A Survey on Mental Health Detection in Online Social Network

Rohizah Abd Rahman[#1], Khairuddin Omar[#2], Shahrul Azman Mohd Noah[#3],
Mohd Shahrul Nizam Mohd Danuri[*4]

Faculty of Information Science &Technology, National University of Malaysia, 43600 Bangi, Selangor, Malaysia
E-mail: [1]rohizah@ukm.edu; [2]myko@ukm.edu;[3]myshahrul@ukm.edu.my

* Faculty of Science & Information Technology, International Islamic University College Selangor, Bandar Seri Putra, 43000
Kajang, Selangor, Malaysia
E-mail: msnizam@kuis.edu.my

Abstract – **Mental health detection in Online Social Network (OSN) is widely studied in the recent years. OSN has encouraged new ways to communicate and share information, and it is used regularly by millions of people. It generates a mass amount of information that can be utilised to develop mental health detection. The rich content provided by OSN should not be overlooked as it could give more value to the data explored by the researcher. The main purpose of this study is to extract and scrutinise related works from related literature on detection of mental health using OSN. With the focus on the method used, machine learning algorithm, sources of OSN, and types of language used for the mental health detection were chosen for the study. The basic design of this study is in the form of a survey from the literature related to current research in mental health. Major findings revealed that the most frequently used method in mental health detection is machine learning techniques, with Support Vector Machine (SVM) as the most chosen algorithm. Meanwhile, Twitter is the major data source from OSN with English language used for mental health detection. The researcher found a few challenges from the previous studies and analyses, and these include limitations in language barrier, account privacy in OSN, single type of OSN, text analysis, and limited features selection. Based on the limitations, the researcher outlined a future direction of mental health detection using language based on user's geo-location and mother tongue. The use of pictorial, audio and video formats in OSN could become one of the potential areas to be explored in future research. Extracting data from multiple sources of OSNs with new features selection will probably improve mental health detection in the future. In conclusion, this research has a big potential to be explored further in the future.**

Keywords— **stress; depression; Twitter; big data; machine learning**

I. INTRODUCTION

Mental health is a leading cause of disability worldwide. It is estimated that nearly 300 million people suffer from depression [1]. Globally, about 450 million people worldwide are mentally ill, in which the disease accounts for 13% of global disease burden [2]. The World Health Organisation estimates that 1 in 4 individuals experiences mental disorders at any stage of their lives. Meanwhile, depression contributes 4.3% of the total global disease burden. In 2015, the Ministry of Health Malaysia found the mental health problems among Malaysians aged 16 and above were 29.2% or nearly an estimated population of 4.2 million Malaysians [3]. The researcher believes that factors that may contribute to mental health probably come from individuals' way of life such as work stress, bad financial situations, family issues, relationship problems, as well as violence and environmental factors.

The traditional method of mental health detection is based on face-to-face interviews, self-reported, or distribution of questionnaires, which is usually labor-intensive and time-consuming [4]. Meanwhile, several studies have used technologies to improve detection of mental health. Some examples are a technological device used to measure blood pressure and heart rate, development of physiological sensors, and use of wearable sensor and smartphone to improve mental health detection [5]–[8]. Furthermore, the mental health detection using data from the online social network (OSN) were explored by several researchers [4], [9], [18], [19], [10]–[17]. These conclude the relationship between OSN and the mental health field of research. Recently, research on mental health detection in OSN was explored by both western and eastern scholars. Data were taken based on specific geographical locations, and some researchers have developed new methods in mental health detection.

This new trend of research is related to the leapfrog of big data research. The growing availability of resources on the internet through social media (e.g., Facebook, Twitter,

Instagram, and Weibo) has become a medium of communication and sharing information which leads to the overflow of data that might be useful for further exploration and analysis [20], [21], [22]. OSN has created a massive social communication among users, and this huge amount of communication data become content generators [23]. Researchers can explore the phenomenon of research by using big data tools for data analysis [24].

