# International Journal of Advanced Trends in Computer Science and Engineering

# Towards a Multimodal Analysis to Predict Mental Illness in Twitter Platform

**Nurul Nadhrah Kamaruzaman[1], Nor Azura Husin, Norwati Mustapha, Razali Yaakob**

Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang,
Selangor Darul Ehsan, Malaysia
[1] nurul.nadhrah0111@gmail.com, {n_azura, norwati, razaliy}@upm.edu.my

## ABSTRACT

The rising number of people with mental illness has become a major concern all over the world. Many efforts have been done to improve the process of detection and surveillance of people with mental illness and one of them is through analysing users' activities on social media platforms. Social media contains multi-content of data such as text, image and social interactions log. By mining all the data, it could help to determine the current mental state of social media users and detect those who are suffering from mental illness. However, most of the prior researches on mental illness in social media have mainly focused on single-based content rather than multi-based content. Due to a rapid growth of multi-content data, considering on single-based content only will ignore the complementary information that offered by other modalities. In consequence, single-based content does not provide an insight knowledge on the phenomenon of interest. This study aims to propose a design of multimodal fusion model, with the exploitation of all the multi-content data including text, image and social interaction data. This multimodal fusion model is expected to produce an accurate result for the prediction of mental illness among social media users.

**Key words:** Fusion, Mental Illness, Multimodal, Twitter, Social Media

## 1. INTRODUCTION

Mental illness has become a rising issue all over the world. Mental illness is associated with a wide range of mental health disorder that affects one's thinking, mood, and behaviour. There are many different types of mental disorders and some common ones include depression, anxiety disorder, bipolar disorder, and schizophrenia. It is estimated that 1 in 5 adult experiences a mental illness problems in every year with one in 17 suffers with serious mental illness such as bipolar disorder or schizophrenia [1]. Half of mental illness conditions considerably begin by age 14, and 75% of the illness develop by age 24 [1].

There is an urgent need to diagnose and treat the mental illness disease. Majority of people with mental illness problem often go undiagnosed and untreated. In a particular disease such as depression, about half of the cases have been diagnosed by primary care physicians [2] and only 13–49% are getting minimally adequate treatment [3]. Early intervention is essential to lessen the severity of illness and avoid major impact of mental illness from arise [4]. Learning about early warning signs, developing symptoms and getting treatment can really help for those who are affected with mental illness problems.

Research has shown that people with mental illness problems, especially those with depression and anxiety disorder, have high usage of social media [5]. They are often turning to social media to share their illness struggles, experiences and seek advice from other people with similar mental health conditions. Besides that, this social media becomes a very important platform for people with mental illness to overcome barriers such as social isolation, reluctance to implement a formal health care services and access timely help.

Twitter is one of the largest platforms where people have a tendency to express their opinion, emotions and share their daily lives. Study shows that approximately 23% of online adults prefer to use Twitter with more than 500 million tweets are sent per day [6]. These rich big data of Twitter have attracted the attention of many researchers to understand the sentiments of social media users. The information extracted from Twitter can provide a real-time and powerful indicator for public sentiment [7]. Many researcher have utilized the data from Twitter as a sensor in many health care applications especially in mental illness surveillance. In other words, it is possible to automate the detection of mental illness problems through social media platform by implementing data mining, natural language processing and machine learning techniques. The difference in actions and patterns of user interaction expressed in social media can be detected through learning information of social media data based on text mining, social interaction analysis, and image analysis [8].

Therefore, this study aims to propose a design of multimodal fusion model, with the exploitation of all the multi-content data including text, image and social

interaction data. The rest of the paper is organized as follows. Section 2 summarize the literature review of previous work. Section 3 presents the overview of proposed design multimodal model. Section 4 presents the expected outcome and future works. Finally, we conclude this study in section 5.

## 2. LITERATURE REVIEW

Utilizing social media data to predict mental illness symptom has gained popularity among researchers to explore further and deeper in this particular area. Various machine learning model has been designed, improved and implemented in order to get the best result in predicting mental illness in social media. Table 1 shows the summaries of prior works relating to prediction of mental illness from social media platform.

