IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 March 2024, accepted 4 May 2024, date of publication 16 May 2024, date of current version 24 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3401704

RESEARCH ARTICLE

Probability-Based Multi-Label Classification Considering Correlation Between Labels–Focusing on DSM-5 Depressive Disorder Diagnostic Criteria

DABIN PARK[1], **GEONJU LEE**[2], **SEONHYEONG KIM**[2], **TAEWOONG SEO**[1], **HAYOUNG OH**[3], **AND SEOG JU KIM**[4]

[1]Department of Applied Artificial Intelligence, Sungkyunkwan University, Seoul 03063, South Korea
[2]Department of Mathematics, Sungkyunkwan University, Seoul 03063, South Korea
[3]Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, Seoul 03063, South Korea
[4]Department of Psychiatry, Sungkyunkwan University School of Medicine, Samsung Medical Center, Seoul 03063, South Korea

Corresponding author: Hayoung Oh (hyoh79@gmail.com)

This work was supported in part by the Culture, Sports and Tourism Research and Development Program through Korea Creative Content Agency Grant funded by the Ministry of Culture, Sports and Tourism (Project Name: Development of Game-Based Digital Therapeutics Technology for Adolescent Mental Health (Psychological and Behavioral Control) Management) under Project RS-2024-00344893, Contribution Rate: 25%; in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government, Ministry of Science and ICT (MSIT), South Korea, under Grant NRF-2022R1F1A1074696; in part by the MSIT through the Graduate School of Metaverse Convergence Support Program, supervised by the Institute for Information and Communications Technology Planning and Evaluation (IITP), under Grant IITP-2024-RS-2023-00254129; in part by the BK21 FOUR Project (Bigdata Research and Education Group for Enhancing Social Connectedness Thorough Advanced Data Technology and Interaction Science Research) funded by the Ministry of Education (MOE), South Korea, and NRF under Grant 5199990913845.

ABSTRACT The incidence of depressive disorder in Korea is the highest among OECD countries. The proportion of patients in their 20s is the highest. However, social gaze and false perception are causing problems such as not visiting the hospital or delaying the visit. Accordingly, we suggest a Korean model for predicting depressive disorders using data from online communities widely used by people in their 20s. In many countries, including South Korea, depressive disorders are diagnosed using DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) published by the American Psychiatric Association. We propose a model that predicts the probability of a user's speech corresponding to nine criteria for diagnosing DSM-5 depressive disorder, following advice obtained through periodic meetings with a psychiatrist. The prediction performance was improved by using the correlation between each criterion in the model implementation stage.

INDEX TERMS Artificial intelligence, correlation, depressive disorder, multi-label classification, natural language processing, psychiatry in AI.

I. INTRODUCTION

The proportion of patients with depressive disorders in Korea, which can be officially confirmed through literature [1] is 36.8%, the highest in OECD countries. In particular, the proportion of patients with depressive disorder in their 20s accounts for 19% of all patients, and data from the Health Insurance Review and Assessment Service showed that the number of patients with depressive disorder in their 20s increased 127.1% in 2021 compared to 2017. Patients in their 20s account for the largest proportion of all age groups, with 177,166 people.

DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) is a manual published by the American Psychiatric Association that provides a standard classification system for mental illness diagnosis and statistical characteristics. Korea, as well as many other countries, use DSM-5 diagnostic criteria to diagnose depressive disorders. However, many existing studies predicting depressive disorder through social media are not based on DSM-5 depressive disorder

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

diagnostic criteria [2], [3], [4], [5], [6]. The criteria for DSM-5 depressive disorder are as follows: 1. Depression mood, 2. decreased interest or enjoyment, 3. weight change, 4. insomnia or excessive sleep, 5. nervousness or delay, 6. fatigue or loss of vitality, 7. feelings of worthlessness and guilt, 8. decreased thinking and concentration, 9. suicidal thoughts

Through periodic meetings with a psychiatrist, we were advised that research on multi-labeling rather than single labeling should be conducted because one sentence does not contain only one emotion. In addition, we were advised that it would be more medically reasonable to find the probability that the speech sentence corresponds to each label, not that it corresponds to or not. Therefore, in this study, data obtained from Everytime, an online community frequently used by college students, are used to get the probability of meeting the DSM-5 depressive disorder diagnostic criteria. The nine criteria for diagnosing DSM-5 depressive disorder are related. For example, a person who feels depressed may have dietary problems, and problems such as insomnia or over sleep may occur [2]. In accordance with the additional advice of a psychiatrist that the higher the probability of one criterion, the higher the probability of meeting another criterion, the more the correlation between labels was introduced to conduct the study. The contributions of our proposed study are as follows:

