

Mental Health Prediction Using Large Language Models

Binu P Chacko

Department of Computer Science

Prajyoti Niketan College, Pudukad,

binupchacko@gmail.com

Abstract : Mental health is a complex disease involving numerous individual, socioeconomic, and clinical variables. Mental illness is a health problem that undoubtedly impacts emotions, reasoning, and social interaction of a person. For children, their ability to learn could be interfered by mental disorders. Early detection of mental health issues is an essential procedure. Traditionally, mental health professionals use face-to-face interviews, self-reporting and questionnaire distribution to distinguish between the mental health status of individuals. Data preprocessing is the first step to be performed after collecting the data. It is used to clean and normalise the data before fitting it into a model. It is necessary because unprocessed raw data may contain noise, missing values, and duplicates that could affect the model's accuracy. The next step is to collect the features and select the most discriminate set to identify the disease. For the purpose of mental health prediction, approaches based on machine learning, deep learning and large language models are explored. Commonly used datasets in this domain are also discussed.

Keywords: Mental health prediction, Datasets, Machine learning, Deep learning, LLM

Introduction

Mental health problem or mental illness is a health issue that affects a person's emotions, thoughts, behaviour, connection with society, and daily activities. It results in serious consequences across families and societies and demands new strategies for prevention and intervention. The types of mental health problems include anxiety disorders, depression, stress, bipolar disorder, BPD, Schizophrenia, posttraumatic stress disorder (PTSD), self-harm, suicide watch, addiction, severe alcoholism, opioids, and autism. The diagnosis of mental health problems will begin with a specific interview that is filled with questions about symptoms, medical history, and physical examination. Besides that, psychological tests and assessment tools are also available [1]. The PHQ-9 questionnaire is used to measure depression in adults. In clinical settings, disorders are identified using the statistical manual of mental disorders (DSM-5). The clinical staging paradigm allows that different levels of care to be assigned and early intervention is needed to slow down or prevent the emergence and recurrence of certain diseases [2]. With proper care and treatment, many individuals will hopefully be able to recover from mental illness or emotional disorder.

The factors of mental health problems can be influenced by biological, psychological, and environmental issues. Nowadays, majority of students grumble about the high level of stress including feelings of anxiety and depression, especially towards the end of the semester. The level of stress increases as the learning process progresses due to the need to balance assessments, workload, and examinations. Students may face a high risk of developing mental health problems due to family issues, uncertainties about their future careers, financial troubles, and difficulties arising out of living away from home. The student is bound to experience loneliness, hence increasing their stress level, which may lead to mental health problems such as depression and anxiety disorder [3].

The increase of mental health problems and the need for effective medical health care have led to an investigation of machine learning (ML) in mental health problems. The algorithms such as support vector machine (SVM), random forest (RF), and artificial neural network (ANN) are used to predict, diagnose, and identify mental health problems. Supervised learning is the most widely applied approach in many types of research, studies, and experiments, especially in predicting illness in the medical field. The performance on the ML algorithms or techniques is being

evaluated based on accuracy, sensitivity, specificity, or area under the ROC curve (AUC) [1]. [4] made use of brain imaging data, electronic health records (ERs), novel smart monitoring systems, mood rating scales, and platforms on social media in classifying and predicting mental illness.

Major source of information collection in mental health has been text-based records. Therefore, the main approach to detect psychiatric disease is via text mining. In [5], 150 text files were manually used for psychiatric symptoms and attributes. A system was designed to automatically detect the attributes, and the workflow made use of a neural network model. In order to treat at an early stage, deep learning tools were used in diagnosing and forecasting children's mental health. A convolutional neural network (CNN) was first used to examine the deep parts of children's behaviour data features, followed by insights into the disruption, classification, and then finally, forecasting [6]. The classification of mental health-related texts, which can serve as early warnings and early diagnoses, is facilitated by analytical and natural language processing techniques. The utilized data set consists of texts on mental health with labels. Text reflects an individual's overt emotional state, such as posts on social media platforms, and interview transcriptions. Natural Language Processing (NLP) has the potential to be utilized in the identification of emotions and the provision of mental health surveillance [7].

Datasets

Existing works on mental health analysis and detection can be categorised by data modality. A line of work seeks to monitor mental health by monitoring physiological data such as heart-rate variability and respiratory rate [8] or motor activity data such as the intensity and duration of movement [9]. Another line of work seeks to detect and analyse an individual's mental health through the use of audio-visual (AV) sources. AV datasets typically include behavioural signals such as facial affect, body gestures and vocal intensity [10].

