

The Application of Natural Language Processing in the Detection and Intervention of Mental Health Fields

Yijun Zhuang *

Maynooth International Engineering College, Fuzhou University, Fuzhou, Fujian, China

* Corresponding Author Email: 832401130@fzu.edu.cn

Abstract. Mental health disorders have become a prominent public health problem worldwide. Traditional mental health services have many deficiencies in diagnostic criteria, resource allocation, and other aspects. The development of Natural Language Processing (NLP) technology offers solutions to solve these challenges. This article conducts a systematic review of the application of NLP in the field of mental health over the past five years. Taking text data as the core research object, it sorts out the application situations such as disease detection and special scenario analysis. It compares the performance and applicable scenarios of three technical routes: rule-based, traditional machine learning, and deep learning. Point out the shortcomings of the current research in terms of language diversity, clinical adaptability, multimodal fusion framework, and ethical and legal aspects. Compared with existing reviews, this article focuses on the perspective of cross-integration between technology and clinical practice, clarifying future research directions such as the development of cross-language models and the construction of a "text + physiological signal" integration framework, with the aim of providing ideas for the intelligent development of mental health services.

Keywords: Natural Language Processing (NLP), Mental Health, Sentiment Analysis, Psychological Intervention.

1. Introduction

Mental health disorders have now become a serious public health challenge on a global scale. According to statistics, the global prevalence rates of depression and anxiety disorders have reached 5% and 4% respectively, and the economic losses caused by mental disorders are expected to be as high as 6 trillion US dollars by 2030 [1]. Traditional mental health services have always been confronted with numerous challenges: the diagnostic process lacks objective standards, the quality of treatment varies greatly, and there is a shortage of professionals. Especially in rural and developing countries, over 85% of patients have difficulty accessing effective intervention [1].

The rise of natural language processing (NLP) technology offers new possibilities for solving these problems. Over the past two decades, the research focus on the mental health of children and adolescents in China has shifted from "epidemiological investigation" to "digital intervention" [2]. NLP is one of the core technologies of "digital intervention". By quantitatively analyzing unstructured data such as conversation transcriptions, clinical notes, and social media texts, NLP can uncover hidden psychological state information. In recent years, research in this field has grown significantly. After 2019, the number of related achievements has increased markedly. Among them, the proportion of studies published from 2020 to 2022 reached 53.9%, and the sample size has expanded from less than 100 treatment records initially to over one million conversations [1]. Based on the research literature of the past five years, this article systematically reviews the application status, technical route performance and clinical value of NLP in the field of mental health, focuses on the cross-integration of technology and clinical practice, and provides a reference for promoting the intelligence of mental health services.

2. A Comparative Analysis of the Methodologies Based on Different Technical Routes

2.1. Rule-based Methods for the Detection and Intervention of Mental Health Disorders

Rule-based methods rely on manually defined dictionaries and grammar rules (such as LIWC dictionaries and sentiment word table) to analyze texts by matching pre-set patterns, understand simple natural language instructions, and thereby simulate human conversations.

The review by Malgaroli shows that 43 studies adopted this method to extract features such as emotion and cognition [1], which is suitable for scenarios with low requirements for semantic complexity and relatively regular text formats, such as structured text (such as standardized scales) and simple emotion classification. Raveau identified psychological stressors by customizing subject word lists (for example, the "violence" theme included "rape, bullying"), and achieved an accuracy rate of 0.73-0.93 in the classification of hotline conversations (0.73 for the self-image theme and 0.93 for the theme of sexual minorities) [3]. This indicates that in topic scenarios where semantic features are more focused and explicit (such as themes related to sexual minorities), using rule-based methods, due to their higher table coverage and classification accuracy, also leads to a relatively higher accuracy rate.

The advantage of this method lies in its strong interpretability and low cost, without the need for large-scale labeled data. However, it lacks the ability to handle complex texts such as metaphors and irony. The research by Zhang shows that the F1 value of rule-based methods in unstructured texts on social media is only 0.5-0.6, which is much lower than that of machine learning methods [4]. Therefore, such methods are more suitable for simple scenarios with limited resources and have limited applicability in complex clinical conversations.

2.2. Detection and Intervention of Mental Health Disorders Based on Traditional Statistical Machine Learning Methods

The core of statistical machine learning methods is to convert unstructured text into numerical vectors recognizable by the model through artificial feature engineering, and then rely on statistical models to learn data patterns. There is no longer a need to manually encode clear grammar rules, but a large amount of corpus accumulation is required to learn language patterns from them.

The core logic of this type of method is to convert text into numerical vectors through feature engineering such as TF-IDF and n-grams, and then complete classification or prediction in combination with models such as SVM and random forest. The review by Malgaroli shows that 30 studies used TF-IDF to model word frequency, and 43 studies used n-grams to capture word order. They performed stably in moderately complex tasks, such as intervention type identification and symptom screening [1].