- We created a more medically reasonable model by finding the probability of each label corresponding to each label, rather than simply labeling it as 1 and 0 whether it corresponds to each label.
- Unlike traditional approaches such as PHQ-9, where patients directly respond to surveys, or black-box AI models that extract features from text to predict depression, our research holds significance as we utilize the DSM-5 diagnostic framework.
- By incorporating label correlations, we developed a more rational model, contributing to the field by achieving more effective results.

II. PREVIOUS WORKS

Reece et al. [3] proposed a method for predicting the onset and progression of mental illness using features extracted from Twitter data. As a result of the study, it was confirmed that the predictive model is effective in predicting the onset of mental illness 2 weeks before and 1 week before progression. This suggests that Twitter data is a useful tool that can be used to prevent and treat mental illness. In addition, the results of the study suggested new possibilities for the prevention and treatment of mental illness, and if Twitter data can be used to predict the onset and progression of mental illness early, it is expected that the patient's treatment will be more effective.

Ghosh & Anwar [4] presented a deep learning-based approach to estimating the intensity of depression using social media data. The intensity of depression was predicted using various functions such as emotional, subject, behavioral, user level, and n-gram related to depression. A small LSTM network was used to predict the intensity of depression.

Cha et al. [6] presented a text-based approach to depression detection on Twitter and in the university community. A dictionary-based method was used to identify words and phrases related to depression, and cross-validation was used to evaluate the performance of the depression detection model.

Farruque et al. [7] presented a multi-label classification approach to identify basic and depression-related emotions on Twitter. Basic and depression-related emotions were identified using various functions such as textual features, language models, emotion dictionaries, and depression dictionaries. Various machine learning algorithms for multi-label classification were evaluated. However, multi-label classification was not performed based on the DSM-5 depressive disorder diagnostic criteria.

Ameer et al. [8] presented a transfer learning approach for multi-label emotion classification in the text. We initialize the emotion classification model using a large pre-trained language model and improve the performance of the model by adding subsequent layers specialized in emotion classification.

Label imbalance and label correlation are problems that can reduce learning efficiency in multi-label classification. Chen et al. [9] showed that it is an effective way to improve learning efficiency by solving these problems through label grouping. However, there is a limitation in that the effect may vary depending on the dataset, and the label grouping method may be complex, resulting in poor practicality.

The KRF framework proposed by Ma et al. [10] performs better than conventional neural network-based models and multi-label classification methods, and significantly improves the original error and F1 score of music style classification. The framework was evaluated on two real-world datasets, Duban Music and Amazon Music, and results showed that KRF improved the micro-F1 score of the best baseline method on the Duban dataset from 64.5% to 70.8%, and from 69.1% to 73.2% on the Amazon dataset. The contribution of this paper is to capture complete style correlations by incorporating statistical relationships between external music knowledge and style labels. It is emphasized that using this method can automatically label multiple styles for all songs with higher accuracy.

Liwen et al. [11] We discussed a wide range of applications for multi-label learning, including text processing and image mining. In addition, it emphasized the need to improve classification accuracy while shortening the task and learning time raised by the curses of the dimension. It provides an overview of the existing multi-label feature selection algorithm and said that considering label correlation, performance can be improved. The authors noted that many researchers are currently working on multi-label learning and achieve important results in fields such as computer vision and natural language processing.

Che et al. [12] proposed a novel method to improve label correlation in multi-label classification. This method efficiently reflected label correlation using local property

TABLE 1. Comparison table of the previous works.

