition (also called DSM-IV) is a manual published by the American Psychiatric Association (APA) that includes almost all currently recognized mental health disorder symptoms [24]. We, therefore, chose the DSM-IV Criteria for Major Depressive Disorder to describe keywords related to the nine depressive symptoms. Pre-trained word2vec (Gensim pre-trained model based on Wikipedia corpus) was used in this study to extend our keywords that are similar to these symptoms. We also extracted “Antidepressant” by creating a complete list of clinically approved prescription antidepressants in the world. These are: 1) Depressed mood, 2) Loss of interest, 3) Weight or appetite change,

4) Sleep disturbance, 5) Psychomotor changes, 6) Fatigue or loss of energy, 7) Feel Worthlessness, 8) Reduced concentration, and 9) Suicidal ideation.

For a given sample of tweet, the multi-modal features are represented as $X_{t1}, X_{t2}, X_{t3}, \dots, X_{tn}$, where $X_{ti} \in \mathbb{R}^d$ is the d -dimensional feature for the i -th modality for each tweet and n is the size of the combined feature space, which is 21 in this study.

C. TF-IDF

We first review the definitions of Term Frequency and Inverse Document Frequency below.

Definition 1. Term Frequency (TF): consider the tweets in Bag-of-Word model (BoW), where each tweet is modeled as a sequence of words without order info. Apparently in BoW scheme, a word in a tweet with occurrence frequency of 10 is more important than a term with that frequency of 1, but not proportional to the frequency. Here we use the log term frequency tf as defined by:

$$tf_{(t,d)} = 1 + \log(1 + tf_{(t,d)}) \quad (1)$$

where $tf_{(t,d)}$ represents occurrence number of the term t in a tweet d .

Definition 2. Inverse Document Frequency (IDF): It uses the frequency of the term in the whole collection for weighting the significance of the term in light of inverse effect. Therefore under this scheme, the IDF value of a rare word is higher, whereas lower for a frequent term, i.e. weighting more on the distinct words. The log IDF which measures the informativeness of a term is defined as:

$$idf_t = \log_{10} \frac{N}{df_t} \quad (2)$$

where N represents the total number of tweets in the tweet corpus, and df_t the number of tweets containing the term t .

Therefore, TF-IDF is calculated by combining TF and IDF as represented in Eq.3.

$$tfidf_t = tf_t * idf_t \quad (3)$$

In order to extract relevant information from each tweet and reduce the amount of redundant data without losing important information, we combined multi-modalities with TF-IDF. TF-IDF is a numerical statistic metric to reflect the importance of a word in a list or corpus, which is widely studied in relevant work. Choudhury et al. [6] applied TF-IDF to words in Wikipedia to remove extremely frequent terms and then used the top words with high TF-IDF. The approach helped them to assess the frequency of uses of depression terms that appear on each user's Twitter posts on a given day. In [25] the authors have used TF-IDF in their model to compare the difference of performance by feeding TF-IDF into five machine learning models. It was found that all of them can achieve very good performance. However, one weakness of BoW is unable to consider the word position in the text, semantics, co-occurrences in different documents. Therefore, TF-IDF is only useful as a lexical level feature.

TABLE II: The performance of tweet depression detection based on multi-modalities only.

Features	Method	Precision	Recall	F1-Score	Accuracy
Multi-Modal	LR	0.842	0.828	0.832	0.833
	LDA	0.843	0.816	0.820	0.824
	GNB	0.873	0.814	0.818	0.825

TABLE III: The performance of tweet depression detection based on TF-IDF only.

Features	Method	Precision	Recall	F1-Score	Accuracy
TF-IDF	LR	0.908	0.896	0.900	0.901
	LDA	0.906	0.893	0.897	0.898
	GNB	0.891	0.873	0.877	0.879

D. Modeling depression in tweets

Two labelled datasets as introduced in the previous section are used to train the depression classification model for tweets. Here we only use English tweets to train the model, and all non-English tweets are excluded. We also exclude any tweets with a length shorter than five words since these tweets could only introduce noise and influence the effectiveness of the model negatively. Three mainstream classification methods are used in this study to compare their performance, namely Logistic Regression (LR), Linear Discriminant Analysis (LDA), and Gaussian Naive Bayes (GNB). We used scikit-learn libraries to import the three classification methods. The classification performance by these three methods were evaluated by 5-fold cross-validation. The experiments are conducted using Python 3.6.3 with 16-cores CPU.