Depresjon and Psykose consist of motor activity data, whereas D-Vlog includes audio-visual recordings. Each dataset has a different distribution breakdown across gender and mental health disorders (MHD) class. In Depresjon and Psykose, clinical experts diagnosed the patients, whereas in D-Vlog, annotators were instructed to identify depressed vlogs based on statements which indicate that they are either suicidal or on MHD medication. Depresjon consists of motor activity data collected from 55 individuals. Data was recorded using an ActiGraph wristband. Out of 55 participants, 23 individuals have been diagnosed with depression ($Y = 1$). This includes both unipolar and bipolar depression. The dataset contains 30 females and 25 males. Psykose consists of motor activity data collected from 54 individuals. Data was recorded using an ActiGraph wristband. Out of 54 participants, 22 individuals have been diagnosed with schizophrenia ($Y = 1$). The dataset contains 23 females and 31 males. D-Vlog consists of vlog data collected from YouTube videos. The dataset contains 555 depressed and 406 non-depressed vlogs of 639 females and 322 males [11].

The mental health text corpus and the dataset consist of comments related to mental health concerns, such as anxiety and depression. Mental Health Corpus is acquired from Kaggle (<https://www.kaggle.com/datasets/reihanenamdari/mental-health-corpus/data>) and the Reddit mental disorders identification dataset from Kaggle, (<https://www.kaggle.com/datasets/kamaruladha/mental-disorders-identification-reddit-nlp>). The Reddit dataset produced seven unique categorizations related to mental health, encompassing borderline personality disorder (BPD), anxiety, depression, bipolar disorder, mental illness, schizophrenia, and poison. BPD is a psychiatric disorder characterized by impulsive behaviour, emotional instability, unstable interpersonal relationships, and a distorted sense of self. Anxiety is a mental health issue characterized by excessive anxiety, worry, and behavioural disturbances. Depression is a psychological condition characterized by feelings of sadness, loss of interest in daily activities, decreased appetite, and difficulty focusing. Bipolar disorder, previously referred to as manic depressive illness, is characterized by abnormal fluctuations in mood, energy levels, activity, and concentration. Mental illness is a condition that impacts an individual's thoughts, mood, behaviour, and cognitive functioning. Schizophrenia is a psychiatric disorder that affects an individual's thoughts, emotions, and behaviour, and is characterized by hallucinations and delusions. The text additionally addresses the topic of poison, which is commonly linked to individuals who are afflicted with symptoms of anxiety, despair, and other mental health disorders [7].

Xu, Xuhai, et al [12] conducted experiments using four datasets (Dreaddit, DepSeverity, SDCNL, CSSRS-Suicide) collected from Reddit. They defined six diverse mental health prediction tasks based on these datasets.

Dreaddit: The dataset contains ten subreddits in five domains (abuse, social, anxiety, PTSD, and financial) and includes 2929 users' posts. This dataset is used for a post-level binary stress prediction (Task 1).

DepSeverity: This dataset is employed for two post-level tasks: binary depression prediction (Task 2), and four-level depression prediction (minimal, mild, moderate, severe - Task 3).

SDCNL: This dataset is employed for the post-level binary suicide ideation prediction (Task 4).

CSSRS-Suicide: This dataset is used for two user-level tasks: binary suicide risk prediction (Task 5), and five-level suicide risk prediction (supportive, indicator, ideation, behaviour, attempt - Task 6).

The following datasets are also used to perform different tasks [12].

Red-Sam: This dataset also collected posts involving five subreddits (mental health, depression, loneliness, stress, anxiety). This dataset is used as an external evaluation dataset on binary depression detection (Task 2).

Twt-60Users: This dataset is used as an external evaluation dataset on depression detection (Task 2).

SAD: This dataset contains SMS-like text messages with nine types of daily stressor categories (work, school, financial problem, emotional turmoil, social relationships, family issues, health, everyday decision making, and other). This dataset is used as an external evaluation dataset on binary stress detection (Task 1).

Preprocessing

Data pre-processing is an essential stage in the development of a classification model. Before applying machine learning techniques, the data needs to be processed by removing the redundancy, inconsistency, noise and missing values from it [2]. A numerical cleaner filter [13] is used to find missing values in the data and evaluate features. By changing numeric data to a present value, this filter cleans numeric data that is either too large or too little. Further, data cleaning helps to ensure that the dataset used to train or evaluate the model is accurate, consistent, and relevant. It also helps to reduce bias and variability in the model, resulting in more accurate and reliable predictions.