title	published	characteristic	dataset	limitations
Forecasting the onset and course of mental illness with Twitter data [3]	2017	- Detected depression and PTSD with Twitter posts - Predicted onset and course of depression and PTSD using Hidden Markov Model	- Survey of CES-D, TSQ responses and Twitter data through Amazon Mechanical Turk portal	- The results are limited to only the Twitter users who have been diagnosed with depression or PTSD, and are willing to provide their social media history.
Depression Intensity Estimation via Social Media: A Deep Learning Approach [4]	2021	- presented a deep learning-based approach to estimate the intensity of depression using social media data. - Various functions such as emotional, subject, behavioral, user level, and n-gram related to depression were used. - A small LSTM network was used to predict the intensity of depression.	- curated by Shen et al. [5] - 6,562 users(1,402 depressed and 5,160 nondepressed), 4,245,747 tweets(292,564 depressed and 3,953,183 nondepressed)	- Weak labeling can make it difficult to accurately measure the intensity of depression. -Deep learning models can be sensitive to bias and noise in data.
A lexicon-based approach to examine depression detection in social media: the case of Twitter and university community [6]	2022	- presented a text-based approach to depression detection on Twitter and in the university community. - Dictionary-based methods were used to identify words and phrases related to depression. - Cross-validation were used to evaluate the performance of the depression detection model.	- Crawling dataset of Twitter and online community of Korean university	- Dictionary-based methods may not identify all words and phrases related to depression. -Twitter datasets may not guarantee representation for depression detection.
Basic and Depression Specific Emotion Identification in Tweets: Multi-label Classification Experiments [7]	2021	- conducted experiments on two new methods for multi-label classification: a cost-sensitive RankSVM model and a deep learning model based on a Bi-LSTM with a self-attention mechanism using the softmax function. - evaluated the performance of the models using Macro-FM and Micro-FM metrics.	- 2 multi-labeled Twitter datasets from "Clean and Balanced Emotion Tweets" (CBET) - Set 1: 31,303 tweets containing 3,000 tweets with 9 emotions and 4,303 tweets with 2 labels of 9 emotions - Set 2: total 50,000 tweets adding 7 emotions Tweets to Set 1	-There is no clear evidence that the proposed model is better than other models.
Multi-label emotion classification in texts using transfer learning [8]	2023	- proposed deep learning methods for multi-label emotion classification, such as LSTM, Bi-LSTM, and Bi-LSTM with single attention. - used transfer learning models, such as XLNet-MA, DistillBERT-MA, and RoBERTa-MA	- English Twitter dataset, Chinese blogs dataset	- The proposed model did not account for the relationships between phrases and emotions. - There was not enough data to train a model to accurately identify emotions such as optimism and disgust. -The data used to train the model was from social media, which means that the text was short and may contain misspelled words and missing punctuation.
Enhancement of DNN-based multilabel classification by grouping labels based on data imbalance and label correlation [9]	2022	- Learning efficiency was improved by grouping labels in consideration of label imbalance and label correlation. - It evaluated the effectiveness of label grouping using a DNN-based multi-label classification model.	- A text dataset containing documents for worldwide river restoration projects	- The effect of label grouping may vary depending on the dataset. -Label grouping methods may be complicated and may be less practical.
Beyond Statistical Relations: Integrating Knowledge Relations into Style Correlations for Multi-Label Music Style Classification [10]	2020	- A complete style correlation was captured by integrating knowledge and statistical relations. - The correlation between music styles was identified by using external music knowledge. - Music style classification was performed by fusion of review expressions and style expressions.	- Dataset of Douban Music, the most popular music review website in China, and dataset of Amazon Music in the United States	- needed to cover more datasets and more diverse music styles. - Utilizing external knowledge can help increase accuracy, but results can vary greatly depending on the accuracy of external knowledge. - Because the style of music was classified using only reviews and style information in this paper, other aspects of music were not considered.

TABLE 1. ***(Continued.)*** **Comparison table of the previous works.**

Multi-label Learning By exploiting Correlations of Label Subsets [11]	2021	- dealt with multi-label learning - proposed a feature selection algorithm considering label correlation to improve the performance of Multi-label Learning.		-Only the method was presented, but the experiment was not conducted with the actual dataset.
Label correlation in multi-label classification using local attribute reductions with fuzzy rough sets [12]	2021	- explored new ways to improve label correlation in multi-label classification using local property reduction and fuzzy rough sets. -proposed a method to efficiently reflect the label correlation by dividing the entire label set into several mutually exclusive label-related subsets consisting of overall label correlations and regional label correlation.	-15 multi-label datasets coming from Mulan Library, that are Reuters, Yeast, Image, Birds, Cal500, Recreation, Computer, Reference, Artificial, Business, Education, Health, Society, Medical and Bookmark	- More experimental verification is needed of how effective the new methods used to improve label correlation are compared to other methods. - Since the datasets used in this paper were not collected in various fields, it is necessary to verify whether this method is applicable to other fields. - need to study whether the method used to improve label correlation is applicable to other multi-label classification problems. - the method used to improve label correlation has a high computational cost.
A novel approach for learning label correlation with application to feature selection of multi-label data [13]	2019	- explored the importance of label correlation in multi-label learning and proposes ways to optimize it. - proposed a new algorithm, correlation-labels-specific features (CLSF), which is different from the existing algorithm. This algorithm reconstructed the binary relationship by incorporating labels within the relevant label group, taking into account label correlation and classification error rate.	- Emotion, Birds, Science, Yeast, Reuters, Recreation, Computer, Reference, Medical, Enron, Cal500 dataset from Mulan Library and Yahoo	- need to perform more datasets and more diverse experiments to obtain more general results. - the paper considered label correlations within label groups, but not between label groups.

