MentalNet: A Individual Level Mental Disease Detection Framework Demonstrated with Users' Social Media Posts

Tzu-Hao Mo[1#], Jialiang Zhou[1#], Salih Selek, M.D[2], Hongfang Liu[2], Ming Huang, PhD[2]
[1]University of Pennsylvania, Philadelphia, PA, USA
[2]University of Texas Health Science Center at Houston, TX, USA
[#]Equal contribution

Abstract
In this study, we propose a novel individual level mental disease detection framework, MentalNet, for identifying social media users with mental health conditions with their social media posts for potential digital intervention. The current detection framework includes and implemented data preprocessing, sematic embeddings generated from pre-trained language models, and classifiers based on Convolutional Neural Networks (CNN). Our results show that the MentalNet framework outperforms existing deep learning methods such as GPT-4 and BERT-based models, showing its feasibility and reliability for the effective detection of mental health conditions in the user level. The detection framework is a highly flexible architecture and enables the incorporation of different health documents, embedding techniques, and classification methods. The framework also allows the applications beyond the mental health field, although it is demonstrated for mental disease detection.

Introduction Mental illness poses a significant challenge to public health and economic stability, affecting millions globally. The early detection and intervention of mental health conditions are crucial, because they can lead to improved outcomes and substantial healthcare savings. The increasing trend of individuals using social media platforms like Reddit to express their life difficulties and health issues highlights the evolving role of digital platforms in mental health support and discussion. The anonymity offered by these platforms encourages users to share personal information and discuss challenging life experiences with less fear of offline situations. This shift has been recognized in recent research, underscoring social media's effectiveness as a medium for self-disclosure and seeking social support for mental health challenges. Reddit, in particular, has emerged as a significant resource in this context. With over 3 million subreddits as of 2023, the substantial volume of discussions on Reddit offers a valuable dataset for identifying individuals with mental disorders for potential intervention. However, the vast quantity of posts and replies presents a substantial challenge for mental health providers attempting to identify individuals in need of help. This challenge, in turn, opens significant opportunities for the application of deep learning (AI) and natural language processing (NLP) methods. These technologies have the potential to automatically parse millions of social media posts, effectively identifying users who may be experiencing mental illness and could benefit from intervention. The development and implementation of such tools could revolutionize the way mental health support is provided on social media platforms, offering a scalable and efficient approach to reaching individuals in need. Recently large language models (LLMs) have revolutionized the field of machine learning and NLP due to their simplicity, efficient processing, and their ability to achieve state-of-the-art results across a wide range of NLP tasks. Current LLMs have limitations in processing extensive text data due to token constraints. For example, while BERT models are typically restricted to 512 tokens, GPT-3.5 and GPT-4 can handle up to 4096 tokens. However, even with this increased capacity, analyzing data from a user's entire social media posts remains challenging for mental health detection at the individual level. To fill in the gap, we propose a novel deep learning framework, MentalNet, for identifying mental health conditions in the individual level with their social media posts for potential intervention. The MentalNet's ability to analyze a large collection of social media posts can offer a scalable solution for mental health screening, reaching a broader population than traditional methods.

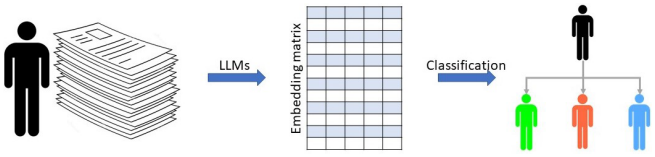

Figure 1. Architecture of MentalNet to detect mental health disorders at the individual level.

Methods The MentalNet framework is designed to detect various mental diseases in the individual level, the architecture of which is shown in **Figure 1**. The current framework includes and implements standardized text preprocessing, semantic embeddings generated from pretrained LLMs, and classification based on CNN. The data preprocessing stage primarily aims to generate the text inputs with appropriate lengths that meets the token limits of LLMs, together with other processes for performance improvement. For example, it includes steps such as filtering English reviews, correcting the grammar and spelling, expanding the word abbreviations, converting emojis and special characters to text, and removing short posts that has little semantic context. For embedding generation, we use a wide range of encoder-based LLMs – BERT and its variants, from

general-purpose pre-trained models such as BERT and sentence-BERT to specialized and purpose-built pre-trained language models such as BioBERT and MentalBERT, to prepare the data for the CNN-based classifier. These pre-trained language models help convert each processed post into a computational vector which learns semantic information and the language patterns in the text. Thus, a user is represented by an embedding matrix with a length of processed posts and a width of embedding dimension. CNN was used for the transformation of the embedding matrix of each user from the pretrained models and the identification of behavioral patterns of users from transformed embeddings to better classify the users' mental health condition more accurately. The MentalNet was trained and tested using the Self-reported Mental Health Diagnoses (SMHD) dataset[1] as shown in **Table 1** to detect 9 types of mental health disorders with users' Reddit posts. The 9 mental diseases include Depression (Dep), Attention-Deficit/Hyperactivity Disorder (ADHD), Anxiety (Anx), Bipolar (Bip), Post-Traumatic Stress Disorder (PTSD), Autism (Aut), Obsessive Compulsive Disorder (OCD), Schizophrenia (Sch), and Eating disorder (Eat). The model performance was measured using accuracy, precision, recall and F1 score.