**Table 1**: Summaries of the Literature Review on Prediction of Mental Illness from Social Media Platform

| Author | Platform | Type of Mental Illness | Data | Method | Performance |
|--------|----------|------------------------|------|--------|-------------|
| [9] | Twitter | Depression | Text Social Int. | SVM | Accuracy: 72.384% |
| [10] | Twitter | Depression, PTSD | Text | Decision List | Precision:70 – .76 |
| [11] | Twitter | Bipolar | Text Social Int. | Random Forest Classifier | Precision:91% and 92% |
| [12] | Twitter | Depression | Text Image | Linear SVM | Accuracy Text : 84.45 Text + Image: 90.04 |
| [13] | Twitter | Depression, Bipolar, Anxiety, PTSD, Self-Harm | Text | Neural Network | AUC>0.8 |
| [14] | Live-journal Facebook Twitter | Depression | Text | SVM Naïve Bayes | Accuracy SVM:57% Naïve Bayes:63% |
| [15] | Facebook | Depression | Text Social Int. | Convolutional Neural Network | Precision: 71.4% |
| [16] | Twitter | Depression | Text | Convolutional Neural Network | Accuracy: 87.957% |

### 2.1 Multimodal Analysis In Predicting Mental Illness

Mining the Twitter big data is certainly challenging and time-consuming process due to the limitations, restrictions and complication in its content. Gathering tweets automatically and simultaneously allow to extract the valuable information for mental illness domain makes the task even more complex. The lack of gold standard datasets is the most distressing problem since we are not able to benchmark any analysis performed in this particular domain. Prior studies show that, most of the research relating to mental illness in social media was mainly focused on text-based content rather than considering others content formats [17]. Study from [12] suggests that by analysing multimodal information from multi-content data of social media, it will allow a better understanding in emotion-related behaviour of user with mental illness. As the result, it will increase the performance of the model

and achieve a more promising result for the prediction of mental illness.

### 2.2 Analysing Text Modality

Social media textual data are huge, noisy and consist of a lot of abbreviation, slang and other types of informal language expressions. One of the major difficulties in text extraction and classification of social media text data is the existence of different types of errors, such as spelling and grammatical errors. Furthermore, to understand linguistic pattern from texts displayed by those experiencing mental illness is confusing as it contains ambiguities that need to be interpreted correctly. Several studies used the established linguistic inquiry and word count (LIWC) to extract potential signals of mental illness from textual content (example: the recurrence word of the first personal pronoun "I" or "me" ; positive and negative emotions used in a post) [9][18][19]. However, this approach does have disadvantages like ignoring the semantic of words and word order in one's sentence. Apart from that, many researchers have utilized different machine learning algorithms for feature extraction of textual data [12][11] [14][13]. Nevertheless, while the machine learning is simple and efficient, it shows certain weaknesses, including the reliance on manual training data and non-replicable results for other domains. Moreover, the feature extraction accuracy of machine learning algorithms is poor due to their deficiencies related to synonymy, polysemy, ambiguities and semantics.

### 2.3 Analysing Image Modality

Apart from text, the users are also allowed to post images on social media platform. Some studies have mentioned that the images also play an important role in understanding the cues of mental illness [12]. According to [20], users who are diagnosed with depression were more likely to post images that are bluer, greyer and darker compared to healthy users who are generally preferred to post images with brighter and more vivid colours [21][22][23][24]. In line with the findings, some study has implemented colour composition extraction based colour image scale (CIS) [12], while some study has implemented pixel-analysis by computing pixel-level average of three colours properties; Hue, Saturation and Value (HSV) which are commonly used in image processing [20].

### 2.4 Analysing Social Interaction Modality

Social media platform contains millions of interactions and relationships among users. It allows ones to connect with others not only by adding friends, follow and un-follow, but also post, tweets, like, retweet and comment to other users of social media. According to [9], depressed users are usually have lower numbers of followers and followees besides reduced reciprocity in communication with other users, indicating their decreased desire for social interaction. Nevertheless, it is certainly difficult to understand the cues relating to symptoms of mental illness due to millions of interactions and relationships among users, which can be obtained in social media platforms [25].

