Migration Letters

Volume: 21, No: S7 (2024), pp. 73-80

ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)

www.migrationletters.com

Mental Disorder Classification And Prediction

[1]Ms.Nivedhitha G, [2]Dhanushya S, [3]Gayathri D, [4]Harismita K

ABSTRACT

While mental illness continues to cast a wide shadow over the world, its classification remains a complex dance. Thankfully, recent years have brought significant leaps forward in understanding and categorizing these conditions, fueled by powerful new diagnostic tools and computational approaches. This snapshot dives into the current state of mental health problem classification, highlighting exciting trends. Imagine machines learning from vast amounts of data, uncovering hidden patterns and boosting diagnostic accuracy like never before. These advancements hold the potential to revolutionize how we identify and address mental health challenges. This research introduces an innovative method for categorizing mental health disorders through machine learning techniques. The study utilizes a diverse dataset encompassing a range of mental health conditions, demographic details, and behavioral patterns. The primary objective is to develop a robust model capable of accurately categorizing individuals into different mental health groups based on their distinctive characteristics. The study addresses a classification challenge, aiming to differentiate individuals among Major Depressive Disorder (MDD), Obsessive-Compulsive Disorder (OCC), Anxiety, Post-Traumatic Stress Disorder (PTSD), sleeping problems, and loneliness. In this research, various ML Algorithms, including the Logistic Regression Algorithm and Random Forest Algorithm, are employed for the classification of mental health disorders. This endeavor aims to broaden the project's scope, showcasing additional capabilities while upholding accuracy.

Keywords: *Mental Health Prediction, Machine Learning, Depression, Logistic Regression, Random Forest.*

I INTRODUCTION

A "syndrome marked by a [1]clinically significant disruption in an individual's cognition, emotion regulation, or behavior, indicative of dysfunction in the psychological, biological, or developmental processes influencing mental functioning" (American Psychiatric Association, 2013). As per the Canadian Mental Health Association (2016), 20% of Canadians across diverse demographics have encountered mental illnesses at some point in their lives, with around 8% of adults having undergone episodes of severe depression [3]. Based on data from the WHO in 2014, around 20% of children and adolescents experienced mental disorders, with half of them onset before reaching the age of 14. Additionally, mental and substance use disorders accounted for roughly 23% of global deaths [3].

According to the WHO (2001), mental illness is the biggest cause of disability worldwide, with depression affecting almost 300 million people. Lifetime prevalence estimates range from 3% in Japan to 17% in the United States. In North America, the likelihood of having a major depressive episode within a year is about 3-5% for men and 8-10% for women (Andrade et al., 2003) [5]. Mental health diseases are recognised as a key source of disability on a global scale, with an estimated 970 million people suffering from various mental or brain disorders.

[1,2,3,4]Department of Computer Science & Engineering, Sri Krishna college of Technology, Coimbatore, Tamil Nadu, India.

Mental health issues have been a longstanding concern in human history, dating back to the fifth century BC. Nevertheless, in the contemporary setting, the prevalence of these issues is on the rise. Government statistics indicate that approximately 130 million individuals in India, constituting a significant portion of the population, may be grappling with various forms of mental illness [11]. The primary cause behind this substantial number of people experiencing mental health challenges is attributed to the deteriorating healthcare system and inadequate governmental support in addressing this issue. In India, mental health remains a sensitive and stigmatized subject, leading to a mere 8 to 10% of individuals seeking treatment for their issues, leaving a significant portion unaddressed and potentially contributing to elevated suicide rates. Medical professionals have identified that more than 35% of individuals seeking medical assistance may be grappling with conditions such as depression, post- traumatic stress disorder (PTSD), anxiety, insomnia, bipolar disorder, and other mental health disorders [11]. Affordability issues further compound the challenge, exacerbating the overall problem. According to data from the World Health Organization, India has 0.75 psychologists and psychiatrists per 100,000 inhabitants, whereas Argentina boasts 106 psychologists for every 100,000 people. Addressing the potential mental health epidemic requires decisive and crucial steps from the government in healthcare, including the allocation of a substantial budget to mental health.

II BACKGROUND STUDY

Mental health concerns, as highlighted by Satvik Gurjar et al [11],stand out as a crucial focus within the healthcare domain in the twenty-first century. A primary contributor to this issueis the pervasive lack of awareness. Thisarticle aims to raise awareness about potential mental health challenges, such as depression, anxiety, PTSD, or insomnia,by utilizing machine learning to identify symptoms. To implement machine learning algorithms, a survey form incorporating questions akin to those employed by psychologists to gain acomprehensive understanding of patients' issues was employed to gather data from individuals of diverse ages, professions, genders, and life styles.