We evaluate the classification models by using measure of Accuracy (ACC.), Recall (Rec.), Macro-averaged Precision (Prec.), and Macro-averaged F1-Measure (F1).

V. CLASSIFICATION EVALUATION RESULTS

We firstly evaluate how well the existing models can detect depressed tweets. After extracting the feature representations of each tweet, three different classification models are trained with the labelled data. We adopted a ratio of 75:25 to split the labelled data into training and testing data. We performed experiments by using TF-IDF, multi-modality, and combined features under three different classification models.

We compare the performance of models using only Multi-Modalities (MM) features with the three different classifiers as shown in Table II. We found that Gaussian Naive Bayes for MM features obtained the highest Precision score compared to other classifiers, and Logistic Regression performs better than the other two classifiers in terms of Recall, F1-Score, and Accuracy.

Furthermore, in order to see how different features impact the classification performance, we use the TF-IDF standalone with three classifiers for comparison and the results are as shown in Table III. It shows that all three classifiers can achieve satisfactory classification results. LR and LDA shared the highest Precision and F1-Score. LR competes with the other two classifiers according to the Recall and Accuracy.

Table IV shows the model performance when we concatenate both MM and TF-IDF features, and we can see that the model has improved the performance further slightly. One

TABLE IV: The performance of tweet depression detection based on Multi-Modalities + TF-IDF.

Features	Method	Precision	Recall	F1-Score	Accuracy
MM+TF-IDF	LR	0.908	0.899	0.902	0.903
	LDA	0.912	0.899	0.903	0.904
	GNB	0.891	0.874	0.878	0.879

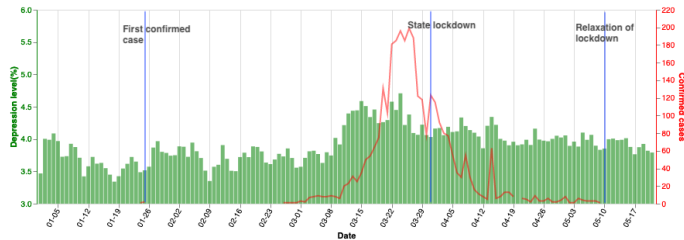


Fig. 4: The community depression dynamics in NSW between 1 January 2020 and 22 May 2020.

conclusion we can draw here is that TF-IDF textual feature can make main satisfactory contribution to detect depression tweets, while other modality can provide additional support. This could be attributed to the lexical importance to depressed-related tweets. Another reason why the overall combination of multi-modal features did not give us a big lift in the results could be because, as mentioned earlier, tweets are highly sparse, poorly phrased, short texts, and noisy content. Therefore, deriving semantic content from tweets, for instance, using the LDA model would always be very challenging to get a huge boost in the results. This is mainly because any statistical model relies on the co-occurrence statistics which might be poor in our case. However, we still see an improvement in the overall recall result, which is important in this case, because we are noticing a reduction in the number of false negative detection. This is mainly because of the interplay between all three features which suggest that these features are important and cannot be ignored.

VI. DETECTING DEPRESSION DUE TO COVID-19

After having testified our classification model, we utilise our approach to detect depressed tweets from different LGAs of NSW, Australia. Since our model deals only with English tweets, we had to exclude tweets in all other languages and input English tweets only into our model to predict depression. We ended up with 49 million tweets, and fed the LR with (MM + TF-IDF) model with these tweets, and the model found that nearly 2 million tweets were classified as depressed tweets. In this section, we show the depression dynamics in NSW during the study period between 1 January 2020 and 22 May 2020. The depression dynamics in different LGAs in NSW are also analysed to demonstrate how COVID-19 pandemic may affect people’s mental health.