## 3. PROPOSED DESIGN OF MULTIMODAL FUSION MODEL

Figure 1 illustrate the design of multimodal model to predict mental illness in social media. Multimodal fusion is

a process of combining various modalities of data for analysis tasks. In this study, the early fusion technique will be implemented whereby the features of all unimodal data will be extracted before combined them into a single representation of vector form. We will concatenate the text vector with the image vector and social interaction vector. We will then obtain the early fusion vector after the feature normalization that serves as the input for a classifier. Fully-connected layer (pooling layer) will be applied whereby the input learns the semantic concept for the early fusion scheme thus, obtain the prediction.
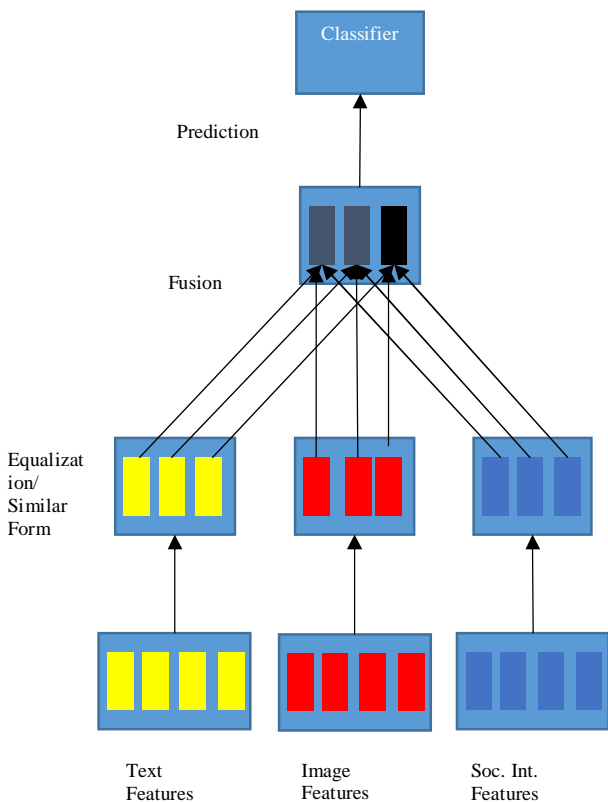


**Figure 1**: Design of Multimodal Fusion Model

The proposed design of multimodal model caters several important elements including capturing the semantic and resolving the ambiguity from textual data and utilizing the image and social interaction data for profiling and better understanding of emotion and behaviour of social media users. In this study, we will focus on utilizing data from Twitter platform. The modalities of data used will include text data, image data and social interaction data. For text data analysis, focus will be given on analysing linguistic pattern of people with mental illness problems while for image data analysis, it will based on colour descriptor or in particular, colour composition. For social interaction data analysis, there are several things that will be looking upon in the analysis which are number of tweets, number of likes, number of retweets, number of comments by users and etc.

### 3.1 Textual Feature Extraction

Figure 2 illustrates the Convolutional Neural Network (CNN) that will be applied on each words for the

extraction of textual feature. The first layer will go through the process of converting from text original format which is words, into a *d*-dimensional word2vec vectors. This embedding layer takes a sequence of words, S= <*w*1, *w*2… *w*n> as input where each words comes from pre –defined vocabulary. The next layer will perform convolutions over the embedded word vector using multiple filter sizes before max pooling the result of the next convolutional layer into a long feature vector and classify the output using a softmax layer.
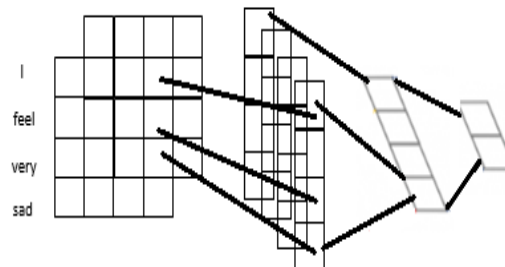


**Figure 2:** Convolutional Neural Network for Textual Features Extraction

### 3.2 Image Feature Extraction

Image feature extraction will be based on the method mentioned in [12]. Basically, the procedure starts with the segmentation of the images using quantization and mean-shift clustering. The colour composition of seed regions with high importance values will be calculated by multiplying the regions' area and Gaussian distance. The measurement of the distance is between the centroid of the region and centre of image. The extracted colour composition will be then compared with specific colour image scale (CIS) and match to the colour compositions with high similarity scores.

### 3.3 Social Interactions Feature Extraction

The feature extraction of social interaction is mainly to measure the behaviour of users on social media. Based on research from [9], this study present a set of attributes that can be used to characterize the behavioural differences of users. The set of attributes consists of engagement, egocentric network and diurnal and nocturnal activity. For engagement attribute, features that will be extracted include the number of tweets per day, number of replies, number of retweets and proportion links (urls) shared over a day. For egocentric network, feature that will be extracted is mainly on the number of replies and favourites tweets per day. Meanwhile, for diurnal and nocturnal activity, the feature extraction will focus on pattern of posting which is based on number of tweets posted at a certain time and also the frequency of tweets posted per hour in certain period of time. All the measurement obtained from each of the attributes will be then used for constructing the feature vectors.