In order to evaluate and mitigate bias, sensitive attributes such as gender, ethnicity and age will be required. Bias mitigation can be conducted at either the pre-processing, in-processing or post-processing stage. Pre-processing methods typically attempt to mitigate bias at the data-level by collecting or resampling data points in order to create a balanced dataset [11]. Zhang et al. [14] proposed a pre-processing mixup data augmentation strategy to generate new samples by mixing up different features and their corresponding labels. They do so for the minority group in order to obtain balanced samples across gender. In-processing methods mainly involve model-level interventions at different stages of ML models or algorithms to mitigate bias. As an in-processing bias mitigation method, Calders et al. [15] proposed an effective loss re-weighting approach. This method works by penalizing misclassified minority classes at a higher weighted proportion. The algorithm is encouraged to pay more attention to the minority class which assists in bias mitigation. Post-processing methods mainly modify the model output to achieve fairer predictions. Kamiran et al. [16] suggested the Reject Option Classification (ROC) method to re-classify the predictions of the minority group in favour of the minority group if predictions fall within a certain decision threshold region.

Feature Selection

Feature Selection is carried out to determine the relevance of each feature. It usually results in improved learning performance, lowering computing cost, and improved model interpretability. It aims at identifying a sample of highly distinguishing variables for diverse mental health diagnoses. Jiang et al. [17] used the contextual representations from Bidirectional Encoder Representations from Transformers (BERT) as input features for mental health issue detection. Feature selection information gain method can be employed as suggested by Claude Shannon's [18] because of its advantage of being fast. For a specific class, the information gain (entropy) of each feature is determined using this

technique. The entropy value goes from 0 to 1, where the highest number indicates the existence of the most relevant information. The highest rating features contribute most to the decision-making process, thus they are kept, and lower scoring features are removed [2]. The importance of each feature can be measured by Chi-square test also [19].

Machine Learning Techniques

Artificial intelligence (AI) is viewed as an innovative tool for organising mental health treatments. AI-driven solutions can employ digital health-care data, which is available in a variety of formats including electronic health records, medical pictures, and handwritten clinical notes, to automate jobs, and support doctors. AI systems have the ability to eliminate unjustified variation in clinical practise, enhance efficiency, and prevent avoidable medical errors. It has great potential to redefine clinical intervention and understanding of mental illnesses. Utilising AI techniques enables the creation of improved pre-diagnosis screening tools and the creation of risk models to assess a person's propensity. To implement personalised mental healthcare as a long-term goal, we need to harness computational approaches best suited for conflict settings [6]. Although much work has been done to predict mental health in clinical and non-clinical settings [20,21], there is still a considerable need to synchronise machine learning predictions with pre-clinical mental health anxiety stages in order to deliver individually tailored mental health recommendations. AI-based innovative diagnostic, monitoring, and therapeutic methods in psychiatry may enhance mental health outcomes, reduce the treatment gap, reduce inequality, public health administration and minimise practitioner workload.

The accuracy of ML approaches is determined by the type of prediction techniques and data acquisitions, such as clinical data, self-reporting data, or screening tools. The hybrid approaches and SVM were the most widely used methods for forecasting PTSD and seasonal affective disorder (SAD). In order to diagnose CKD in patients, Sawhney et al. [22] suggested an approach based on deep neural network. The machine learning techniques such as CNN, SVM, linear discriminant analysis, and K-nearest neighbour have been used to categorise the intensity level of the anxiety and depression [2]. Priya et al. [23] discovered that the Naive Bayes classifier performed the best for predicting depression.

Benton et al. [24] utilized multi-task neural learning (MTL), regression, and multi-layer perceptron single-task learning (STL) models to classify different forms of mental health illnesses. In [25], an SVM model was trained to categorize 200 text messages into two distinct groups: ADHD or non-ADHD. An essential stage in the procedure was eliminating ADHD-associated keywords from the messages prior to learning, with the objective of assessing SVM's ability to learn in the absence of pertinent keywords and semantic content.

