Mental Health Detection Using Texual Data Literature Survey

Sigilapally Akshay
School of Computer Science and Engineering
RV University
Bangalore, Karnataka
Sigilapallyakshaybtech24@rvu.edu.in

Shoveet Pal
School of Computer Science and Engineering
RV University
Bangalore, Karnataka
shoveetpalbtech24@rvu.edu.in

Samart S Bangalore
School of Computer Science and Engineering
RV University
Bangalore, Karnataka
samartsbangalorebtech24@rvu.edu.in

Priyangshu Mukherjee
School of Computer Science and Engineering
RV University
Bangalore, Karnataka
priyangshumukherjeebtech24@rvu.edu.in

***Abstract*—The detection of mental health conditions such as stress and depression is a critical public health priority. Leveraging machine learning and natural language processing (NLP) on survey-based textual data offers scalable, objective approaches for screening and early intervention. This survey synthesizes the latest developments in dataset selection, preprocessing, feature engineering, modeling, evaluation, and the challenges inherent to real-world deployment. Approaches suitable for academic reproduction with Python libraries are emphasized.**

Keywords—mental health, depression, stress, survey data, machine learning, natural language processing, SVM, BERT, Python

I. Introduction (*Heading 1*)

Depression and stress are leading causes of disability worldwide, with profound impacts on individual well-being and societal productivity [1]. Traditional diagnosis relies on clinical interviews and standardized surveys, but these methods are limited by human subjectivity and resource constraints. As a result, ML-based approaches for detecting mental health disorders from textual data are growing in popularity, providing cost-effective and scalable alternatives. Recent work demonstrates machine learning's capability in classifying mental health states using self-reported survey responses [2].

II. DATASETS AND SURVEY INSTRUMENTS

The validity of automated detection models depends heavily on the quality of survey instruments and datasets. Widely adopted tools include the Depression Anxiety Stress Scales (DASS-21) [3], which quantify states of depression, anxiety, and stress through 21 questions, and the PHQ-9, a nine-item depression screener [5]. The DAIC-WOZ corpus, offering conversational interviews along with depression severity scores, is widely used for benchmarking dialogue-based approaches [6]. Open-access resources like Kaggle further facilitate reproducibility [7]. Most datasets contain both raw text responses and categorical labels, enabling supervised ML tasks.

III. PREPROCESSING AND FEATURE EXTRACTION

Preprocessing is crucial for extracting meaningful features from free text. Steps include tokenization, stopword removal, lemmatization, and normalization [4, 8]. The NLTK [8] and spaCy [9] libraries are the most commonly used tools for these tasks in Python research environments. Survey text can be vectorized using Bag-of-Words or TF-IDF to quantify lexical content [4], while psycholinguistic dictionaries like LIWC provide psychologically interpretable features [10]. Recent studies advocate deeper representations using word embeddings (Word2Vec) [11], GloVe, or contextual transformers (BERT/RoBERTa) [12, 16]. Combining traditional features (text length, sentiment, readability scores) with modern embeddings often improves model robustness.

IV. MACHINE LEARNING ALGORITHMS AND ARCHITECTURES

A. Traditional Algorithms

Classical ML classifiers, including logistic regression, KNN, decision trees, SVMs, and random forests, are favored for their ease of implementation (scikit-learn) and interpretability [13]. Random forests exhibit resilience to class imbalance, a common trait in mental health datasets. SVMs, particularly with RBF kernels, are robust for smaller, structured datasets. Naive Bayes works well for shorter survey items due to its probabilistic treatment of text.

B. Deep Learning and Transformer

Deep learning models significantly improve accuracy, especially on larger or more nuanced datasets. CNNs and LSTMs can model local and sequential patterns in narrative responses. Transformer networks, especially BERT and RoBERTa, now dominate published results in survey and social media-based depression detection [12, 15, 17]. Advanced architectures utilize attention mechanisms to better capture long-range dependencies within text and even multimodal (text+audio/video) signals. Hugging Face Transformers have democratized access to these models, enabling easy fine-tuning and deployment on survey data [16].

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

C. Ensemble and Hybrid Approaches

Recent studies deploy ensemble strategies, combining multiple models for improved generalization and stability. Stacked generalization, where outputs of weak learners feed into a meta-classifier, often achieves state-of-the-art results on benchmark datasets.

V. EVALUATION METRICS AND EXPERIMENTAL DESIGN

Owing to prevalent class imbalance (e.g., far fewer depressed than non-depressed samples), accuracy alone is insufficient [14]. Precision, recall, and F1-score are standard, with AUC-ROC frequently reported for binary classification. Stratified k-fold cross-validation ensures robust error estimates across splits. Best practice involves validating models on external, out-of-distribution datasets whenever possible [18].

VI. LIMITATIONS, ETHICS, AND CHALLENGES

Research in this field faces several challenges. Small dataset sizes hamper generalization [19], while class imbalance and subtle language cues complicate learning [20]. Demographic and cultural differences can introduce bias. Ethics are paramount: privacy, informed consent, and the responsible communication of results are needed at every stage. Interpretability is vital in clinical and educational deployments—explaining predictions using SHAP, LIME, or inherently explainable models (e.g., decision trees) is encouraged [21]. Finally, model reproducibility and transparency in preprocessing and hyperparameter choices are essential for academic progress.