A. Depression dynamics in NSW

Fig. 4 presents the overall community depression dynamics in NSW with the confirmed cases of COVID-19 together during the study period between 1 January 2020 and 22 May 2020. “Depression level” refers to the proportion of the

number of depressed tweets over the whole number of tweets each day. From this figure, we can find that people showed a low level of depression during the period before the significant increase in the confirmed cases of COVID-19 until 8 March 2020 in NSW. People’s depression level was significantly increased with a significant increase in confirmed cases of COVID-19. The depression level reached to the peak during the peak outbreak period of COVID-19 on 26 March 2020. After that, people’s depression level decreased significantly for a short period and then kept relatively stable with some short fluctuations. Overall, the analysis clearly shows that people became more depressed after the outbreak of COVID-19 on 8 March 2020 in NSW.

When we drill down into details of Fig. 4, it was found that people’s depression was much sensitive to sharp changes of confirmed cases of COVID-19 on that day or after. For example, people’s depression level had sharp changes around days of 10 March, 25 March, 30 March, and 14 April 2020 (there were sharp changes of confirmed cases of COVID-19 on these days). The sharp increase of confirmed cases of COVID-19 usually resulted in the sharp increase of depression levels (i.e. people became more depressed because of the sharp increase of confirmed cases of COVID-19).

B. Depression under implemented government measures and big events

This subsection investigates the links between people’s depression and implemented government measures for COVID-19 (such as lockdown) as well as big events during the pandemic.

By investigating the labelled topics with the hashtag in the collected Twitter data, it was found that the topics of “lockdown” and “social-distancing” were started to be discussed actively from 9 March when the government encouraged people to increase social-distancing in the life, and the NSW government officially announced the state lockdown on 30 March and the restrictions were begun from 31 March⁹. The NSW government announced the ease of restrictions on 10 May¹⁰. When we link these dates with depression levels as shown in Fig. 4, it was found that people felt significantly more depressed when they started to actively discuss the lockdown restriction on 9 March. People were slightly more depressed after the official implementation of the state lockdown in NSW. The results revealed that the lockdown measure may make people more depressed. However, people still became more depressed even if after the relaxation of lockdown. This is maybe because people still worried about the spread of this severe virus due to the increased community activities.

We are also interested in whether people’s depression was affected by big events during the COVID-19 period. For example, the Ruby Princess Cruise docked in Sydney Harbour on 18 March 2020. About 2700 passengers were allowed to disembark on 19 March without isolation or other measures although some passengers had COVID-19 symptoms at that

⁹https://gazette.legislation.nsw.gov.au/so/download.w3p?id=Gazette_2020_2020-65.pdf

¹⁰<https://www.nsw.gov.au/media-releases/nsw-to-ease-restrictions-week-0>

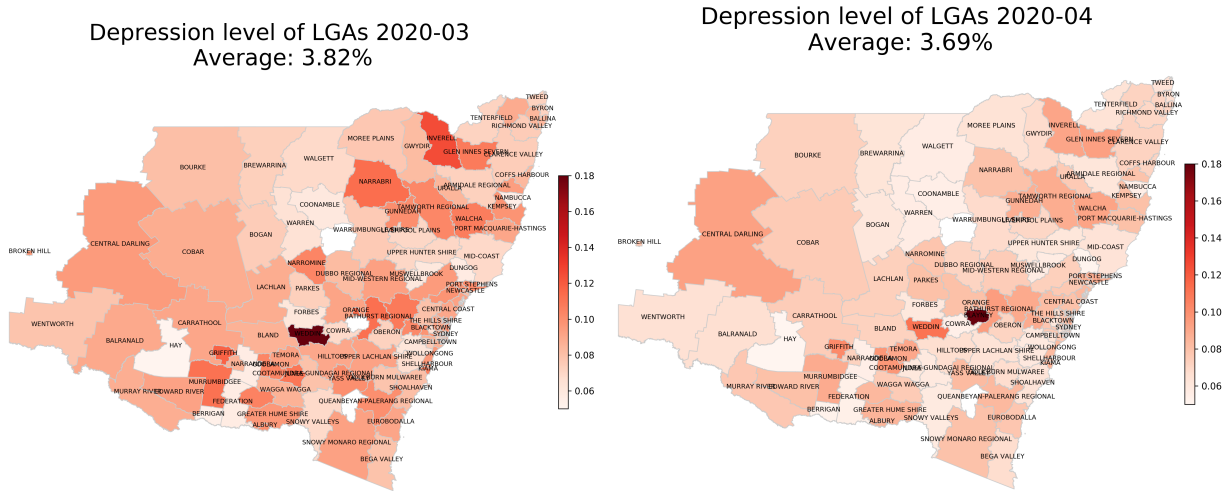


Fig. 5: The choropleth maps of community depression in LGAs in NSW in March 2020 (left) and April 2020 (right).