Mohamed et al. [2] proposed a risk prediction model based on pre-clinical multistage classification strategy to classify individuals early before the onset of disorders based on severity level, i.e., Stage_1: Minimal Anxiety (MA); Stage_2: Mild Anxiety (MIA); Stage_3: Moderate Anxiety (MDA), Stage_4: Severe Anxiety (SA) and Stage_5: Very Severe Anxiety (VSA). The purpose of the study is to pinpoint the stage of anxiety a person is experiencing and to identify the major triggers of anxiety. It also aims to shorten the screening period for anxiety stages. The article intends to use real-world data to predict various anxiety stages using a carefully chosen machine learning approach that comprises SVM, MLP and RF. Additionally, a methodical feature selection process was used to identify 10 best characteristics out of 20 that had the highest accuracy. The proposed methodology is a hybrid of two models: Mental Health Anxiety Stage Disorder Prediction Model (MH_ASD_PM) and Anxiety Stage Mental Health Recommendation Model (AS_MH_RM) based on statistical analysis that can assist psychiatrists and individuals in making quick clinical decisions regarding mental health disorder diagnosis and medical recommendations. Feature extraction selects the most essential and informative attributes to aid the multi-classification process. Finally, one of the five stages is predicted using the specified features of General Anxiety disorder stages. The AS_MH_RM then uses the prediction information, as well as the medical history, to recommend general mental health advice to individuals suffering from a particular predicted anxiety level. The objective is to generate a model by estimating the risk with each variable against a set of vulnerabilities.

The information of severity ranges is used to spot and classify one anxiety stage from another. The rule predefined stipulates that whenever a new individual/patient seeks mental health treatment arrive, tests are performed to determine the anxiety stage based on the severity score. Following computational predictions, spectrum of exposure

parameters is established, and calculations are determined based on the spectrum, risk factor, and likelihood. Lastly, the acquired data is compared to the rules predefined to determine the severity stage. The final score is obtained using the severity values and score from the base of knowledge [2].

The methods that are most ideal for classification and regression operations are known as ensemble methods. This combines several models to form one good model via reducing variance and bias in order to enhance the prediction accuracy [26]. Common methods used for these are boosting and bagging, of which boosting produces better results than bagging. A boosting algorithm is generally a learning algorithm in the field of machine learning. The algorithm was further improved by Payal et al. [27] and it is named as AdaBoost algorithm. Sharma and Verbeke [28] made use of ML in developing a model that can enhance the diagnosis of mental depression. Extreme Gradient Boosting (XGBoost), a type of ML algorithm, was applied to each sample, and classification was carried out. In [29], the gradient-boosting machine learning algorithm has been trained with 25 features on datasets in STAR*D and COMED AXA to predict the treatment outcomes.

In [19], correlation analysis has been applied to identify the correlation among attributes and regression analysis was applied for the prediction of the mental health status of the students based on the given attribute. It was observed that anxiety and depression were positively correlated and social connectedness and depression were negatively correlated. The hyperparameter-tuned logistic regression model provided better result in comparison to the other existing models in the literature.

During the prediction, an intuitive and user-friendly recommendation model examines the data of the individual/patient to calculate the risk. Risk factors might be the features, variables, or hazards that, if present in a specific individual, increase the likelihood of that individual developing a disorder. Depending upon the severity and likelihood of disorder, accurate clinical psychiatric suggestions for future course of action are then made. Final, recommendations are drawn based upon knowledge base created with the help of a clinical psychiatrist. It assists in identifying the importance and primary concern of environmental factors that plays a significant role in predicting anxiety stage disorders. This recommendation model would benefit both individuals and psychiatric practitioners by reducing workload and time costs and will significantly help low-income individuals from huge financial loss [2].

Deep Learning Approaches

Text analysis pertaining to mental health has employed a diverse range of methodologies, which includes both ML techniques and advanced deep learning approaches. Deep feedforward neural network has been utilized in the analysis of clinical and genetic data to forecast mental health diseases. As an illustration, the diagnosis of depression included word embedding in conjunction with neural networks like CNN and recurrent neural networks (RNN) [30]. In the domain of binary classification for mental health-related text posts, Gkotsis et al. [31] employed Feed Forward Neural Networks, CNN, conventional machine learning techniques like SVM, and linear classification. A hierarchical attention network was trained to perform binary classification for each mental disorder, including depression, ADHD, anxiety, and others [32].

In [7], the model generation procedure utilizes a combination of CNN and Bidirectional Long Short Term Memory (BiLSTM). CNN is employed to extract local features from the text, while BiLSTM offers the advantage of comprehending the contextual information of preceding and subsequent words through its forward and backward layers. In addition, the utilization of FastText word weighting is employed in tandem with CNN-BiLSTM to generate word representations for terms that are absent from the training data or to tackle out-of-vocabulary challenges.