Glen Coppersmith. et al.[7] emphasized the predominant reliance of traditional mental health research on information obtained through direct interactions with healthcare professionals. While recent studies have highlighted the effectiveness of utilizing social media datato examine depression, there is a notable scarcity of research on other mental health conditions. Our focus centers on post- traumatic stress disorder (PTSD), a debilitating condition affecting millions globally, particularly prevalent among military veterans [7]. In addition, we offer a novel technique to developing a PTSD classifier for social media, employing simple searches of publicly available Twitter data. When compared to earlier research efforts, this strategy greatly minimises the cost of training data. Wevalidated its effectiveness by comparingdifferences in language usage between those with PTSD and people who werechosen at random. We develop classifiers to distinguish between these two categories and use them to identify heightened cases of PTSD on and around US military bases.

Although mental health has grown in relevance in the medical sector, its examination is difficult due to privacy issues and the lack of objectively quantifiable assessments such as lab tests or physical exams [1]. The available statistics on mental health is primarily reliant on patients' subjective accounts of their experiences, which are mostly expressed in written form. Online platforms and direct input from patients, including various kinds of social media, are primary sources of this textual data. We use datasets supplied by the CLPsych shared tasks in 2016 and 2017, which come from classified ReachOut online forum postings that have been rigorously categorised based on mental health severity [1].

Using a combination of machine and deep learning approaches, we created an automatic severity labelling system. Our solution combines supervised and semi-supervised embedding approaches, using labelled and unlabeled data from ReachOut as well as unlabeled data from WebMD corpora [1]. Posts were categorized into four groups

(green, amber, red, and crisis) based on metadata, syntactic, semantic, and embedding criteria. The resulting systems, incorporating the mentioned features, demonstrated superior performance compared to state-of-the-art systems developed on the ReachOut dataset. Specifically, our systems achieved maximum micro-averaged F-scores of 0.86 and 0.80 for the Clinical Psychology (CP) 2016 and 2017 test datasets, respectively. In preparation for the 2016 Computational Linguistics and CP Shared Task, we offer our approach for judging the severity of user posts on a mental health forum. Our submission includes a meta-classifier that employs a collection of base classifiers built from lexical, syntactic, and metadata aspects [1]. These classifiers are intended for both the focal points and their surroundings, and they include both preceding and following postings. The outputs of these classifiers were employed in the training of a meta-classifier, surpassing the performance of each individual classifier as well as an ensemble classifier. Subsequently, this meta- classifier was expanded into a Random Forest of meta-classifiers, leading to improved classification accuracy. We achieved impressive results, securing the first position among the 60 entries submitted in the competition [1].

III PROBLEM DEFINITION

Developing models and systems to classify individuals according to their mental health status and anticipate the probability of experiencing specific mental disorders plays a crucial role in categorizing and predicting mental health conditions. This aspect is pivotal in the realm of mental healthcare, as it facilitates early intervention, personalized treatment strategies, and efficient resource allocation. Developing an effective system for categorizing individuals into specific mental health groups, considering their symptoms, behavior, and other relevant criteria, can significantly enhance personalized mental healthcare, facilitate early intervention, and overall, improve mental health outcomes. The successful implementation of such a system aims to reduce the stigma associated with mental health disorders while concurrently enhancing the effectiveness of mental health care and treatment.

IV MATERIALS AND METHODS

Predicting mental disorders involves employing machine learning algorithms that leverage data to create models capable of forecasting mental health outcomes. In this study, three crucial ML algorithms—LR and RF— were utilized for predicting mental health. The Kaggle dataset, encompassing measurements of pollutants across various environments, served as the foundation for training and assessing our models.