reduction and fuzzy rough set. In this paper, experiments were performed using 15 multi-label datasets, and a novel method called LRFS-α was proposed and compared to conventional multi-label methods. In this paper, we need to study whether the method used to improve label correlation is applicable to other multi-label classification problems, and the high computational cost is mentioned as a limitation.

Che et al. [13] explored the importance of label correlation in multi-label learning and proposed methods to optimize it. In the paper, we propose a new algorithm, CLSF, which reconstructs binary relations by incorporating labels within related label groups, taking into account label correlations and classification error rates. This has improved the effectiveness and accuracy of classification. In this paper, we evaluated CLSF and six other multi-label classification algorithms using 11 benchmark datasets. As a result of the experiment, CLSF performed better than other algorithms. However, there is a limitation in this paper that more datasets and more diverse experiments need to be conducted to obtain more general results.

III. RELATED TECHNIQUES

A. WORD2VEC

Word2Vec is a widely used technology in the field of natural language processing and is a method of converting words into vectors. Information on the meaning of words and their surrounding relationships is calculated and expressed in vector form. It is a way to reflect the relationship, similarity, and context between words by being able to express them numerically. The Word2Vec model is learned by predicting words by considering the context of surrounding words, and there are two main ways: CBOW, Skip-gram. Continuous Bag of Words (CBOW) is a method of predicting target words based on the context of surrounding words. Skip-gram is a semi-c, a method of predicting surrounding words from target words. After the Word2Vec model is learned, it can be used by calculating the semantic similarity between words or analyzing the semantic relationship between words. It can also be used for natural language processing tasks that generate natural sentences in a given context.

B. PARAGRAPH VECTOR

Paragraph Vector, PV is a similar technique to Word2Vec, but it is a method of vectorizing sentences as well as words. Word2Vec uses the context of the surrounding words to represent each word as a vector, but PV maps that paragraph or sentence into a vector, taking into account the entire context of a given paragraph or sentence. You can use PV to calculate document similarity or classify the subject of a document. You can also generate a response to the query.

C. KoBERT

KoBERT is a language model developed by SKT Brain for Korean natural language processing and is based on the BERT (Bidirectional Encoder Representations Transformers) architecture. BERT is a natural language processing model that understands text in both directions and shows excellent performance. KoBERT is a pre-trained model with Korean data and shows excellent performance in understanding the meaning of Korean texts. KoBERT can be used for a wide variety of natural language processing tasks in Korean.

IV. DATASET

In this study, based on depressive disorder representation data, wellness conversation script data sets and everytime crawling data were multi-label classified and used for KoBERT training. The data originates from Korean communities and is in the Korean language. No information was provided to distinguish between patients or users.

A. DEPRESSION DISORDER REPRESENTATION DATA

The expression data related to depressive disorder is data provided by Gangbuk Samsung Hospital, which summarizes 388 Korean expressions for nine conditions of depressive disorder diagnosis criteria defined by DSM-5. This is the data that a psychiatrist personally organized and delivered expressions often used by patients with depressive disorders. Nine states refer to depressed moods, decreased interest or pleasure, weight and appetite changes, sleep changes, mental arousal and retardation, fatigue or energy loss, self-criticism and worthlessness or guilt, reduced thinking and concentration and difficulty deciding, repeated thoughts about death and suicidal thoughts or plans. Table 2 compiles examples from this dataset.

B. WELLNESS CONVERSATION SCRIPT DATASET

The wellness conversation script dataset is data provided by Gangnam Severance, which extracts 4,200 out of 16,000 counseling data, separates them by sentence, and classifies them by conversation intention. The data can be downloaded from the AI Hub. Only conversational text data were provided.