FastText is a word weighting technique that was derived from the Word2vec method[33]. The AI research team at Facebook is influenced from Mikolov's study [34] in the development of their model. The study presented evidence that FastText has the capability to effectively train a corpus of 1 billion words within a time frame of 10 minutes, while yet achieving comparable outcomes to other existing models [35]. The architectural design of the FastText model, akin to the Continuous Bag-of-Words (CBOW) approach employed in Word2Vec, has a hierarchical structure and encodes words as dense vectors. After applying the FastText word weighting technique, the textual input was converted into vector representations.

In [36], researchers utilized datasets obtained from Reddit and Twitter to examine the performance of LSTM, Bi-LSTM, and BERT+KD models in classifying depression and anxiety labels. The study conducted by Trotszek et al. [37] centered on the timely identification of depression using FastText features, glove embedding, and CNN models. The utilization of FastText techniques and XGBoost models was implemented on datasets related to depression and suicidal risk [38]. Ameer et al. [39] utilized a range of advanced deep learning models, including GRU, Bi-GRU, CNN, LSTM, and BiLSTM, in addition to employing transfer learning techniques with BERT, XLNet, and RoBERTa.

De Choudhury et al. [40] extracted a number of linguistic styles and other features to build an SVM model to perform depression prediction. They have also explored deep-learning-based models for mental health prediction. Tadesse et al. [41] employed an LSTM-CNN model and took word embedding as the input to detect suicide ideation on Reddit. More recently, pre-trained language models have become a popular method for NLP tasks, including mental health prediction tasks [42].

Large Language Models

In the past decade, there has been a profusion of research in NLP and computational social science on detecting mental health issues via online text data such as social media content [43, 44]. However, most of these studies have focused on building domain-specific ML models such as stress detection, depression prediction, or suicide risk assessment. Some studies have also explored the multi-task setup [45], such as predicting depression and anxiety at the same time [46]. Another line of research has been exploring the application of chatbots for mental health services [47]. Despite the growing research efforts of empowering AI for mental health, it's important to note that existing techniques can sometimes introduce bias and even provide harmful advice to users.

Large language model (LLM) is a powerful tool for understanding end-users' mental states through their written language. Xu et al. [12] employed LLM for a variety of mental health-related tasks, such as multiple question-answering, reasoning, and inference. It opened a wide range of opportunities for mental health communities, such as mental-health-aware personal chatbots, intelligent assistants for mental health therapists, daily mental health counsellors and supporters. Authors focussed on two categories of mental health prediction targets: (1) predicting critical mental states, such as stress or depression, and (2) predicting high-stake risk actions, such as suicide. For both categories, binary and multi-class classification is performed. There are some studies on the evaluation of LLMs (e.g., ChatGPT) on mental-health-related tasks, most of which are in zero-shot settings with simple prompt engineering [48, 49]. Researchers have shown preliminary results that LLMs have the initial capability of predicting mental health disorders using natural language. [12] experimented with six LLMs with different sizes, pre-training targets, and availability – Alpaca, Alpaca-LoRA, FLAN-T5, LLaMA2, GPT-3.5 and GPT-4.

Alpaca (7B): An open-source large model finetuned from another open-sourced LLaMA 7B model.

Alpaca-LoRA (7B): This model leverages a different finetuning technique called low-rank adaptation (LoRA), with the goal of reducing finetuning cost by freezing the model weights and injecting trainable rank decomposition matrices into each layer of the Transformer architecture.

FLAN-T5 (11B): FLAN-T5 focuses more on task solving and is less optimized for natural language or dialogue generation.

LLaMA2 (70B): Open-source large model released by Meta.

GPT-3.5 (175B): This large model is a closed-source and available through API provided by OpenAI.

GPT-4 (1700B): This is the largest closed-source model available through OpenAI API.

The experiments explained in [12] contained three stages: (1) zero-shot prompting, (2) few-shot prompting, and (3) instruction-finetuning. Zero-shot setting begins with prompt design. The goal of prompt design is to empower a pre-trained general-purpose LLM to achieve good performance on tasks in the mental health domain. Eq (1) gives a general zero-shot prompt template (Prompt_{zs}) that consists of four parts.

$$\text{Prompt}_{ZS} = \text{TextData} + \text{Prompt}_{\text{Part1-S}} + \text{Prompt}_{\text{Part2-Q}} + \text{OutputConstraint} \quad (1)$$

where *TextData* is the online text data generated by end-users. *Prompt_{Part1-S}* provides specifications for a mental health prediction target. *Prompt_{Part2-Q}* poses the question for LLMs to answer. And *OutputConstraint* controls the output of models.