## 4. EXPECTED OUTCOME AND FUTURE WORK

The integration of multimodal in designing mental illness prediction model is a promising approach to complement various channels of information specifically in mental health analysis and sentiment analysis. Based on Figure 3, it is clearly shows that multimodal method often outperform the unimodal method in term of accuracy of the model. The finding thus motivate this study to implement multimodal analysis in creating a model to predict mental illness in social media.
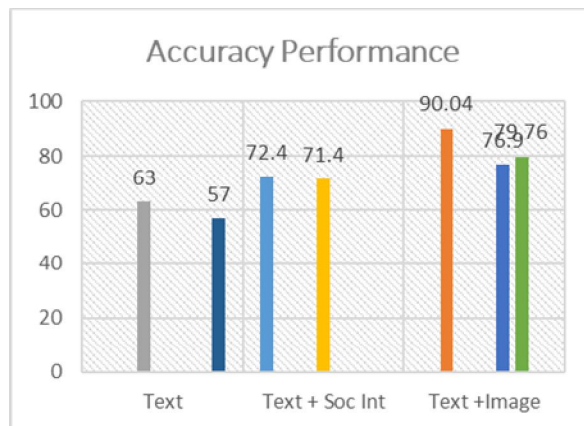


**Figure 3**: Accuracy Performance of Prediction Model

Apart from that, it is certainly challenging to implement early fusion as the features need to be brought into same format before the fusion process takes place. Besides, the features obtained which belongs to diverse modalities can differ widely in many aspects thus, it may take a longer time to process and extracted. However, despite the difficulty to combine features into common representations, early fusion yields a good feature representation as the correlation between various multimodal features at the early stage can potentially serve better task accomplishment besides avoid the potential loss of correlation in mixed feature space.

As for the future work, a set of experiments will be performed in order to evaluate the effectiveness of the multimodal model by comparing the performance of baseline and the proposed model. To prove the effectiveness of the proposed model, experiments with large amount of tweets are necessary and labelled with the accurate ground truth. The performance of the model will be then measured based on accuracy and F1-score that include precision and recall metrics.

## 5. CONCLUSION

In conclusion, the implementation of multimodal information from multi-content data on social media will allow for a better understanding in emotion-related behaviour of users. In compliance with this statement, this study propose a design of multimodal fusion model to predict mental illness among users in social media based on the proposed framework stated earlier in this paper. The exploitation of the multi-content data for the development of the model will be expected to increase the accuracy and the performance of the prediction. Therefore, with the creation of the multimodal fusion model, this study will help in the diagnosis of mental illness through social media big data thus, contribute to the field of public health by providing an early intervention for people with potential mental illness problems, across the world.

## REFERENCES

[1] NAMI, **National Alliance on Mental Illness**, *Natl. Alliance Ment. Illn.*, p. Mental Health Conditions, 2019.

[2] M. Cepoiu, J. McCusker, M. G. Cole, M. Sewitch, E. Belzile, and A. Ciampi, **Recognition of depression by non-psychiatric physicians—a systematic literature review and meta-analysis**, *J. Gen. Intern. Med.*, vol. 23, no. 1, pp. 25–36, 2008. https://doi.org/10.1007/s11606-007-0428-5

[3] P. S. Wang, M. Lane, M. Olfson, H. A. Pincus, K. B. Wells, and R. C. Kessler, **Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication**, *Arch. Gen. Psychiatry*, vol. 62, no. 6, pp. 629–640, 2005. https://doi.org/10.1001/archpsyc.62.6.629

[4] MHA, **Prevention and Early Intervention in Mental Health,** 2018. [Online]. Available: http://www.mentalhealthamerica.net/issues/preventio n and-early-intervention-mental-health.

[5] J. Davila, R. Hershenberg, B. A. Feinstein, K. Gorman, V. Bhatia, and L. R. Starr, **Frequency and quality of social networking among young adults: Associations with depressive symptoms, rumination, and corumination.**, *Psychol. Pop. media Cult.*, vol. 1, no. 2, p. 72, 2012. https://doi.org/10.1037/a0027512

[6] M. Duggan, N. B. Ellison, C. Lampe, A. Lenhart, and M. Madden, **Social media update 2014. Pew Research Center**, 2017.