time, which was considered to create a coronavirus hotbed in Australia. The Ruby Princess Cruise has been reported to be linked to at least 662 confirmed cases and 22 deaths of COVID-19¹¹.

We link the depression dynamics as shown in Fig. 4 with the important dates of the Ruby Princess Cruise (e.g. docking date, disembarking date) and actual timeline of the public reporting of confirmed cases and deaths as well as other events (e.g. police in NSW announced a criminal investigation into the Ruby Princess Cruise debacle on 5 April) related to the Ruby Princess¹¹. We have not found significant changes in people's depression on those dates. This may imply that the big events did not cause people's depression changes significantly.

C. Depression dynamics in LGAs

We further analysed community depression dynamics in LGAs in NSW. Fig. 5 shows examples of choropleth maps of community depression dynamics in LGAs in NSW for two different months which are March 2020 and April 2020. We can observe from the maps that the community depression level was different across different LGAs in each month. Furthermore, the community depression of each LGA changed in different months. On average, people in LGAs were more depressed in March than in April. This is maybe because the number of daily confirmed cases of COVID-19 was significantly increased to a peak and it was gradually decreased in April.

We also dig into more details of depression changes of LGAs around Sydney City areas. For example, Ryde, North Sydney, and Willoughby are three neighbouring LGAs in Northern Sydney. Their community depression dynamics and corresponding confirmed cases of COVID-19 are shown in Fig. 6, respectively. When comparing the dynamics in this figure, we can see that different LGAs showed different depression dynamics maybe because some events specifically related to that LGA. For example, in an aged care centre

in Ryde LGA, a nurse and an 82-year-old elderly resident were first tested positive for coronavirus at the beginning of March¹². After that, a number of elderly residents in this aged care centre were tested positive for COVID-19 or even died. At the same time, a childcare centre and a hospital in this LGA have been reported positive COVID-19 cases in March. Many staff from the childcare centre and the hospital were asked to test the virus and conduct home isolation for 14 days. All these may result in the significant community depression changes in March 2020 as shown in Fig. 6. For example, the depression level in Ryde LGA was changed significantly to a very high level on 10 March 2020 and 16 March 2020 respectively.

However, it was not found that the community depression dynamics in an LGA showed close relations with the dynamics of confirmed cases of COVID-19 in that LGA as we found in the state level. Maybe this is because the community depression dynamics in an LGA was affected largely by the confirmed cases in the overall state but not the local government area. This aligns with our common sense during the COVID-19 pandemic: even if our family is currently safe from COVID-19, we are still worrying about the life because of the continuing significant increases of COVID-19 all over the world especially in big countries.

D. Discussion

COVID-19 pandemic has affected people's lives all over the world. Due to social-distancing measures and other restrictions implemented by the government, people often use social media such as Twitter for socialising.

The results of this study showed that our novel depression classification model can successfully detect community depression dynamics in NSW state level during the COVID-19 pandemic. It was found that people became more depressed after the outbreak of COVID-19 in NSW. People's depression level was much sensitive to sharp changes in confirmed cases of COVID-19, and the sharp increase of confirmed cases made

¹¹<https://www.theguardian.com/world/2020/may/02/australia-coronavirus-lockdown-the-first-50-days>

¹²<https://www.smh.com.au/national/woman-catches-coronavirus-in-australia-40-sydney-hospital-staff-quarantined-20200304-p5461f.html>