C. EVERYTIME CRAWLING DATA

Everytime is a university online community targeting 400 universities nationwide, where anonymity is guaranteed so that university students can share their opinions honestly and is used by most Korean university students. Everytime crawling data is the data that crawls through everytime posts at Hankuk University of Foreign Studies. From March 22, 2018, to January 25, 2023, 1,852 posts were crawled on the depression bulletin board and 38,585 posts on the free bulletin board, and 14,334 posts were crawled on the free bulletin board, including expressions related to depression disorder.

TABLE 2. Examples of expressions.

Criteria	examples
1. Depressed mood	Hopeless; worthless; stuffy; painful; abyss; discouraged; depressed; somber
2. Loss of interest / pleasure	Tiresome; quit; get tired of; give up; do not want to go out; tedious; tough
3. Weight loss or gain	Starve; fat; tasteless; lazy to eat; indigestion; gluttony; binge eating
4. Insomnia or hypersomnia	do not sleep well; toss and turn; sleep like a newborn; Even if I sleep, I don't feel refreshed; I just sleep
5. Psychomotor agitation or retardation	Panic; tension; Difficulty in speaking; dazed; helpless; irritation
6. Fatigue	Burnout; want to rest; feel drained; lack of energy; poor condition; feel physically heavy
7. Feeling worthless or excessive / inappropriate guilt	It is because of me; my fault; I hate myself; failure; looser; useless; self-loathing; defeated
8. Decreased concentration	Decision paralysis; heavy-headed; complicated; distracted; zoned out; forgetful
9. Thoughts of death / suicide	Want to stop; want to run away; want to escape; feel worthless; want to feel at ease now; farewell letter; self-harm; death seems better

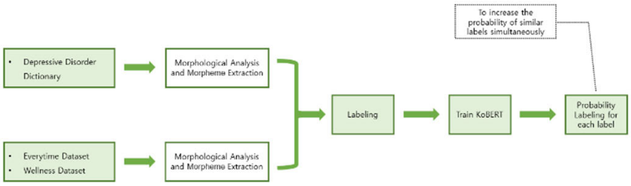
To increase the probability of similar labels simultaneously
Depressive Disorder Dictionary
Morphological Analysis and Morpheme Extraction
Everytime Dataset
Wellness Dataset
Morphological Analysis and Morpheme Extraction
Labeling
Train KoBERT
Probability Labeling for each label

FIGURE 1. Overall process of this study.

V. MODELING

The overall structure is shown in Figure 1.

A. MORPHOLOGICAL ANALYSIS AND MORPHEME EXTRACTION

First, the expression in the depressive disorder expression data (hereinafter referred to as depression dictionary) is analyzed in morpheme to extract only the morpheme that has meaning. After that, each morpheme is assigned a label of

TABLE 3. Instances with labels.

example sentences	Label 1	Label 2	Label 3	Label 4	Label 5	Label 6	Label 7	Label 8	Label 9
I'm so sad and tired.	0.6	0.7	0.1	0.2	0.1	0.2	0.2	0.1	0.2
I'm not motivated since I'm pregnant.	0.4	0.5	0.2	0.1	0.2	0.85	0.2	0.1	0.1
I want to stop.	0.4	0.3	0.1	0.2	0.1	0.6	0.1	0.1	0.9

the criteria for diagnosing depressive disorders to which the existing expression belongs.

Next, after morphological analysis of the speech of the wellness conversation data set and the Everytime data set, only the morpheme that has meaning is extracted.

B. LABELING

When the morpheme extracted from the utterance is compared and matched with the morpheme of the depression dictionary, the label of the expression in the depression dictionary is given to the utterance.

C. CORRELATION

Using the features of the PV model, the correlation between labels is analyzed. By using the PV model, the similarity of the vector can be derived for each label. The cosine similarity between labels is expressed as a value between 0 and 1.

D. KoBERT LEARNING

The similarity obtained through the process above is reflected in the last layer of KoBERT to train KoBERT. We trained KoBERT by splitting the data set at a ratio of 0.75:0.25. The learning parameters were used in the same way except for the only one parameter, epoch, which were used in the Naver review classifications example released by SKT Brain. max_len was set to 128, batch size was set to 64, learning rate was set to 5e-5, and, epoch was adjusted from 5 to 10 to prevent underfitting. Sigmoid was used as an activation function.