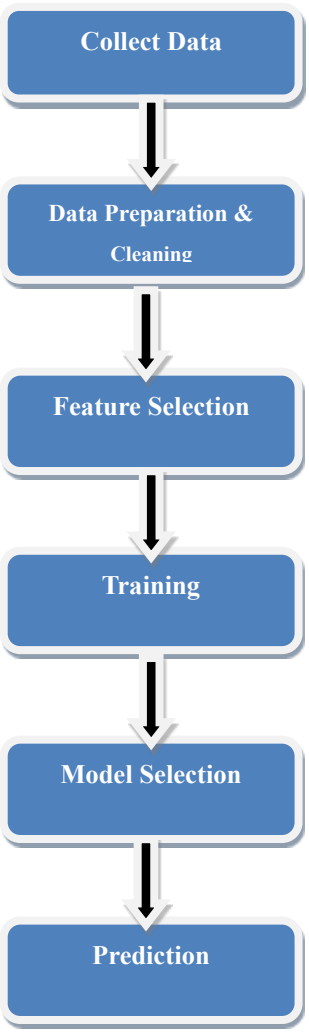

Figure 1: Proposed Workflow architecture

4.1 Dataset collection

The dataset was acquired from the Kaggle website and is accessible at https://www.kaggle.com/datasets/baselbakeer/mental-disorders-dataset. We have developed an extensive dataset that encompasses symptoms related to various mental disorders, including ADHD, OCD, PTSD, and more. As we all know, timely identification of these disorders is critical for improving the quality of life for many people.

4.2 Dataset preprocessing

The data for the project on mental conditions was sourced from Kaggle, a platform offering a variety of statistics. The dataset, stored in a CSV file, encompasses features like ADHD, OCD, PTSD, among others, serving as the project's labels. The data was imported using the pandas package to manage any discrepancies. Data that is missing, incomplete, or corrupted might skew operations like count, average, and mean, demanding its removal before any data analysis. Cleaning involved deleting confusing values to ensure the accuracy of subsequent analyses.

4.3 Mental health disorder using Logistic Regression and Random Forest

This research aims to forecast mental health outcomes by employing two Machine Learning (ML) methodologies, namely Logistic Regression and Random Forest. The dataset, sourced

from Kaggle, encompasses various conditions such as ADHD, OCD, PTSD, and others. Following a comprehensive preparation phase addressing missing values, outlier detection, data consistency checks, and feature engineering, the dataset is meticulously refined for training and testing the models. The Random Forest technique is employed to predict the association between mental health and the provided characteristics. Due to its optimization for binary classification tasks, this approach contributes to a better understanding of how recovery impacts mental health. Leveraging machine learning algorithms stands out as a highly efficient approach for improving prediction accuracy. This comparative analysis aims to provide insights into the effectiveness of Logistic Regression and Random Forest in predicting mental health (MH), contributing valuable information to enhance the classification of mental health through machine learning techniques.

4.4 Random Forest

The RF algorithm holds significant prominence in machine learning as it addresses both classification and regression tasks, which are fundamental aspects of this field. Operating on
supervised learning principles, Random Forest is widely acknowledged for its effectiveness in solving classification and regression challenges. This algorithm leverages ensemble learning, a technique that amalgamates multiple classifiers to address intricate problems, thereby enhancing overall model performance. Accuracy in Random Forest is determined by dividing the total accurately predicted cases by the total number of instances in the test set.

Precision (P) is the ratio of true positive predictions to the total predicted positives.
Recall (R) is the ratio of true positive predictions to the total actual positives.
F1-score is the harmonic mean of precision and recall.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of prediction}}$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{F1} - \text{Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Final Prediction} = \frac{1}{N} \sum_{i-1}^{N} \text{Tree}$$

where N is the number of trees, Tree_i is the prediction from the i-th tree, and X is the input data. In this case, we'll refer to W as the independent variable and Y as the dependent variable. For this reason, the most basic random forest model's equation for a straight line is:

$$Y = \gamma 0 + \gamma 1\, W + \varepsilon$$

Regression coefficients, where 0 denotes the intercept and 1 the slope of the regression line, constitute the random errorcomponent.

Thus for n observations, $yi = \gamma 0 + \gamma 1\, wi + i, i = 1, 2 \ldots n.$

Here, yi and I are treated as random variables, while wi values are fixed. Equation provides a regression model that is valid under the following conditions.

V RESULTS AND DISCUSSION

The core of the research lies in the Results and Discussion section, where the study's findings are presented and meticulously analyzed. This part unveils essential facts, patterns, and insights uncovered during a comprehensive investigation. The ensuing discussion

interprets these findings in the context of existing literature, theoretical frameworks, and the overall objectives of the study.