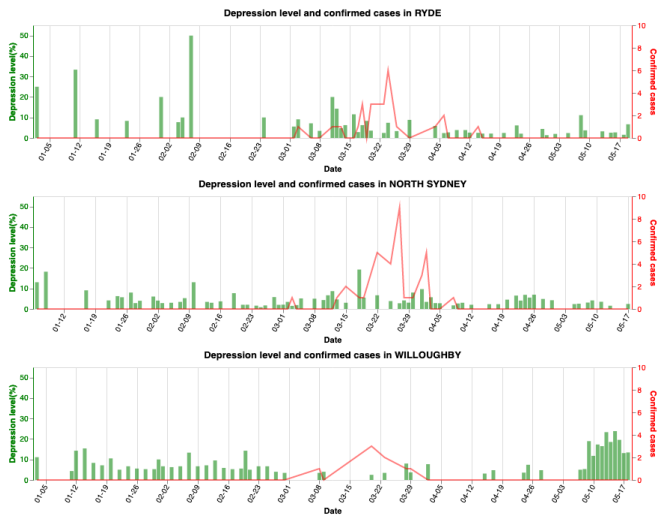


Fig. 6: The community depression dynamics in Ryde, North Sydney, and Willoughby in Northern Sydney between 1 January 2020 and 22 May 2020.

people more depressed. When we conducted a fine-grained analysis of depression dynamics in LGAs in NSW, our novel model can also detect the differences of people’s depression in LGAs as well as depression changes in each LGA at a different period.

The study found that the policies/measures implemented by the government such as the state lockdown has had obvious impact on community depression within a short time period. The implementation of the state lockdown made people more depressed; however, the relaxation of restriction also made people more depressed. This could be primarily because people are still worried about the spread of COVID-19 due to the increased community activities after the relaxation of restrictions.

This study did not find the significant effects of big events such as the Ruby Princess Cruise ship coronavirus disaster in Sydney on the community depression. This is maybe because the data related to the disaster is small and passengers were also from other Australian states and even overseas besides NSW.

Since the proposed approach does not have constraints on locations of Twitter data, the proposed approach can be used to analyse the depression dynamics in other regions or countries. The proposed approach can also be generalised without special considerations to analyse depression dynamics over time and locations because of other events besides COVID-19.

VII. CONCLUSION AND FUTURE WORK

This paper conducted a comprehensive examination of the community depression dynamics in the state of NSW in Australia due to the COVID-19 pandemic. A novel depression classification model based on multi-modal features and TF-IDF was proposed to detect depression polarities from the Twitter text. By using Twitter data collected from each LGAs of NSW from 1 January 2020 until 22 May 2020 as input to our novel model, this paper investigated the fine-grained

analysis of community depression dynamics in NSW. The results showed that people became more depressed after the outbreak of the COVID-19 pandemic. People’s depression was also affected by the sharp changes in confirmed cases of COVID-19. Our model successfully detected depression dynamics because of the implementations of measures by the government. When we drilled down into LGAs, it was found that different LGAs showed different depression polarities during the timeframe of the tweets used in our study, and each LGA may have different depression polarity on different days. It was observed that the big health emergencies in an LGA had a significant impact on people’s depression. However, we did not find significant effects of the confirmed cases of COVID-19 in an LGA on people’s depressions in that LGA as we observed in the state level. These findings could help authorities such as governmental departments to manage community mental health disorders more objectively. The proposed approach can also help government authorities to learn the effectiveness of policies implemented.

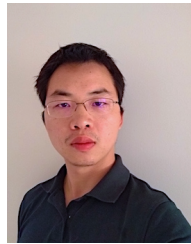
VIII. ACKNOWLEDGEMENT

This work was partially supported by the Australian Research Council Grant No. DP200101374, LP170100891; and the Asian Office of Aerospace Research & Development (AOARD) grant No. 216624.