VII. FUTURE DIRECTIONS

New work explores multimodal depression detection, fusing text signals with context or sensor data (speech, facial expressions). Research on cross-lingual and culturally adaptive models aims to expand generalizability. Explainable AI and bias mitigation remain active areas of needed improvement [21, 22]. Open-source toolkits and collaborative benchmarking will continue to shape progress in this important domain.

REFERENCES

[1] World Health Organization, “Depression and other common mental disorders: Global health estimates,” Geneva, 2017. [Online]. Available: https://www.who.int/publications/i/item/depression-global-health-estimates

[2] S.A. Alison et al., “Early detection of mental health disorders using machine learning models,” Sci. Rep., vol. 15, p. 386, 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-00386-8

[3] S.H. Lovibond and P.F. Lovibond, Manual for the Depression Anxiety & Stress Scales, Psychology Foundation, Sydney, 1995. [Online]. Available: https://novopsych.com/assessments/depression/depression-anxiety-stress-scales-short-form-dass-21/

[4] A. Le Glaz et al., “Natural language processing in clinical neuroscience and psychiatry: systematic review,” Front. Psychiatry, vol. 13, p. 946387, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyt.2022.946387/full

[5] K. Kroenke et al., “The PHQ-9: validity of a brief depression severity measure,” J. Gen. Intern. Med., vol. 16, pp. 606–613, 2001. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1495268/

[6] J. Gratch et al., “The Distress Analysis Interview Corpus of human and computer interviews,” in Proc. LREC, 2014. [Online]. Available: https://www.semanticscholar.org/paper/The-Distress-Analysis-Interview-Corpus-of-human-and-Gratch-Artstein/ebb4c2e0cb5c5aa0ccdf8882f3607c79f3b00fe5

[7] Kaggle, “Mental Health Dataset,” 2024. [Online]. Available: https://www.kaggle.com/datasets/bhavikjikadara/mental-health-dataset

[8] S. Bird et al., Natural Language Processing with Python, O’Reilly, 2009. [Online]. Available: https://www.nltk.org/book/

[9] M. Honnibal and I. Montani, “spaCy 2: Natural language understanding with Bloom embeddings,” 2017. [Online]. Available: https://sentometrics-research.com/publication/72/

[10] J.W. Pennebaker et al., Linguistic Inquiry and Word Count: LIWC 2015. [Online]. Available: https://liwc.app/static/documents/LIWC2015%20Manual%20-%20Operation.pdf

[11] T. Mikolov et al., “Efficient estimation of word representations in vector space,” ICLR, 2013. [Online]. Available: https://arxiv.org/abs/1301.3781

[12] S. Senn et al., “Ensembles of BERT for depression classification,” IEEE EMBS Conf. Proc., pp. 4691–4694, 2022. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/36085764/

[13] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011. [Online]. Available: https://scikit-learn.org/stable/about.html

[14] U. Madububambachu, A. Ukpebor, and U. Ihezue, "Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review," Clin Pract Epidemiol Ment Health, vol. 20, p. e17450179315688, Jul. 2024. doi: 10.2174/0117450179315688240607052117. PMCID: PMC11443461. PMID: 39355197. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11443461/

[15] *W. Zhang, K. Mao, and J. Chen, "A Multimodal Approach for Detection and Assessment of Depression Using Text, Audio and Video," Phenomics, vol. 4, no. 3, pp. 234-249, Jun. 2024. doi: 10.1007/s43657-023-00152-8. PMCID: PMC11467147. PMID: 39398421. Available: https://pubmed.ncbi.nlm.nih.gov/39398421/*

[16] T. Wolf et al., “Transformers: State-of-the-art NLP,” in Proc. EMNLP, 2020. [Online]. Available: https://arxiv.org/abs/1910.03771

[17] T. Kallstenius et al., “Comparing traditional NLP and large language models for mental health classification,” Sci. Rep., vol. 15, 2025. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/40619512/

[18] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” J. Mach. Learn. Res., vol. 13, pp. 281-305, 2012. [Online]. Available: https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf

[19] Y. Cao et al., “ML approaches for mental illness detection on social media: Biases and challenges,” arXiv preprint 2410.16204, 2024. [Online]. Available: https://arxiv.org/abs/2410.16204

[20] D.A. Scherbakov et al., “NLP and social determinants in mental health research,” JMIR Mental Health, 2025. [Online]. Available: https://mental.jmir.org/2025/1/e67192

[21] C. Rudin, “Stop explaining black box ML models for high stakes decisions and use interpretable models instead,” Nat. Mach. Intell., vol. 1, no. 5, pp. 206–215, 2019. [Online]. Available: https://www.nature.com/articles/s42256-019-0048-x

[22] V. Aggarwal et al., “Leveraging LLMs for mental health: Detection from social discussions,” arXiv preprint 2503.01442, 2025. [Online]. Available: https://arxiv.org/abs/2503.01442