However, previous research has outlined some limitations such as the language barrier, account privacy policy, data extraction format and sources from another OSNs service provider. In order to determine mental health from OSN, previous studies used the English language instead of other languages [9]–[11], [15]–[19]. The researcher believes that future research should focus on using other languages, especially the local languages, for better results in mental health detection. Meanwhile, the researchers also should be aware of the changes of account privacy policy by the OSN service providers before data extraction to avoid any difficulty in the future. The format of data extraction also needs to be widely explored and should not be limited to text format extraction. The researchers may explore other types of format such as pictorial, audio or video format, and also from other sources such as WhatsApp, Instagram and other OSN service providers in the future. These interesting studies influence researcher to explore further research on improving the method of mental health detection in OSN.

II. MATERIAL AND METHOD

A. Current Work

Recently, there is a myriad of research works conducted on mental health detection in OSN. Previous research emphasized the use of language, implementation of algorithms in machine learning, type of OSN service providers and users' geo-location detection of OSN. In 2017 in China, a hybrid model that combines Factor Graph and Convolutional Neural Network algorithm was used to improve the performances of stress detection based on social interactions among Twitter users. Within one week, the researcher collected 3200 tweets. The results showed that the hybrid model, with the proposed selection features, greatly improved the detection performances with F1-Score 6-9% compared to other machine learning algorithms [4].

Meanwhile, in the same country, the researcher collected tweets from 124 students at Weibo between January 1, 2012, and February 2, 2015. The researcher used the previous stress detection method and discovered a correlation between stressor events and stress, and the significant causes of stress were self-cognition domain and school life domain [13].

In 2014, another researcher from China compared several machine learning techniques to prove that their proposed method is better or not. The data were harvested within three years, i.e., from October 2009 to October 2012 using Weibo. The experimental results showed that the proposed method is effective and efficient in detecting psychological stress from microblog data [12]. The researcher believes that language is one of the limitations in detecting stress. However, implementing natural language processing technique, such Unigram can help in understanding the corpus. Through this effort, the studies have contributed a new dataset with new proposed features that can be applied in other methods in the future.

Another researcher from Japan has evaluated the effectiveness of using a user's social media activities for estimating the degree of depression. Data were collected from Twitter. The results showed that depression could be recognized in users with an accuracy of approximately 69%, topics extracted are useful features and timing for observation was important in improving the accuracy [14].

In 2017, an effective method was developed to determine the stress level among Facebook users in Greece. The researcher applied for permission from the same respondents to access their Facebook accounts after giving a set of questionnaires. Data were analyzed using statistical analysis and compared with several machine learning algorithms. The results showed that the Support Vector Machine algorithm had better F1-score value for most categories and all the values for precision, recall, and f-score are greater than 70% [9].

There is also a researcher who distributed the questionnaire before he/she accessed the respondents' social network. The researcher collected data from Twitter from September to October 2012 to predict depression via social media users. The results showed that the users with depression had lower social activities, greater negative emotion, high self-attentional focus, increased relational and medicinal concerns and heightened expression of religious thoughts [18]. Kandias et al. [9] used a questionnaire before getting permission to access their respondents' social network account. Detecting mental health users through questionnaire and accessing their social accounts giving advantages to monitoring the mental health users' social interactions. Meanwhile, some researcher had found that mental health users had less communication in OSN [4], [16], [18].

In the United States, several studies have been conducted regarding stress. Data were harvested from June to July 2009 from Twitter to detect mood depression in 2012. The research found that users in the depressed group were more likely to post tweets about themselves rather than interacting with other users compared to typical Twitter users [16]. In the same country, a researcher collected data from Web Forum using English as the data language. The results showed significant improvements in multi-label classification accuracy using human-generated rationales in support of annotated distress labels [17].

Meanwhile, a study was conducted among the United States military officers using data harvested from their Twitter accounts with posting of 3200 tweets in two weeks. The data were collected from the users who had already mentioned that they had Post Traumatic Stress Disorder (PTSD) among US military officers. The results showed that approximately 1% more PTSD like tweets in military areas than civilian areas and that approximately 0.7% more PTSD like tweets in frequently deploying military areas compared to less frequently deploying areas [15].

This current study of mental health detection is not specified their geo-location of tagging data in the online social network in general. The researcher collected data from the Twitter from 18th February to 23rd April 2014. The

researcher examined whether the level of concern for a suicide-related post on Twitter could be determined based solely on the content of the post, as judged by human coders and then replicated by machine learning [19].