The research of Zhang pointed out that the F1 value of SVM combined with TF-IDF in the detection of depression in clinical notes can reach 0.6-0.7, and the stability is outstanding in scenarios with a sample size of less than 1000 [4]. Its advantage lies in having a low requirement for data volume and having better domain adaptability than rule-based methods. For instance, it can adapt to different languages or text types by adjusting the feature weights. However, there are also obvious shortcomings, namely, excessive reliance on artificial feature design. Malgaroli[1] found that in unstructured texts such as patients' free narratives, the accuracy rate is 10%-15% lower than that of deep learning.

Overall, this method stands out for its cost-effectiveness in clinical scenarios with limited data. However, the subjectivity of feature engineering may introduce biases and affect the stability of the results.

2.3. Detection and Intervention of Mental Health Disorders Based on deep Learning Methods.

When dealing with mental health texts based on deep learning methods, there is no longer a reliance on manually designed features. The core lies in automatically mining text patterns through neural networks without the need for manual intervention in feature extraction.

Neural network models represented by LSTM and Transformer automatically learn text features through pre-training (such as BERT) and fine-tuning. Malgaroli's review mentioned that 40 studies adopted deep learning, among which the Transformer architecture demonstrated significant advantages in context-dependent tasks, such as conversation intent recognition and complex sentiment analysis, with an F1 value of 0.75 in CBT conversation quality assessment [1].

Zhang's analysis confirmed that BERT's F1 value in the detection of depressive symptoms on social media reached 0.8-0.85, far exceeding traditional methods, and is particularly suitable for large-scale, unstructured texts (such as social media posts and long conversation transcribed texts) [4]. Its greatest advantage is that it does not require artificial feature engineering and can capture deep semantics, such as identifying suicide risks from the phrase "Life is meaningless". Miao Danmin et al.'s research also shows that the deep learning model combining text and eye movement data can achieve a depression recognition accuracy rate of 91.80%.

However, the limitations are equally prominent, that is, it requires large-scale labeled data of over 10,000 samples. Wang Hui et al. pointed out that in the detection of small-sample rare diseases such as post-traumatic stress disorder, its performance would decline by 15% to 20% [5]. In addition, the "black box" feature may reduce clinical trust, and it is necessary to combine interpretability tools such as SHAP to enhance practicality.

3. Limitations and Prospects

3.1. Limitations

The application of NLP in the field of mental health has accumulated many research achievements, but there is still obvious room for exploration.

Malgaroli conducted a systematic review of 102 studies and found that the application of NLP in mental health intervention has grown rapidly [1]. The proportion of studies from 2020 to 2022 reached 53.9%, and 87.3% of the studies confirmed that the contribution of text features to model accuracy far exceeded that of audio features. This provides a basis for this paper to focus on text data. However, the review also points out that the existing research has two major limitations: insufficient language diversity (87.3% of the data is in English) and low reproducibility (only 15.7% is in public code), which seriously restricts the cross-cultural promotion of the technology.

In terms of disease detection, Zhang analyzed 399 studies and pointed out that deep learning models such as BERT achieved an F1 value of 0.8-0.85 in the detection of depression and suicide risk, significantly outperforming traditional machine learning (F1 value of 0.6-0.7) [4]. However, these models have poor performance in detecting rare diseases such as post-traumatic stress disorder, and 81% of the data relies on social media. The adaptability to clinical scenarios needs to be verified urgently, which is precisely the direction that this article hopes to expand.

The research on special scenarios is also quite inspiring. Raveau [3] based on 10,986 conversations from the Chilean Mental Health Hotline, confirmed that NLP can effectively identify psychological stressors during the pandemic, with the correlation coefficient between the theme of emotional crisis and depressive symptoms as high as 0.802 ($p < 0.01$). However, the dictionary method adopted in this study has limited ability to capture complex semantics such as metaphor and irony, suggesting that deep learning needs to be combined to improve the analysis accuracy.

Multi-quality fusion technology provides ideas for breaking through the limitations of a single mode. Miao Danmin et al. proposed that integrating multi-source data such as text, eye movement, and facial blood flow can enhance the objectivity of assessment [6]. The eye movement feature model developed by their team achieved a depression recognition accuracy rate of 91.80%. However, current

research mostly focuses on the integration of a single technology and lacks a systematic exploration of the framework for integrating NLP and cognitive neural data. This is precisely the methodological gap that this paper aims to fill.