In order to provide more domain-specific information, researchers have also explored few-shot prompting with LLMs by providing few-shot demonstrations to support in-context learning. The intuition is to present few examples for the model to learn domain-specific knowledge. The design of the few-shot prompt (*Prompt_{FS}*) is by using eq (2).

$$\text{Prompt}_{FS} = [\text{Sample Prompt}_{ZS} - \text{label}]_M + \text{Prompt}_{ZS} \quad (2)$$

where *M* is the number of prompt-label pairs and is capped by the input length limit of a model.

The goal of instruction finetuning is closer to the traditional few-shot transfer learning, where the model is trained with a small amount of domain-specific data. They experiment with multiple finetuning strategies. They first conducted basic finetuning on a single dataset (training set). This finetuned model can be tested on the same dataset (test set) to evaluate its performance and different datasets to evaluate its generalizability. On task A, the model becomes specific to task A and is only suitable for solving that particular task. In contrast, LLMs are finetuned on several mental health datasets, employing diverse instructions for different tasks across these datasets in a single iteration. This enables them to handle multiple tasks without additional task-specific finetuning. With online social media data, the results reveal that finetuning on a variety of datasets can significantly improve LLM's capability on multiple mental-health-specific tasks.

Mental-Alpaca & Mental-FLAN-T5: Zero-shot prompting and few-shot prompting do not update model parameters during the experiment. In contrast, instruction finetuning will update model parameters and generate new models. To enhance their capability in the mental health domains, Alpaca and FLAN-T5 are updated on six tasks across the four datasets using the multi-dataset instruction finetuning method, which leads to the new models, Mental-Alpaca and Mental-FLAN-T5 [12].

Conclusion

Mental health is widely discussed topic in every community. Early detection of mental illness is very essential as it will severely affect the individual, family and society. Apart from the traditional method of collecting information through questionnaire, symptoms can be detected through the online text posted by the patient. In this instinct, advances in AI can be utilized for mental health prediction. The different techniques in AI are best utilized for this purpose. The large language models are now capable of analyzing the contents of social media post. The recommendation model can make use of this information and predict the current mental status of the individual. In the future, it is planned to do mental health prediction using multimodal data.

Reference

- [1] Chung, Jetli, and Jason Teo. "Mental health prediction using machine learning: taxonomy, applications, and challenges." *Applied Computational Intelligence and Soft Computing* 2022.1 (2022): 9970363.
- [2] Mohamed, E. Syed, et al. "A hybrid mental health prediction model using Support Vector Machine, Multilayer Perceptron, and Random Forest algorithms." *Healthcare Analytics* 3 (2023): 100185.
- [3] Mutalib, Sofianita, Nor Safika Mohd Shafiee, and Shuzlina Abdul-Rahman. "Mental health prediction models using machine learning in higher education institution." *Turkish Journal of Computer and Mathematics Education* 12.5 (2021): 1782-1792.
- [4] S. Graham, C. Depp, E. E. Lee, C. Nebeker, X. Tu, H.-C. Kim, and D. V. Jeste, "Artificial intelligence for mental health and mental illnesses: an overview," *Current psychiatry reports*, vol. 21, no. 11, pp. 1–18, 2019.