VI. RESULT AND EVALUATION

The results of calculating the similarity between labels through PV are as follows. Label 1 has similarities of labels 2 and 0.45, label 5 and 0.4, label 7 and 0.49, and label 9 and 0.3. Label 2 has the similarity between label 6 and 0.3, and label 4 has the similarity between label 6 and 0.35. Label 5 has similarity between label 7 and 0.33, and label 9 and 0.36. Label 7 has a similarity of 0.3 with label 9.

In Table 3, the few instances with labels in our dataset can be identified.

Multi-label classification problems have multiple labels, and data often do not exist balanced. Therefore, in the multi-label classification problem, rather than using accuracy as a performance evaluation metric, the F-1 score is used. After adjusting parameters and conducting various experiments, the model achieved its best performance with a learning rate of 5e-5 and an epoch of 10. Under these settings, the F-1 score for the training set reached a perfect 1.0, while the F-1 score for the test set was an impressive 0.9747, demonstrating excellent performance. This is illustrated in Figure 2.

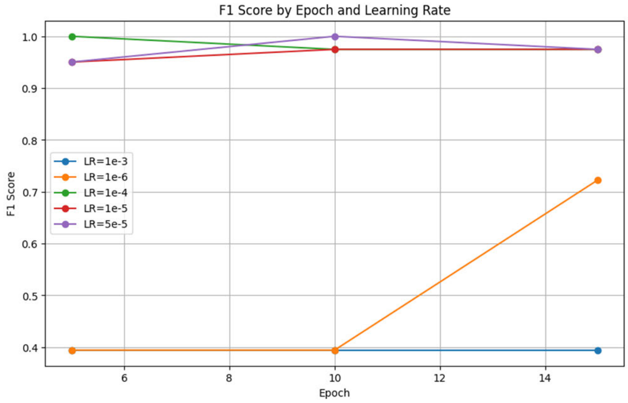

FIGURE 2. F1 Score by epoch and learning rate.

VII. DISCUSSION AND ANALYSIS

By addressing these research questions, filling critical gaps in the literature, and providing a detailed analysis of the methodology and model performance, this research contributes significantly to the field of probability-based multi-label classification for mental health diagnosis.

A. RESEARCH QUESTIONS AND THE ANSWERS

Research Questions of this study are below:

1. How can probability-based multi-label classification be utilized for DSM-5 depressive disorder diagnostic criteria?
2. What is the significance of considering label correlation in the classification process?
3. What are the implications of this research for the field of psychiatry and mental health diagnosis?

The answers to research questions are below:

1. The research demonstrates the application of probability-based multi-label classification to assess the likelihood of meeting the DSM-5 depressive disorder diagnostic criteria. By leveraging data from various sources, including wellness conversation scripts and online community discussions, the study provides a comprehensive approach to understanding and predicting depressive disorder criteria.
2. The consideration of label correlation in the classification process is crucial as it aligns with medical advice to determine the probability of a speech sentence

corresponding to each label, rather than a binary classification. This approach enhances the medical relevance of the classification process and provides a more nuanced understanding of the interrelated nature of depressive disorder criteria.

3. The implications of this research for the field of psychiatry and mental health diagnosis are significant. By incorporating machine learning and deep learning techniques, the study offers a novel approach to understanding and predicting depressive disorder criteria. This has the potential to enhance early detection, personalized treatment, and overall management of depressive disorders.

B. RESEARCH GAPS ADDRESSED

This research fills several critical gaps in the existing literature:

- It addresses the need for a more nuanced approach to multi-label classification, particularly in the context of mental health diagnosis.
- The study incorporates label correlation, which is essential for understanding the complex interplay of depressive disorder criteria, thus bridging a gap in existing methodologies.
- By leveraging data from diverse sources and applying advanced natural language processing techniques, the research offers a comprehensive and interdisciplinary approach to understanding mental health conditions, addressing a gap in traditional psychiatric research methodologies.

C. METHODOLOGY ANALYSIS

The methodology employed in this research demonstrates a robust and interdisciplinary approach to understanding and predicting depressive disorder criteria. By integrating data from wellness conversation scripts, online community discussions, and psychiatric expertise, the study ensures a comprehensive representation of depressive disorder expressions and experiences. Furthermore, the utilization of advanced machine learning and deep learning techniques, such as KoBERT, Word2Vec, and Paragraph Vector, showcases the research's commitment to leveraging cutting-edge technologies for mental health research. The consideration of label correlation within and between label groups further enhances the methodological rigor of the study, ensuring a holistic and medically relevant approach to multi-label classification.