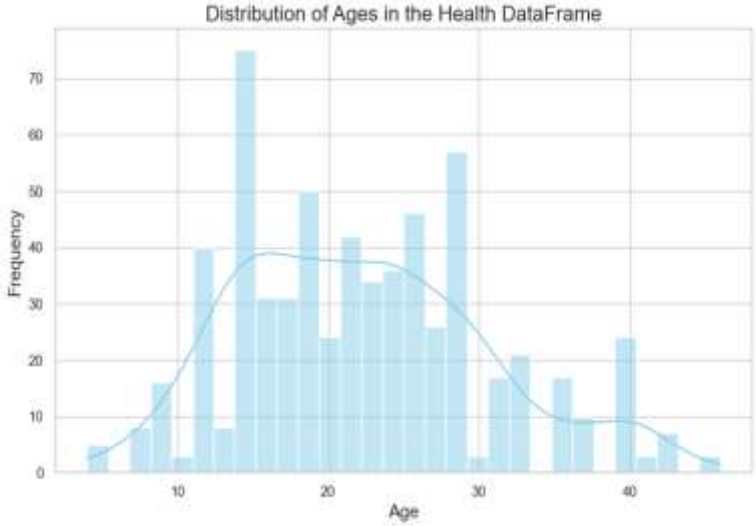

Figure 2: Age wise Distribution

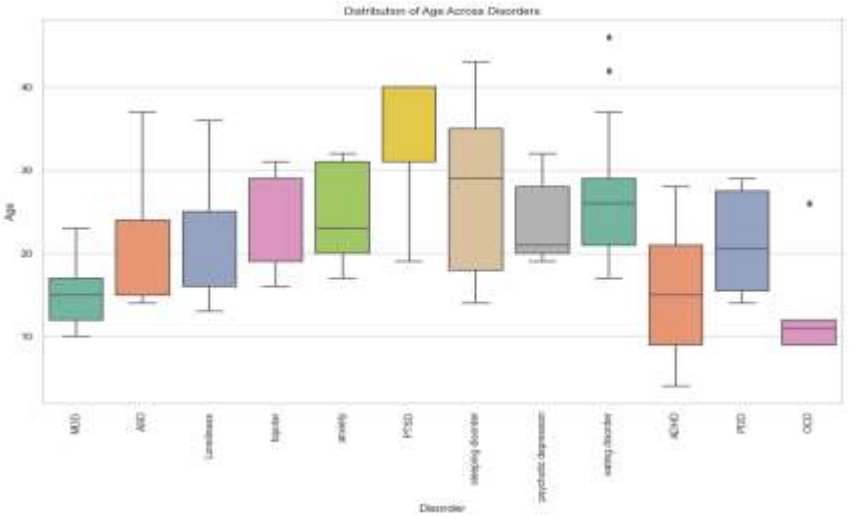

Figure 3: Age wise disorder

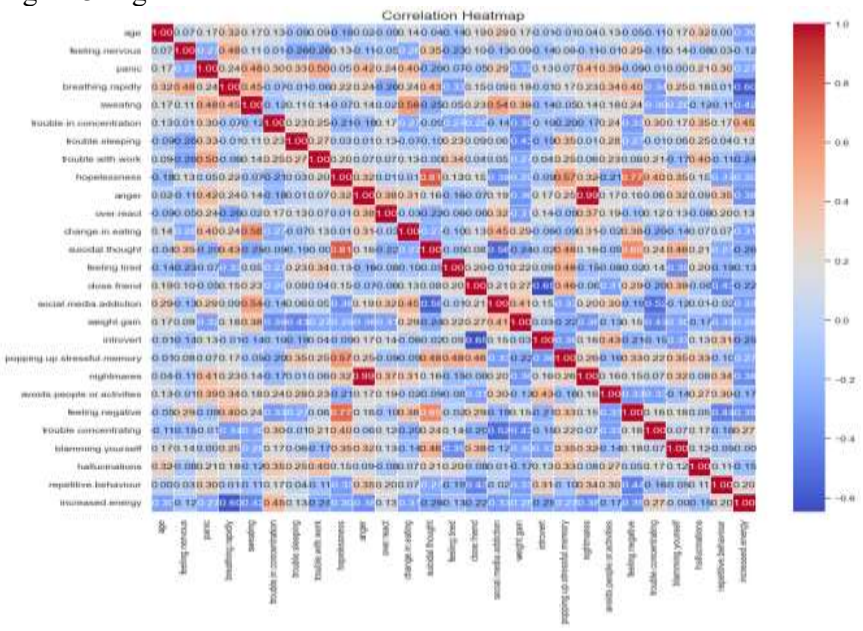

Figure 4: Correlation Heatmap

Figure 5: comparison chart

Table 1 and Figure 5 showcase a notable enhancement in performance through the recommended "ML classification" approach compared to established methods, particularly RF and LR. The proposed ML classification achieved a significantly superior accuracy of 98.95%, surpassing RF (98.95%) and LR (97.35%). Precision, recall, and F- measure metrics all demonstrated superior performance with the ML classification. These results suggest that the enhanced ML classification strategy surpasses existing methods, underscoring its potential for more precise and reliable predictions in the assessed task. The observed improvements across various metrics underscore the efficacy of the suggested method in enhancing the overall performance of the categorization task.