- [5] S. Madan, F. J. Zimmer, H. Balabin, S. Schaaf, H. Fröhlich, J. Fluck, I. Neuner, K. Mathiak, M. Hofmann-Apitius, and P. Sarkheil, "Deep learning-based detection of psychiatric attributes from German mental health records," *International Journal of Medical Informatics*, vol. 161, p. 104724, 2022.
- [6] Ogunseye, Elizabeth Oluyemisi, et al. "Predictive analysis of mental health conditions using AdaBoost algorithm." *ParadigmPlus* 3.2 (2022): 11-26.
- [7] Fudholi, Dhomas Hatta. "Mental health prediction model on social media data using CNN-BiLSTM." *Kinetik: Game technology, information system, computer network, computing, electronics, and control* (2024): 29-44.
- [8] Joanna C Yau, Benjamin Girault, Tiantian Feng, Karel Mundnich, Amrutha Nadarajan, Brandon M Booth, Emilio Ferrara, Kristina Lerman, Eric Hsieh, and Shrikanth Narayanan. Tiles-2019: A longitudinal physiologic and behavioral data set of medical residents in an intensive care unit. *Scientific Data*, 9(1):536, 2022
- [9] Petter Jakobsen, Enrique Garcia Ceja, Lena Antonsen Stabell, Ketil Joachim Oedegaard, Jan Oystein Berle, Vajira Thambawita, Steven Alexander Hicks, Pål Halvorsen, Ole Bernt Fasmer, and Michael Alexander Riegler. Psykose: A motor activity database of patients with schizophrenia. In 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS), pages 303–308. IEEE, 2020
- [10] Lang He, Mingyue Niu, Prayag Tiwari, Pekka Marttinen, Rui Su, Jiewei Jiang, Chenguang Guo, Hongyu Wang, Songtao Ding, Zhongmin Wang, et al. Deep learning for depression recognition with audiovisual cues: A review. *Information Fusion*, 80:56–86, 2022.
- [11] Cheong, Jiaee, et al. "Towards gender fairness for mental health prediction." *International Joint Conferences on Artificial Intelligence Organization*, 2023.
- [12] Xu, Xuhai, et al. "Mental-LLM: Leveraging large language models for mental health prediction via online text data." *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 8.1 (2024): 1-32.
- [13] B. Stark, C. Knahl, M. Aydin, K. Elish, A literature review on medicine recommender systems, *Int. J. Adv. Comput. Sci. Appl.* 10 (8) (2019) 6–13.
- [14] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations*, 2017.
- [15] Toon Calders, Faisal Kamiran, and Mykola Pechenizkiy. Building classifiers with independence constraints. In 2009 IEEE international conference on data mining workshops, pages 13–18. IEEE, 2009.
- [16] Faisal Kamiran, Asim Karim, and Xiangliang Zhang. Decision theory for discrimination-aware classification. In 2012 IEEE 12th International Conference on Data Mining, pages 924–929. IEEE, 2012.
- [17] Zheng Ping Jiang, Sarah Ita Levitan, Jonathan Zomick, and Julia Hirschberg. 2020. Detection of mental health from reddit via deep contextualized representations. In *Proceedings of the 11th international workshop on health text mining and information analysis*. 147–156.
- [18] R.B. Pereira, A. Plastino, B. Zadrozny, L.H.C. Merschmann, Information gain feature selection for multi-label classification, *J. Inf. Data Manag.* 6 (1) (2015) 48–58.
- [19] Sahu, Barnali, et al. "Mental health prediction in students using data mining techniques." *The Open Bioinformatics Journal* 16.1 (2023).
- [20] A. Goodman, R. Joyce, J.P. Smith, The long shadow cast by childhood physical and mental problems on adult life, *Proc. Natl. Acad. Sci. USA* 108 (15) (2011) 6032–6037
- [21] A. Coyle, C. Thieme, M. Linehan, J. Balaam, D. Wallace, S. Lindley, Emotional wellbeing, *Int. J. Hum. Comput. Stud.* 72 (8–9) (2014) 627–628