Meanwhile, a few researchers developed a new method in stress detection using Twitter as a data source. They proposed a new system call Tensi-Strength, which is to detect the strength of stress and relaxation expressed in social media text messages. The results showed that Tensi-Strength could detect expressions of stress and relaxation in tweets with a reasonable level of accuracy compared to human coders [10].

Another researcher had proposed a new method in mental health detection, which is Social Network Mental Disorders Detection (SNMDD). Data were collected via Amazon Mechanical Turk (MTurk) and one set of questionnaires was distributed before data collection. However, the researcher argued that it was not easy to detect mental problems via online social activity logs [11].

Proposing a new method in detecting mental health via online social network also contributes to early detection of mental health. Either the method is new, or based on the previous research, the researcher believes there is a potential of implementing various methods in this research. There are many techniques available in the literature survey which have shown the possibility of mental health detection using OSN.

B. General Architecture of Mental Health Detection

Numerous research works have applied a few steps for mental health detection in OSN. The researcher believes that this general architecture is a common step of future research implementation in mental health detection. The general architecture for mental health detection consists of several steps such as social network, data extraction using the keyword, data pre-processing, features selection, data classification using machine learning algorithms and early mental health detection. Figure 1 illustrates the steps in the general architecture based on the input-process-output in the principles of information systems [25], [26].
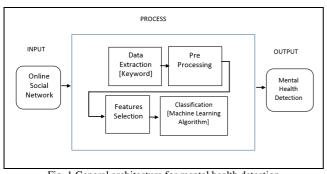

Fig. 1 General architecture for mental health detection

The most previous research found the OSN service providers as data input. Then, for the process part, a few steps were implemented such as data extraction based on keyword. After extracting the data, the pre-processing data must be conducted to eliminate the outliers before the features selection steps. The final process is modeling the data using the machine learning technique.

III. RESULTS AND DISCUSSION

A. Analysis of Existing Works in Mental Health Detection

Recently, research in mental health detection increases intensely. Many researchers focus on mental health detection through data from online social media. These analyses of the existing works focus on the implementation method of mental health detection, list of machine learning algorithm tested, and the matrix of using the online social network, language, and the data set collection.

TABLE I
THE IMPLEMENTATION OF METHOD IN MENTAL HEALTH DETECTION LITERATURE

Study	Author	Year	Method					
			Questionnaire	Machine Learning Algorithm	Hybrid Algorithm	Statistical/Empirical Analysis	New Method	Natural Language Processing
[4]	Huijie Lin	2017		√	√			√
[9]	Miltiadis Kandias	2017	√	√		√		
[10]	Mike Thelwall	2017		√			√	√
[11]	Hong-Han Shuai	2017	√		√			
[12]	Huijie Lin	2014		√				
[13]	Qi Li	2017				√		
[14]	Sho Tsugawa	2015	√	√				
[15]	Glen A. Coppersmith	2014					√	√
[16]	Minsu Park	2012	√					√
[17]	Shirin Saleem	2012		√		√		√
[18]	Munmun De Choudhury	2014	√	√				√
[19]	Bridianne O'Dea	2015		√		√		√

Table 1 shows the existing studies using different methods in mental health detection. All these research works to highlight the comparison between machine learning algorithms to show a better technique [4], [9], [11]–[14], [16]–[19]. Besides that, a few researchers have implemented a hybrid method in mental health detection [4], [11].

In addition, another researcher conducted a survey using a questionnaire and compared between machine learning algorithms [9], [14], [18]. Data from the questionnaire were analyzed using statistical or empirical analysis. On the other hand, some researchers have proposed new methods in mental health detection [4], [11]. Implementing natural language processing as a method will also benefit in detecting mental health problems [4], [10], [15]–[19].

Table 2 shows a summary of the machine learning algorithms tested in mental health detection literature. Machine learning is a subset of artificial intelligence in computer science. Machine, or computer, will become intelligent once data are being trained by the machine learning algorithm. Machine learning algorithms have been widely used in text classification such as Naïve Bayes, Linear Regression, Support Vector Machine, Random Forest, Gradient Boosted Decision Tree, K-Means and Deep Neural Network.