From an ethical and legal perspective, Ye Haosheng pointed out that in the application of mental health education for teenagers, artificial intelligence poses ethical risks such as privacy protection and data bias [7]. In the field of mental health, the vast amount of user privacy data collected by NLP technology is highly likely to be leaked if not stored or used properly. When general artificial intelligence is applied in the field of psychology, it faces legal predicaments such as an unclear responsible party [8]. Wang Hui et al. also emphasized that there is a gap in ethical review and supervision in the application of artificial intelligence in psychological intervention [5]. For instance, in cases of AI infringement related to emotions, issues such as the infringement of personal rights and mental damage are involved. There are difficulties in determining the responsible party and proving the causal relationship. The current research, when applying NLP technology, has not paid sufficient attention to these ethical and legal issues and explored the corresponding countermeasures.

3.2. Prospects

References are cited in the text just by square brackets [1]. (If square brackets are not available, slashes may be used instead, e.g. /2/.) Two or more references at a time may be put in one set of brackets [3, 4]. The references are to be numbered in the order in which they are cited in the text and are to be listed at the end of the contribution under a heading References, see our example below.

Different technical routes have their own suitable scenarios in the field of mental health: rule-based methods are suitable for simple structured text analysis, traditional machine learning is more cost-effective in small sample clinical scenarios, and deep learning performs best in large-scale unstructured text. However, the existing research still has obvious shortcomings: insufficient language diversity (mainly English), a lack of verification of clinical scenario adaptability, and the absence of a multimodal fusion framework.

In the future, breakthroughs can be made in three aspects: First, promote the development of cross-language models, and optimize multi-language adaptation in combination with the knowledge graph technology mentioned by Yuan Zhitao et al. [9], not just English, because there are a large number of children and adolescents with mental health problems in rural areas of China, and a large dialect database is needed to support it. Second, enhance the accumulation of clinical data and leverage the advantages of Transformer pre-training pointed out by Hao Litao et al. to improve the accuracy of disease detection through fine-tuning of clinical texts [10]. Third, referring to the idea of multi-quality integration proposed by Miao Danmin et al. [6], a framework of "text + physiological signals" is constructed to enhance the objectivity of the assessment. At the same time, ethical norms cannot be ignored. Standards should be established in areas such as data privacy protection and algorithm transparency to promote the secure implementation of technology.

4. Conclusion

The social impact of mental health issues is becoming increasingly severe, and the global mental health burden continues to grow. Meanwhile, traditional mental health services are confronted with numerous challenges and are unable to meet the current demands for research and services related to mental health. Under such immense pressure, natural language processing technology has emerged as a new turning point. By quantitatively analyzing various unstructured data, NLP can not only more accurately uncover the hidden psychological state information of the researchers but also significantly enhance efficiency. However, at the same time, there are also many limitations that need to be further improved in the future. This article divides the application of NLP in the field of mental health into two aspects: research and intervention. It analyzes from three perspectives: application status, performance comparison of three different technical routes, and clinical value, providing new improvement ideas for the future application of NLP in the field of mental health.

References

- [1] Malgaroli M, Hull T D, Zech J M, Althoff T. Natural language processing for mental health interventions: A systematic review and research framework. *Translational Psychiatry*, 2023, 13 (1): 309.
- [2] Wan Y C, Meng Y. Twenty years of research on mental health of Chinese children and adolescents: Research hotspots, knowledge mapping, and development trends. *Psychological Research*, 2024, 17 (1): 87-96.
- [3] Raveau M P, Goñi J I, Rodríguez J F, Paiva-Mack I, Barriga F, Hermosilla M P, et al. Natural language processing analysis of the psychosocial stressors of mental health disorders during the pandemic. *npj Mental Health Research*, 2023, 2 (1): 17.
- [4] Zhang T, Schoene A M, Ji S, et al. Natural language processing applied to mental illness detection: A narrative review. *NPJ Digital Medicine*, 2022, 5 (1): 46.
- [5] Wang H, Hu Y H, Feng X D, et al. Application of artificial intelligence in psychological interventions: Effects, challenges, and prospects. *Chinese General Practice*, 2025, 28 (25): 3209-3216.
- [6] Miao D M, Cao S, Liu Z, et al. Development and challenges of multi-modality fusion psychometric techniques. *Journal of Air Force Medical University*, 2023, 44 (10): 909-915.
- [7] Ye H S. How artificial intelligence empowers adolescent mental health education: Adaptation analysis based on Chinese social culture. *Chinese Moral Education*, 2025, (09): 35-41.
- [8] Zhu T S. Analysis of the application of general artificial intelligence in psychology. *People's Forum · Academic Frontiers*, 2023, (14): 86-91+101.
- [9] Yuan Z T, Liang S Y, Wang Q Y, et al. Visual analysis of natural language technology applications in psychology based on knowledge graph. *Journal of Air Force Medical University*, 2025, 1-10.
- [10] Hao L T, Yu Z S. Development and application of artificial intelligence-based natural language processing technology. *Heilongjiang Science*, 2023, 14 (22): 124-126.