REFERENCES

- [1] N. Montemurro, “The emotional impact of COVID-19: From medical staff to common people,” *Brain, Behavior, and Immunity*, March 2020.
- [2] J. P. Rogers, E. Chesney, D. Oliver, T. A. Pollak, P. McGuire, P. Fusar-Poli, M. S. Zandi, G. Lewis, and A. S. David, “Psychiatric and neuropsychiatric presentations associated with severe coronavirus infections: a systematic review and meta-analysis with comparison to the COVID-19 pandemic,” *The Lancet Psychiatry*, May 2020.
- [3] World Health Organization, “Depression,” 2020, [Online; accessed 2020-06-15]. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/depression>
- [4] J. Zhou, S. Yang, C. Xiao, and F. Chen, “Examination of community sentiment dynamics due to covid-19 pandemic: a case study from australia,” *arXiv:2006.12185 [cs]*, June 2020. [Online]. Available: <http://arxiv.org/abs/2006.12185>
- [5] A. Leis, F. Ronzano, M. A. Mayer, L. I. Furlong, and F. Sanz, “Detecting signs of depression in tweets in spanish: Behavioral and linguistic analysis,” *JMIR*, vol. 21, no. 6, June 2019.
- [6] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, “Predicting depression via social media,” in *AAAI*, 2013.
- [7] S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, “Recognizing depression from twitter activity,” in *CHFCS*, 2015, pp. 3187–3196.
- [8] W. Yang and L. Mu, “Gis analysis of depression among twitter users,” *Applied Geography*, vol. 60, pp. 217–223, 2015.
- [9] A. Hussein Orabi, P. Buddhitha, M. Hussein Orabi, and D. Inkenp, “Deep learning for depression detection of twitter users,” in *CLCP*, New Orleans, LA, Jun. 2018, pp. 88–97.
- [10] G. Shen, J. Jia, L. Nie, F. Feng, C. Zhang, T. Hu, T.-S. Chua, and W. Zhu, “Depression detection via harvesting social media: A multimodal dictionary learning solution,” in *IJCAI*, 2017, pp. 3838–3844.
- [11] I. Li, Y. Li, T. Li, S. Alvarez-Napagao, D. Garcia-Gasulla, and T. Suzumura, “What are we depressed about when we talk about COVID19: Mental health analysis on tweets using natural language processing,” *arXiv:2004.10899 [cs]*, June 2020. [Online]. Available: <http://arxiv.org/abs/2004.10899>
- [12] S. Yang, J. Jiang, A. Pal, K. Yu, F. Chen, and S. Yu, “Analysis and insights for myths circulating on twitter during the covid-19 pandemic,” *IEEE OJCS*, vol. 1, pp. 209–219, 2020.

- [13] H. Zogan, X. Wang, S. Jameel, and G. Xu, "Depression detection with multi-modalities using a hybrid deep learning model on social media," *CoRR*, vol. abs/2007.02847, 2020. [Online]. Available: <https://arxiv.org/abs/2007.02847>
- [14] X. Wang, C. Zhang, Y. Ji, L. Sun, L. Wu, and Z. Bao, "A depression detection model based on sentiment analysis in micro-blog social network," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2013, pp. 201–213.
- [15] T. Shen, J. Jia, G. Shen, F. Feng, X. He, H. Luan, J. Tang, T. Tiropanis, T. S. Chua, and W. Hall, "Cross-domain depression detection via harvesting social media," in *IJCAI*, 2018.
- [16] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *NIPS*, 2013, pp. 3111–3119.
- [17] T. Gui, L. Zhu, Q. Zhang, M. Peng, X. Zhou, K. Ding, and Z. Chen, "Cooperative multimodal approach to depression detection in twitter," in *AAAI*, vol. 33, 2019, pp. 110–117.
- [18] T. H. Nguyen and R. Grishman, "Relation extraction: Perspective from convolutional neural networks," in *VSM for NLP*, Denver, Colorado, Jun. 2015, pp. 39–48.
- [19] S. Galea, R. M. Merchant, and N. Lurie, "The mental health consequences of covid-19 and physical distancing: The need for prevention and early intervention," *JAMA*, vol. 180, no. 6, pp. 817–818, 2020.
- [20] Y. Huang and N. Zhao, "Generalized anxiety disorder, depressive symptoms and sleep quality during covid-19 outbreak in china: a web-based cross-sectional survey," *Psy Research*, p. 112954, 2020.
- [21] M. Y. Ni, L. Yang, C. M. Leung, N. Li, X. I. Yao, Y. Wang, G. M. Leung, B. J. Cowling, and Q. Liao, "Mental health, risk factors, and social media use during the covid-19 epidemic and cordon sanitaire among the community and health professionals in wuhan, china: Cross-sectional survey," *JMIR*, vol. 7, no. 5, p. e19009, 2020.
- [22] P. K. Novak, J. Smailović, B. Sluban, and I. Mozetič, "Sentiment of emojis," *PloS one*, vol. 10, no. 12, 2015.
- [23] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.
- [24] E. A. O'Connor, E. P. Whitlock, B. Gaynes, and T. L. Beil, "Screening for depression in adults and older adults in primary care: An updated systematic review," Agency for Healthcare Research and Quality (US), Evidence Syntheses 75, December 2009. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK36406/table/ch1.1/>
- [25] R. Singh, J. Du, Y. Zhang, H. Wang, Y. Miao, O. A. Sianaki, and A. Ullah, "A framework for early detection of antisocial behavior on twitter using natural language processing," in *Conference on Complex, Intelligent, and Software Intensive Systems*. Springer, 2019, pp. 484–495.