Study	Author	Year	Machine Learning Algorithm					
			Support Vector Machine	Random Forest	Linear Regression	Deep Neural Network	Decision Tree	Naïve Bayes
[4]	Huijie Lin	2017	√	√		√	√	
[9]	Miltiadis Kandias	2017	√		√			√
[10]	Mike Thelwall	2017	√				√	√
[11]	Hong-Han Shuai	2017	√					
[12]	Huijie Lin	2014	√			√		
[14]	Sho Tsugawa	2015	√					
[17]	Shirin Saleem	2012	√					
[18]	Munmun De Choudhury	2014	√					
[19]	Bridianne O'Dea	2015	√		√			

TABLE III
THE SIMPLE MATRIX USED OF ONLINE SOCIAL NETWORK IN MENTAL
HEALTH DETECTION

Study	Author	Year	Online Social Network				
			Twitter	Facebook	Weibo	Web Forum	Others
[4]	Huijie Lin	2017	√				
[9]	Miltiadis Kandias	2017		√			
[10]	Mike Thelwall	2017	√				
[11]	Hong-Han Shuai	2017					√
[12]	Huijie Lin	2014			√		
[13]	Qi Li	2017			√		
[14]	Sho Tsugawa	2015	√				
[15]	Glen A. Coppersmith	2014	√				
[16]	Minsu Park	2012	√				
[17]	Shirin Saleem	2012				√	
[18]	Munmun De Choudhury	2014	√				
[19]	Bridianne O'Dea	2015	√				

TABLE IV
THE SIMPLE MATRIX USED OF LANGUAGE AND DATA COLLECTION IN
MENTAL HEALTH DETECTION

Study	Author	Year	Language		Data Set	
			English	Others	New Data Set	Previous Data
[4]	Huijie Lin	2017		√	√	
[9]	Miltiadis Kandias	2017	√		√	
[10]	Mike Thelwall	2017	√		√	
[11]	Hong-Han Shuai	2017	√		√	
[12]	Huijie Lin	2014		√	√	
[13]	Qi Li	2017		√	√	
[14]	Sho Tsugawa	2015		√		√
[15]	Glen A. Coppersmith	2014	√		√	
[16]	Minsu Park	2012	√		√	
[17]	Shirin Saleem	2012	√		√	
[18]	Munmun De Choudhury	2014	√		√	
[19]	Bridianne O'Dea	2015	√		√	

The researcher believes that comparing the machine learning algorithm will improve the quality of mental health detection. Most researchers have selected any of the available machine learning techniques and applied it in the proposed approaches or algorithms [27]. Support Vector Machine is one of the techniques in machine learning algorithm that is commonly used on mental health detection [4], [9]–[12], [14], [17]–[19].

Table 3 shows a summary of the OSN used in data crawling based on the literature. OSN is an online platform which people use to build social networks or social relations with others. Previous research used a few types of OSN such Facebook, Twitter and Weibo as data sources in mental health detection. This type of OSN example is mentioned in Table 3.

Most of the researchers extracted the data from OSN and classified the contents using machine learning algorithms. Twitter is one of the popular OSN that is widely used as the data source and covers a considerable scale of data [4], [10], [14]–[16], [18], [19]. Furthermore, Facebook and Weibo are also used as data sources in mental health detection [9], [12], [13].

Table 4 shows a summary of the language and data set collected based on the literature. Generally, most studies used Twitter as the main data source and English as the main language in mental health detection. English is the most common language used in harvesting data using keywords. The data collection normally crawled in certain duration based on the keywords. Most researchers prefer proposing new data set in mental health detection [4], [9], [19], [10]–[13], [15]–[18]. Generally, a few elements need to be taken

into consideration in the study of mental health detection in OSN. For example, the type of the OSN, language of the data source and types of machine learning algorithm. The analysis of these general elements is important in early mental health detection.

B. Challenges

Generally, there are difficulties involved in determining mental health in OSN due to a few issues in non- face-to-face communication and human-computer interaction. The main difficulty is usually from the language barrier that commonly happened while determining the exact meaning of mental health behind the words and language written in the OSN [4], [12]–[14]. However, there are several ways to solve this issue, and one of them is through machine learning to learn, understand and determine the possibility of mental health behind the words and language written in OSN.

Other than that, the account privacy policy that is currently being enforced in most OSN service providers has also made it difficult for most researchers to extract data from OSN [9]. It is recommended that the researchers fully understand the ethical code of conduct before collecting data from OSN and apply good practice in research by sending a permission request to the users and OSN providers.