D. SOCIETAL IMPACT

Our study aims to screen and monitor depressive disorders, which can be used as an assistant tool for doctors through multi-label classification based on DSM-5 diagnostic criteria for depressive disorders. This can help manage and treat depressive disorders through early detection and timely intervention. Our research focuses on not only accurate diagnosis, but also supporting doctors who can detect and take action against early symptoms of depressive disorders. This can minimize negative effects from depressive disorders and improve an individual's quality of life. At the same time, our study highlights that it can also be helpful for people who lack an understanding of depressive disorders or are afraid to visit hospitals or counseling centers. By allowing them to recognize their condition early and take appropriate measures without the need for a direct visit from a medical professional, they can help with early detection and management of depressive disorders. This is expected to allow more individuals to overcome difficulties associated with depressive disorders and eventually have a positive social impact

VIII. CONCLUSION

In Korea, the incidence of depressive disorders among people in their 20s is becoming a serious social problem. Through this study, we presented a model to predict depressive disorders using data from online communities frequently used by university students in their 20s. A model was proposed in which the entered utterance returns the probability value to the nine criteria for diagnosing DSM-5 depressive disorder after periodically meeting with a psychiatrist professor. The prediction performance was further improved by using the correlation between labels. We expect to contribute to early detection and treatment of depressive disorders through data from online communities or social media. There are several limitations to our study. The limitations are as follows:

- The limitation of our proposed model is that it cannot completely replace the doctor. Our model is an assistive tool that helps doctors diagnose or treat patients, such as recommending users visit hospitals or counseling centers, or informing doctors of patient symptoms. It is essential to be diagnosed by a doctor.
- There were fewer seed words. We think that with more seed words, we could have achieved higher quality results.
- We were unable to collect text for each user, so we could not confirm whether the symptoms lasted more than two weeks for each user.

We proposed a model for detecting depression in text based on the widely used diagnostic tool in the medical field, DSM-5, with the aim of minimizing the efforts of physicians in diagnosing depressive disorders. We found that there was a correlation between the nine diagnostic criteria for DSM-5 depressive disorder and applied them to our study. We hope that to further improve the performance of the model, studies that utilize more data and apply different machine learning algorithms will emerge. If the research is conducted by augmenting the depressive disorder expression dataset in the future, a more practical model will be made. We made a model specialized for Korean. It is hoped that it will be developed in other languages so that diverse language speakers can also promote the maintenance of their mental health. We believe that our proposed modeling method could be applied to the development of similar models in other languages, provided that training data in the languages of other countries are sufficient. We could not confirm whether each user has depressive symptoms lasting more than two weeks. In the future, we hope to collect data from each user and conduct research closer to actual medical diagnosis. If our proposed model is used in a depressive disorder counseling chatbot, it is expected that screening will be possible for a long cycle between counseling and counseling, so that better treatment can be proceeded.

REFERENCES

[1] *Organisation for Economic Co-operation and Development, A New Benchmark for Mental Health Systems: Tackling the Social and Economic Costs of Mental Ill-Health*, OECD Publishing, Paris, France, 2021.

[2] Y. Sun, Z. Fu, Q. Bo, Z. Mao, X. Ma, and C. Wang, "The reliability and validity of PHQ-9 in patients with major depressive disorder in psychiatric hospital," *BMC Psychiatry*, vol. 20, no. 1, p. 474, Dec. 2020, doi: 10.1186/s12888-020-02885-6.

[3] A. G. Reece, A. J. Reagan, K. L. M. Lix, P. S. Dodds, C. M. Danforth, and E. J. Langer, "Forecasting the onset and course of mental illness with Twitter data," *Sci. Rep.*, vol. 7, no. 1, p. 13006, Oct. 2017, doi: 10.1038/s41598-017-12961-9.

[4] S. Ghosh and T. Anwar, "Depression intensity estimation via social media: A deep learning approach," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 6, pp. 1465–1474, Dec. 2021, doi: 10.1109/TCSS.2021.3084154.

[5] 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 *Proc. 26th Int. Joint Conf. Artif. Intell.* Melbourne, VIC, Australia: International Joint Conferences on Artificial Intelligence Organization, Aug. 2017, pp. 3838–3844, doi: 10.24963/ijcai.2017/536.

[6] J. Cha, S. Kim, and E. Park, "A lexicon-based approach to examine depression detection in social media: The case of Twitter and university community," *Humanities Social Sci. Commun.*, vol. 9, no. 1, p. 325, Sep. 2022, doi: 10.1057/s41599-022-01313-2.