	Algorithm	**Accuracy**	**Precision**	**Recall**	**F-measure**
Existing methods	LSTM	92.41	88	85	87
Proposed methods	RF	98.95	99.04	98.95	98.94

Table 1: Performance metrics comparison table

VI CONCLUSION

In our study, we employed the capabilities of Random Forest, a robust machine learning technique, to address the crucial task of categorizing and predicting mental diseases. The Kaggle dataset, encompassing information on mental health conditions, served as a comprehensive basis for training and assessing our models. Random Forest exhibited several advantages in this context, such as robustness, interpretability, and the capacity to manage intricate, non-linear associations within the data. Throughout our experiments, Random Forest demonstrated high accuracy in classifying individuals into different mental health categories. The algorithm's resilience was demonstrated by its adept handling of diverse datasets characterized by varying features and complexities. In contrast, the suggested Random Forest (RF) approach surpasses all existing methods in terms of accuracy (98.95%), precision (99.04%), and F- measure (98.94%). The interpretability of the Random Forest proves instrumental in identifying crucial factors contributing to the classification of mental disorders. This newfound knowledge serves as a valuable resource for mental health providers, enabling targeted intervention and personalized treatment planning. It is crucial to note that the success of the Random Forest model hinges on the quality and representativeness of the training data. Ensuring the model's ability to generalize requires addressing biases within the dataset and promoting diversity. Employing Random Forest for the classification and prediction of mental disorders holds

promise for advancing personalized mental healthcare. Although implementing this approach may pose challenges, the potential advantages, such as early intervention, tailored treatment, and enhanced patient outcomes, underscore the value of integrating data science and mental health care.

REFERENCES

[1] B. G. Patra, R. Kar, K. Roberts, and H. Wu, "Mental health severity detection from psychological forum data using domain-specific unlabelled data," in Proc. AMIA Summits Transl. Sci., 2020, p. 487.

[2] A. H. Yazdavar, M. S. Mahdavinejad, G. Bajaj, K. Thirunarayan, J. Pathak, and A. Sheth, "Mental health analysis via social media data," in Proc. IEEE Int. Conf. Healthcare Informat. (ICHI), Jun. 2018, pp. 459–460, doi: 10.1109/ICHI.2018.00102.

[3] A. H. Orabi, P. Buddhitha, M. H. Orabi, and D. Inkpen, "Deep learning for depression detection of Twitter users," in Proc. 5th Workshop Comput. Linguistics Clin. Psychol., From Keyboard Clinic, 2018, pp. 88–97.

[4] Kroenke K, Spitzer RL, Williams JB, Linzer M, Hahn SR, deGruy III FV, Brody D. Physical symptoms in primarycare: predictors of psychiatric disorders and functional impairment. Archives of family medicine, 1994; 3(9):774.

[5] De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. In Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media, 2013 (pp. 128–137).

[6] Malmasi S, Zampieri M, Dras M. Predicting post severity in mental health forums. In Proceedings of the third workshop on computational lingusitics and clinical psychology, 2016 (pp. 133–137).

[7] Glen Coppersmith, Mark Dredze, and Craig Harman. 2014a. Measuring Post Traumatic Stress Disorder in Twitter. In In Proceedings of the 7th International AAAI Conference on Weblogs and Social Media (ICWSM)., volume 2, pages 23–45.

[8] Glen Coppersmith, Mark Dredze, Craig Harman, Hollingshead Kristy, and Margaret Mitchell. 2015b. CLPsych 2015 Shared Task: Depression and PTSD on Twitter. In Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, pages 31–39.

[9] A. Sharma, I. W. Lin, A. S. Miner, D. C. Atkins, and T. Althoff, "Towards facilitating empathic conversations in online mental health support: A reinforcement learning approach," in Proc. Web Conf., Apr. 2021, pp. 194–205, doi: 10.1145/3442381.3450097.

[10] M. De Choudhury, S. S. Sharma, T. Logar, W. Eekhout, and R. C. Nielsen, "Gender and cross-cultural differences in social media disclosures of mental illness," in Proc. ACM Conf. Comput. Supported Cooperat. Work Social Comput., Feb. 2017, pp. 353–369. [Online]. Available: http://dl.acm.org/citation.cfm?id=2998220.

[11] Satvik Gurjar, Chetna Patil, Ritesh Suryawanshi, Madhura Adadande, Ashwin Khore, Noshir Tarapore "Mental Health Prediction Using Machine Learning" Volume: 09 Issue: 12 | Dec 2022, IRJET.