Most previous researchers used the single type of OSN as the source of data extraction. They also used text analysis in OSN for mental health detection. It is difficult to interpret the mental health detection as the OSN users nowadays usually use other types of data to express themselves in OSN. Some of them introduce features selection which is limited to their current study. A good feature selection will improve the performance of learning algorithms [28], [29], [30].

C. Future Direction

The researcher believes that the challenges outlined are possible to become future direction of this research. The language barrier issue and extracting other languages from the OSN form the potential future direction of this research based on users' geo-location and mother tongue. This is because many OSN users use their native languages for texting than using the only the English language in OSN.

Furthermore, the researchers also believe that the use of pictorial, audio and video formats in OSN could become one of the potential areas to be explored in future research. The vastly rich contents available in OSN should not be overlooked as they could give more value to the data that are explored by the researchers in the future.

The researcher believes that extracting data from multiple sources of OSNs will improve the results of mental health detection. The researchers will observe any possibility to extract data from other OSN service providers, other than the common OSN used in the previous research. This may lead the researchers to new features of the selection process before implementing the machine learning algorithms in mental health detection.

IV. CONCLUSION

This research is a literature survey of mental health detection using OSN that is widely studied in the recent years. Many techniques in the literature survey showing the possibility of mental health detection using OSN are viable and feasible to be identified quickly using more general techniques. The literature survey also found the possibility of mental health detection that is not only limited to one language such as English, as many researchers are using other languages as well. Most of the works available in the literature also used Twitter as their OSN platform for data collection instead of other types of OSN. Therefore, it can be concluded that this research work has a big potential in the early detection of mental health.

REFERENCES

[1] World Health Organization, *The World Health Report 2001*, vol. 36, no. 10. WHO, 2001.

[2] Global Burden of Disease Study 2013 Collaborators, "Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013," *Lancet*, vol. 386, no. 9995, pp. 743–800, 2015.

[3] Institute for Public Health (IPH) 2017, *National Health and Morbidity Survey 2017 (NHMS 2017): Adolescent Mental Health (DASS-21)*. 2011.

[4] H. Lin, J. Jia, J. Qiu, Y. Zhang, G. Shen, L. Xie, J. Tang, L. Feng, and T. S. Chua, "Detecting stress based on social interactions in social networks," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 9, pp. 1820–1833, 2017.

[5] T. G. M. Vrijkotte, L. J. P. van Doornen, and E. J. C. de Geus, "Effects of Work Stress on Ambulatory Blood Pressure, Heart Rate, and Heart Rate Variability," *Hypertension*, vol. 35, no. 4, pp. 880–886, 2000.

[6] F. Sun, C. Kuo, H. Cheng, S. Buthpitiya, P. Collins, and M. Griss, "Activity-aware Mental Stress Detection Using Physiological Sensors," *Mob. Comput. Appl. Serv.*, vol. 76, pp. 1–20, 2012.

[7] E. Garcia-Ceja, V. Osmani, and O. Mayora, "Automatic Stress Detection in Working Environments from Smartphones' Accelerometer Data: A First Step," *IEEE J. Biomed. Heal. Informatics*, vol. 20, no. 4, pp. 1053–1060, 2016.

[8] A. Sano and R. W. Picard, "Stress Recognition Using Wearable Sensors and Mobile Phones," *2013 Hum. Assoc. Conf. Affect. Comput. Intell. Interact.*, pp. 671–676, 2013.

[9] M. Kandias, D. Gritzalis, V. Stavrou, and K. Nikoloulis, "Stress level detection via OSN usage pattern and chronicity analysis: An OSINT threat intelligence module," *Comput. Secur.*, vol. 69, pp. 3–17, 2017.

[10] M. Thelwall, "TensiStrength: Stress and relaxation magnitude detection for social media texts," *Inf. Process. Manag.*, vol. 53, no. 1, pp. 106–121, 2017.

[11] H.-H. Shuai, C.-Y. Shen, D.-N. Yang, Y.-F. Lan, W.-C. Lee, P. Yu, and M.-S. Chen, "A Comprehensive Study on Social Network Mental Disorders Detection via Online Social Media Mining," *IEEE Trans. Knowl. Data Eng.*, vol. 4347, no. c, pp. 1–1, 2017.