[7] A. Signorini, A. M. Segre, and P. M. Polgreen, **The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic,** *PLoS One*, vol. 6, no. 5, p. e19467, 2011. https://doi.org/10.1371/journal.pone.0019467

[8] M. A. Moreno *et al.*, **Feeling bad on Facebook: Depression disclosures by college students on a social networking site**, *Depress. Anxiety*, vol. 28, no. 6, pp. 447–455, 2011.

[9] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, **Predicting depression via social media**, *ICWSM*, vol. 13, pp. 1–10, 2013.

[10] T. Pedersen, **Screening twitter users for depression and ptsd with lexical decision lists,** in *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*, 2015, pp. 46–53.

[11] E. Saravia, C.-H. Chang, R. J. De Lorenzo, and Y.-S. Chen, **MIDAS: Mental illness detection and**

**analysis via social media**, in *Advances in Social Networks Analysis and Mining (ASONAM), 2016*

[12] K. Kang, C. Yoon, and E. Y. Kim, **Identifying depressive users in Twitter using multimodal analysis**, in *Big Data and Smart Computing (BigComp), 2016 International Conference on*, 2016, pp. 231–238.

[13] A. Benton, M. Mitchell, and D. Hovy, **Multi-task learning for mental health using social media text**, *arXiv Prepr. arXiv1712.03538*, 2017. https://doi.org/10.18653/v1/E17-1015

[14] M. M. Aldarwish and H. F. Ahmad, **Predicting depression levels using social media posts,** in *Autonomous Decentralized System (ISADS), 2017 IEEE 13th International Symposium on*, 2017, pp. 277–280.

[15] M. Y. Wu, C.-Y. Shen, E. T. Wang, and A. L. P. Chen, **A deep architecture for depression detection using posting, behavior, and living environment data,** *J. Intell. Inf. Syst.*, pp. 1–20, 2018.

[16] A. H. Orabi, P. Buddhitha, M. H. Orabi, and D. Inkpen, **Deep Learning for Depression Detection of Twitter Users**, in *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, 2018, pp. 88–97.

[17] L. Zhou, D. Zhang, C. C. Yang, and Y. Wang, **Harnessing social media for health information management**, *Electron. Commer. Res. Appl.*, vol. 27, pp. 139–151, 2018. https://doi.org/10.1016/j.elerap.2017.12.003

[18] G. Coppersmith, M. Dredze, and C. Harman, **Quantifying mental health signals in Twitter,** in *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic*

*IEEE/ACM International Conference on*, 2016, pp. 1418–1421.
*Signal to Clinical Reality*, 2014, pp. 51–60.

[19] M. R. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, and A. Ulhaq, **Depression detection from social network data using machine learning techniques**, *Heal. Inf. Sci. Syst.*, vol. 6, no. 1, p. 8, 2018.

[20] A. G. Reece and C. M. Danforth, **Instagram photos reveal predictive markers of depression**, *EPJ Data Sci.*, vol. 6, no. 1, p. 15, 2017. https://doi.org/10.1140/epjds/s13688-017-0110-z

[21] C. J. Boyatzis and R. Varghese, **Children's emotional associations with colors,** *J. Genet. Psychol.*, vol. 155, no. 1, pp. 77–85, 1994.

[22] M. Hemphill, **A note on adults' color–emotion associations,** *J. Genet. Psychol.*, vol. 157, no. 3, pp. 275–280, 1996.

[23] C. B. Barrick, D. Taylor, and E. I. Correa, **Color sensitivity and mood disorders: biology or metaphor?**, *J. Affect. Disord.*, vol. 68, no. 1, pp. 67–71, 2002. https://doi.org/10.1016/S0165-0327(00)00358-X

[24] H. R. Carruthers, J. Morris, N. Tarrier, and P. J. Whorwell, **The Manchester Color Wheel: development of a novel way of identifying color choice and its validation in healthy, anxious and depressed individuals**, *BMC Med. Res. Methodol.*, vol. 10, no. 1, p. 12, 2010.

[25] A. Wongkoblap, M. A. Vadillo, and V. Curcin, **Researching Mental Health Disorders in the Era of Social Media: Systematic Review**, *J. Med. Internet Res.*, vol. 19, no. 6, pp. e228–e228, Jun. 2017. https://doi.org/10.2196/jmir.7215