Table 1. SMHD dataset used in this study

Mental Condition	Training Set			Validation Set			Testing Set		
	Users	Case	Control	Users	Case	Control	Users	Case	Control
Dep	2,662	378,175	653,322	2,573	356,636	634,840	2,611	358,561	640,150
ADHD	1,768	252,580	432,613	1,747	250,206	430,012	1,779	248,794	442,564
Anx	1,711	238,778	418,978	1,592	219,570	382,977	1,675	227,133	415,264
Bip	1,216	169,712	299,470	1,182	156,992	289,206	1,247	165,550	308,217
PTSD	528	76,378	131,191	516	71,848	126,713	558	75,408	136,778
Aut	479	65,567	119,044	480	72,704	116,373	517	74,915	127,552
OCD	409	53,768	100,628	477	66,162	115,649	390	53,667	96,640
Sch	238	31,777	59,006	278	37,322	67,827	267	35,498	66,009
Eat	104	14,199	27,660	115	16,210	26,859	112	16,427	26,381

Results Figure 1 shows the performance of MentalNet with sentence BERT for embedding generation in detecting the 9 mental diseases, comparing with other LLMs reported in the literature[2]. MentalNet has demonstrated remarkable performance in identification of users with mental health problems. Evaluated by F1 score, MentalNet is ranked number one in 6 out of 9 mental health conditions (ADHD, Anx, Bip, Aut, Sch, and Eat) and ranked number two in two mental illness (Dep and PTSD) with small difference with the best F1. Notably, the performance of MentalNet is significantly higher than other methods for detecting 3 mental health problems - Aut, Sch and Eat. This performance underscores MentalNet's effectiveness in precise categorization of the user's mental health issue. Besides sentence BERT for embedding generation, we also explored 9 BERT-based LLMs for generating embedding for MentalNet. We tested the framework with Depression detection. The results in **Figure 2** show MentalBERT) demonstrated a higher performance with a F1 score of 0.85, compared to other tested models including GPT-4 (0.84).

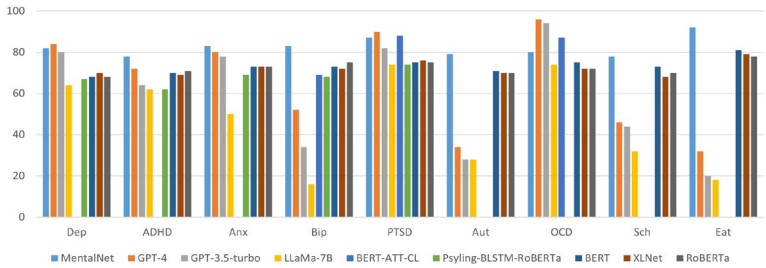

Figure 1. Performance (F1) of MentalNet compared to other models reported

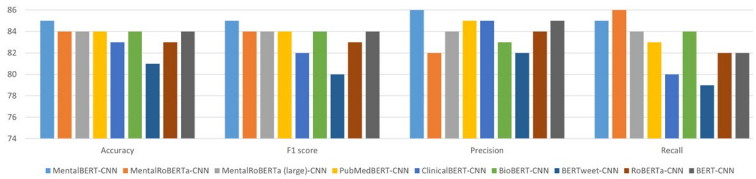

Figure 2. Performance of MentalNet on Depression detection with different LLMs

Discussion and Conclusions This study showcases MentalNet's capability and effectiveness in identifying mental health conditions at the individual level by analyzing extensive social media data. MentalNet demonstrates a scalable approach to mental health screening, capable of processing large volumes of user-generated content on social media. This approach significantly extends the reach of mental health detection efforts beyond the capacities of conventional screening methods, potentially engaging a much wider and more diverse population. While our research primarily utilizes the SMHD dataset to validate MentalNet's performance in detecting various mental health conditions, the framework's design is inherently flexible. MentalNet's architecture is highly adaptable, allowing for the incorporation of diverse health documents, advanced embedding techniques, and various classification methodologies. This versatility not only underlines MentalNet's robustness in mental health applications but also opens avenues for its use in broader health and research fields, demonstrating its potential as a multifaceted tool in health informatics.

References

1. Arman Cohan, Bart Desmet, Andrew Yates, Luca Soldaini, Sean MacAvaney, and Nazli Goharian. 2018. SMHD: a Large-Scale Resource for Exploring Online Language Usage for Multiple Mental Health Conditions. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1485–1497
2. Jin H, Chen S, Wu M, Zhu KQ. PsyEval: A Comprehensive Large Language Model Evaluation Benchmark for Mental Health. arXiv preprint arXiv:2311.09189. 2023 Nov 15.