[7] N. Farruque, C. Huang, O. Zaiane, and R. Goebel, "Basic and depression specific emotion identification in tweets: Multi-label classification experiments," 2021, *arXiv:2105.12364*.

[8] I. Ameer, N. Bölücü, M. H. F. Siddiqui, B. Can, G. Sidorov, and A. Gelbukh, "Multi-label emotion classification in texts using transfer learning," *Exp. Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 118534, doi: 10.1016/j.eswa.2022.118534.

[9] L. Chen, Y. Wang, and H. Li, "Enhancement of DNN-based multilabel classification by grouping labels based on data imbalance and label correlation," *Pattern Recognit.*, vol. 132, Dec. 2022, Art. no. 108964, doi: 10.1016/j.patcog.2022.108964.

[10] Q. Ma, C. Yuan, W. Zhou, J. Han, and S. Hu, "Beyond statistical relations: Integrating knowledge relations into style correlations for multi-label music style classification," in *Proc. 13th Int. Conf. Web Search Data Mining*. Houston, TX, USA: ACM, Jan. 2020, pp. 411–419, doi: 10.1145/3336191.3371838.

[11] P. Liwen, Z. Xiaolin, and Z. Yun, "Multi-label learning by exploiting correlations of label subsets," in *Proc. 9th Int. Conf. Inf. Technology: IoT Smart City*. Guangzhou, China: ACM, Dec. 2021, pp. 203–206, doi: 10.1145/3512576.3512613.

[12] X. Che, D. Chen, and J. Mi, "Label correlation in multi-label classification using local attribute reductions with fuzzy rough sets," *Fuzzy Sets Syst.*, vol. 426, pp. 121–144, Jan. 2022, doi: 10.1016/j.fss.2021.03.016.

[13] X. Che, D. Chen, and J. Mi, "A novel approach for learning label correlation with application to feature selection of multi-label data," *Inf. Sci.*, vol. 512, pp. 795–812, Feb. 2020.

DABIN PARK is currently pursuing the degree in applied artificial intelligence with Sungkyunkwan University, South Korea. Her research interests include data analysis, social network analysis, and natural language processing techniques, especially those in psychiatry using machine learning and deep learning.

GEONJU LEE is currently pursuing the bachelor's degree in mathematics with Sungkyunkwan University, South Korea. Her research interests include natural language processing (NLP) and psychiatry using machine learning and deep learning.

SEONHYEONG KIM is currently pursuing the bachelor's degree with the Department of Mathematics, Sungkyunkwan University, South Korea. Her research interests include data analysis, machine learning, and natural language processing techniques.

TAEWOONG SEO is currently pursuing the bachelor's degree in applied artificial intelligence with Sungkyunkwan University, South Korea. His research interests include AI, NLP, computer vision, and metaverse.

HAYOUNG OH received the Ph.D. degree in computer science from Seoul National University, Seoul, Republic of Korea, in 2013.

She was a Visiting Scholar with the University of California, Berkeley, in 2010. Her major is in artificial intelligence with big data analysis. From 2001 to 2004, she was a Researcher and a Developer with the Institute of Shinhan Financial Group, Seoul, Republic of Korea, and an Assistant Professor of computer science with Soongsil University and Ajou University, from 2014 to 2019. Since 2020, she has been an Associate Professor with the College of Computing and Informatics and the Department of Human-Artificial Intelligence, Sungkyunkwan University. Her research interests include social network analysis, recommender systems, spam detection, and natural language processing techniques using machine learning and big data analysis.

SEOG JU KIM received the Ph.D. degree in psychiatry from Seoul National University, Seoul, Republic of Korea, in 2006. He has an extensive professional background, having served as a Professor with Sungkyunkwan University School of Medicine and Samsung Seoul Hospital, since 2015. In 2017, he assumed the role of the Director of the Psychiatry Department, Samsung Seoul Hospital's Mental Health Clinic, until March 2023. Previously, he was an Associate Professor with Seoul National University College of Medicine and Seoul National University Hospital, from 2011 to 2015. His earlier experiences include roles as an Associate Professor with Gachon University Gil Medical Center, from March 2005 to August 2011, and various positions at Seoul National University Hospital, including Clinical Instructor, Psychiatry Specialist, and Intern, spanning from 1997 to 2002. His areas of expertise in patient care include insomnia, stress/trauma, anxiety disorders, panic disorders, and depression.