Shuiqiao Yang received his B.S. degree in 2012 from Shanghai Ocean University, China, M.S. degree in 2016 from Zhejiang University, China, and Ph.D. degree in 2020 from School of Information Technology, Deakin University, Australia. He currently works as a postdoc at Data Science Institute, University of Technology Sydney, Australia. His research interests include text mining, machine learning and data analysis.



Shoaib Jameel is a Lecturer in Computer Science and Artificial Intelligence at the School of Computer Science and Electronic Engineering, University of Essex in the United Kingdom (UK). His research interests span from text mining, natural language processing and computer vision. He received his PhD from the Chinese University of Hong Kong in 2014. He works with various technology startups in the UK spearheading their technical sphere, where his research outputs are directly applied to their production systems. He is the Fellow of the Higher

Education Academy.



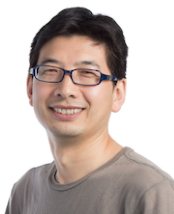
Guandong Xu is a Full Professor and Program Leader at School of Computer Science and Advanced Analytics Institute, University of Technology Sydney, Australia and he received PhD degree in Computer Science from Victoria University, Australia. His research interests cover Data Science, Data Analytics, Recommender Systems, Web Mining, User Modelling, NLP, Social Network Analysis, and Social Media Mining. He has published three monographs in Springer and CRC press, and 250 + journal and conference papers. He is the assistant

Editor-in-Chief of World Wide Web Journal, and served as guest editors for Pattern Recognition, IEEE Transactions on Computational Social Systems, Journal of Software and Systems, World Wide Web Journal, Multimedia Tools and Applications, and Online Information Review..

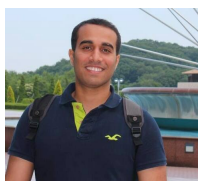


Fang Chen is a Distinguished Professor in AI and data science at the University of Technology Sydney (UTS), Australia. She is the Executive Director Data Science and Distinguished Professor, the University of Technology Sydney. Prof. Chen's research interests include machine learning, AI ethics, behaviour analytics and their applications. She has created many world-class AI innovations while working in Intel, Motorola, NICTA, CSIRO and now UTS, and helped governments and industries utilising data and significantly increasing productivity, safety and

customer satisfaction.



Jianlong Zhou is a Senior Lecturer in Data Science Institute, University of Technology Sydney, Australia. His research work focuses on ethics of AI, AI fairness, AI explainability, data analytics, visual analytics, behaviour analytics, human-computer interaction, and related applications. Dr. Zhou is a leading senior researcher in trustworthy and transparent machine learning, and has done pioneering research in the area of linking human and machine learning.



Hamad Zogan received his BSc degree from King Abdulaziz University, Kingdom of Saudi Arabia, and MSc from Saint Joseph's University, United States. He is currently a PhD student at the University of Technology Sydney with full scholarship from Jazan University, Saudi Arabia. His research interests in Data Analytics, Social Network Analysis, and explores novel methods in machine learning to discover mental illness on social media.