- [22] R. Sawhney, A. Malik, S. Sharma, V. Narayan, A comparative assessment of artificial intelligence models used for early prediction and evaluation of chronic kidney disease, *Decis. Anal. J.* 6 (2023) 100169
- [23] A. Priya, S. Garg, N.P. Tigga, Predicting anxiety, depression and stress in modern life using machine learning algorithms, *Procedia Comput. Sci.* 167 (2019) (2020) 1258–1267
- [24] A. Benton, M. Mitchell, and D. Hovy, “Multi-Task Learning for Mental Health using Social Media Text,” 2017
- [25] M. Fekihal, J. Diederich, and A.-A. A, Machine learning, text classification and mental health. 2004.
- [26] M.Zounemat-Kermani, O. Batelaan, M. Fadaee, and R. Hinkelmann, “Ensemble machine learning paradigms in hydrology: A review,” *Journal of Hydrology*, vol. 598, p. 126266, 2021
- [27] M. Payal, A. Kumar, and S. A. Ajagbe, “Support vector machines (svms) based advanced health care system using machine learning techniques,” *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 10, no. 5, pp. 3007–3014, 2022.
- [28] A. Sharma and W. J. Verbeke, “Improving diagnosis of depression with xgboost machine learning model and a large biomarkers dutch dataset (n= 11,081),” *Frontiers in big Data*, p. 15, 2020.
- [29] Chekroud AM, Zotti RJ, Shehzad Z, et al. Cross-trial prediction of [12] treatment outcome in depression: a machine learning approach. *Lancet Psychiatry* 2016; 3(3): 243-50.
- [30] D. I. Ahmed Hussein Orabi, Prasadith Buddhitha, Mahmoud Hussein Orabi, “Deep Learning for Depression Detection of Twitter Users,” *Proc. of the Fifth Work. Comput. Linguist. Clin. Psychol. From Keyboard to Clin.*, pp. 88–97
- [31] G. Gkotsis et al., “Characterisation of mental health conditions in social media using Informed Deep Learning,” *Sci. Rep.*, vol. 7, pp. 1–10, 2017
- [32] I. Sekulic and M. Strube, “Adapting deep learning methods for mental health prediction on social media,” *W-NUT@EMNLP 2019 - 5th Work. Noisy User-Generated Text, Proc.*, pp. 322–327, 2019
- [33] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching Word Vectors with Subword Information,” vol. 5, pp. 135–146, 2017
- [34] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *1st Int. Conf. Learn. Represent. ICLR 2013 - Work. Track Proc.*, pp. 1–12, 2013
- [35] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, “Bag of tricks for efficient text classification,” *15th Conf. Eur. Chapter Assoc. Comput. Linguist. EACL 2017 - Proc. Conf.*, vol. 2, pp. 427–431, 2017
- [36] K. Zeberga, M. Attique, B. Shah, F. Ali, Y. Z. Jembre, and T. S. Chung, “A Novel Text Mining Approach for Mental Health Prediction Using Bi LSTM and BERT Model,” *Comput. Intell. Neurosci.*, vol. 2022, 2022
- [37] M. Trotzek, S. Koitka, and C. M. Friedrich, “Utilizing Neural Networks and Linguistic Metadata for Early Detection of Depression Indications in Text Sequences,” *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 3, pp. 588–601, 2020
- [38] S. Ghosal and A. Jain, “Depression and Suicide Risk Detection on Social Media using fastText Embedding and XGBoost Classifier,” *Procedia Comput. Sci.*, vol. 218, pp. 1631–1639, 2023
- [39] I. Ameer, M. Arif, G. Sidorov, H. Gómez-Adorno, and A. Gelbukh, “Mental Illness Classification on Social Media Texts using Deep Learning and Transfer Learning,” 2022
- [40] Munmun De Choudhury, Michael Gamon, Scott Counts, and Eric Horvitz. 2013. Predicting Depression via Social Media. *Proceedings of the International AAAI Conference on Web and Social Media* 7, 1 (2013), 128–137

- [41] Michael Mesfin Tadesse, Hongfei Lin, Bo Xu, and Liang Yang. 2019. Detection of Suicide Ideation in Social Media Forums Using Deep Learning. *Algorithms* 13, 1 (Dec. 2019), 7.
- [42] Sooji Han, Rui Mao, and Erik Cambria. 2022. Hierarchical attention network for explainable depression detection on Twitter aided by metaphor concept mappings. *arXiv preprint arXiv:2209.07494* (2022)
- [43] Johannes C Eichstaedt, Robert J Smith, Raina M Merchant, Lyle H Ungar, Patrick Crutchley, Daniel Preotiuc-Pietro, David A Asch, and HAndrew Schwartz. 2018. Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences* 115, 44 (2018), 11203–11208.
- [44] Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, and Johannes C Eichstaedt. 2017. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences* 18 (Dec. 2017), 43–49
- [45] Adrian Benton, Margaret Mitchell, and Dirk Hovy. 2017. Multi-task learning for mental health using social media text. *arXiv preprint arXiv:1712.03538* (2017)
- [46] Shailik Sarkar, Abdulaziz Alhamadani, Lulwah Alkulaib, and Chang-Tien Lu. 2022. Predicting depression and anxiety on reddit: a multi-task learning approach. In *2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 427–435
- [47] Gillian Cameron, David Cameron, Gavin Megaw, Raymond Bond, Maurice Mulvenna, Siobhan O'Neill, Cherie Armour, and Michael McTear. 2017. Towards a chatbot for digital counselling
- [48] Mostafa M. Amin, Erik Cambria, and Björn W. Schuller. 2023. Will Affective Computing Emerge from Foundation Models and General AI? A First Evaluation on ChatGPT
- [49] Bishal Lamichhane. 2023. Evaluation of ChatGPT for NLP-based Mental Health Applications