[12] H. Lin, J. Jia, Q. Guo, Y. Xue, J. Huang, L. Cai, and L. Feng, "Psychological stress detection from cross-media microblog data using Deep Sparse Neural Network," *Proc. - IEEE Int. Conf. Multimedia. Expo*, vol. 2014–Sept, no. September, pp. 1–6, 2014.

[13] Q. Li, Y. Xue, L. Zhao, J. Jia, and L. Feng, "Analyzing and Identifying Teens' Stressful Periods and Stressor Events from a Microblog," *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 5, pp. 1434–1448, 2017.

[14] S. Tsugawa, Y. Kikuchi, F. Kishino, K. Nakajima, Y. Itoh, and H. Ohsaki, "Recognizing Depression from Twitter Activity," *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst. - CHI '15*, pp. 3187–3196, 2015.

[15] G. A. Coppersmith, C. T. Harman, and M. H. Dredze, "Measuring Post Traumatic Stress Disorder in Twitter," *Proc. 7th Int. AAAI Conf. Weblogs Soc. Media (ICWSM).*, vol. 2, no. 1, pp. 23–45, 2014.

[16] Park Minsu, Chiyoung Cha, and Meeyoung Cha, "Depressive Moods of Users Portrayed in Twitter," *ACM SIGKDD Work. Healthc. informatics*, pp. 1–8, 2012.

[17] S. Saleem, R. Prasad, S. Vitaladevuni, M. Pacula, M. Crystal, B. Marx, D. Sloan, J. Vasterling, and T. Speroff, "Automatic Detection of Psychological Distress Indicators and Severity Assessment from Online Forum Posts," *Proc. COLING 2012*, vol. 5, no. December 2012, pp. 2375–2388, 2012.

[18] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting Depression via Social Media.," *Icwsm*, vol. 2, pp. 128–137, 2013.

[19] B. O'Dea, S. Wan, P. J. Batterham, A. L. Calear, C. Paris, and H. Christensen, "Detecting suicidality on Twitter," *Internet Interv.*, vol. 2, no. 2, pp. 183–188, 2015.

[20] M. K. Othman and M. S. N. M. Danuri, "Proposed conceptual framework of Dengue Active Surveillance System (DASS) in Malaysia," *ICICTM 2016 - Proc. 1st Int. Conf. Inf. Commun. Technol.*, no. May, pp. 90–96, 2017.

[21] R. Myung, H. Yu, and D. Lee, "Optimizing Parallelism of Big Data Analytics at Distributed Computing System," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 5, p. 1716, 2017.

[22] M. K. M. Nasution and S. A. M. Noah, "Extracting Social Networks from Web Documents (Extended Abstract)," *1 st Natl. Dr. Semin. Artif. Intell. Technol.*, pp. 278–281, 2010.

[23] M. A. Al-Garadi, K. D. Varathan, and S. D. Ravana, "Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network," *Comput. Human Behav.*, vol. 63, pp. 433–443, 2016.

[24] F. Nasution, N. E. Nazira Bazin, - Daliyusmanto, and A. Zulfikar, "Big Data's Tools for Internet Data Analytics: Modelling of System Dynamics," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 3, p. 745, 2017.

[25] K. C. Laudon and J. P. Laudon, *Management information systems: managing the digital firm*, 4th ed. New York, 2012.

[26] R. M. Stair and G. W. Reynolds, *Principles of Information Systems*, 13th ed. Boston: Cengage Learning, 2018.

[27] J. Ghazali, S. A. Noah, and L. Zakaria, "Classification of images for automatic textual annotation: A review of techniques," *Journal of Applied Sciences*, vol. 13, no. 6. pp. 760–767, 2013.

[28] R. Alhutaish and N. Omar, "Feature Selection for Multi-label Document Based on Wrapper Approach through Class Association Rules," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 2, p. 642, 2017.

[29] H. Alshalabi, S. Tiun, N. Omar, and M. Albared, "Experiments on the Use of Feature Selection and Machine Learning Methods in Automatic Malay Text Categorization," *Procedia Technol.*, vol. 11, no. Iceei, pp. 748–754, 2013.

[30] A. Alsaffar and N. Omar, "Study on feature selection and machine learning algorithms for Malay sentiment classification," *Proc. 6th Int. Conf. Inf. Technol. Multimedia.*, no. November, pp. 270–275, 2014.