© September 2025| IJIRT | Volume 12 Issue 4 | ISSN: 2349-6002

Prediction And Detection of Future Mental Disorders Using Social Media Data

D Sravanthi[1], D.Murali[2]
[1]*PG Student, QUBA College of engineering and technology*
[2]*Associate Professor, QUBA College of engineering and technology*

Abstract: Mental disorders affect millions of people worldwide, significantly impacting their thoughts, emotions, and behaviours. These conditions often go undetected until they become severe, making early detection a critical challenge. Timely identification, however, opens the door to early intervention, which can help individuals receive support before their condition worsens. One promising approach to address this challenge involves analysing how people express themselves on social media. These platforms have become digital diaries, where users often share their feelings, experiences, and emotional states. By examining these emotional cues, researchers can uncover patterns that may indicate underlying mental health conditions. In this study, we focus on two computational representations aimed at capturing emotional presence and variability in social media posts. These representations help in modelling not just what users feel, but also how their emotions change over time. We tested our models using two recent public datasets focused on Depression and Anorexia. The findings reveal that emotional signals extracted from user posts can effectively highlight individuals at risk. When both representations are combined, the model's performance improves significantly. It matches the top-performing method for depression detection and trails the best anorexia model by just 1%.A key benefit of these emotional models is their interpretability. Unlike black-box deep learning methods, they offer clearer, more understandable insights.

Index Terms— **Mental health, Social media analysis, Depression detection, Anorexia detection, Emotion variability, Computational representation, Early detection, Interpretability, Natural Language Processing (NLP), Machine Learning**

1. INTRODUCTION

A mental disorder causes different interferences in the thinking and behavior of the affected person [1]. These interferences could vary from mild to severe, and could result in an inability to live routines in daily life and ordinary demands [2]. Common mental disorders such as depression and anorexia affect millions of people around the world. They may be related to a single incident causing excessive stress on the person or by a series of different stressful events. It is also well known that mental disorders tend to increase in countries experiencing generalized violence or recurrent natural disasters. For example, in 2018 a study of mental disorders in Mexico revealed that 17% of its population has at least one mental disorder and one in four will suffer a mental disorder at least once in their life [3]. In another vein, in the modern world, we take for granted that social life could be experienced either in the physical world or in a virtual world created by social media platforms like Face book, Twitter, Redd it, or similar platforms. This reality presents some challenges, but also great opportunities which, if properly addressed, could contribute to the understanding of what and how we communicate. In this regard, the goal of this study is to analyze, via the automatic identification of emotional patterns, social media documents 1 with the purpose of detecting the presence of signs of depression or anorexia in the population of that area [4]–[6]. Previous works have addressed the analysis of emotions of social media users by paying attention to their contrast and tone. They have mainly applied this analysis to predict users' age and gender as well as a range of sensitive personal attributes including sexual orientation, religion, political orientation [7], [8], income [9], and personality traits [10], [11]. According to these studies, the analysis of emotions in social media allows capturing important information related to users.

1.1. PROBLEM OVERVIEW

This research aims in analyzing social media data to detect and predict potential future mental disorders in individuals. By examining linguistic patterns, emotional expressions, and user behavior on platforms like Twitter, Reddit, and Instagram, the system aims to identify early warning signs of mental health issues. Utilizing machine learning algorithms, the project seeks to develop a predictive model capable of classifying and forecasting mental health conditions with high accuracy. The system is designed to work in real-time, continuously monitoring social media activity to flag at-risk individuals.

1.2 PROBLEM STATEMENT

Mental health disorders are increasingly affecting individuals worldwide, often going undetected until they become severe. Traditional methods of diagnosis rely heavily on self reporting and clinical observation, which may delay timely intervention. With the rise of social media, individuals frequently express their thoughts, emotions, and behaviors online, offering a potential window into their mental state.

II SYSTEM ANALYSIS

2.1 LITERATURE SURVEY

[1] "The burden of depressive illness, Public Health Perspectives on Depressive Disorders " R. Kessler, E. Bromet, P. Jonge, V. Shahly, and Marsha., 2017.

Depressive disorders represent one of the most significant public health challenges globally, both in terms of prevalence and the burden they place on individuals and healthcare systems. The work of Kessler et al. (2017) titled "The Burden of Depressive Illness" offers a comprehensive examination of the magnitude, persistence, and multifaceted impact of depressive disorders across populations. Their research synthesizes epidemiological findings, global health statistics, and clinical insights to underline the seriousness of depression as a chronic and often disabling condition.

[2]" "What about mood swings? identifying depression on twitter with temporal measures of emotions,"" C. Xuetong, D. Martin, W. Thomas, and E. Suzanne, 2019.

The World Health Organization's (WHO) 2019 Mental Health Fact Sheet provides a concise yet impactful overview of the global mental health landscape, emphasizing the rising concern over the increasing prevalence of mental health disorders and the urgent need for comprehensive action. The fact sheet consolidates key data and trends related to common mental illnesses such as depression, anxiety, schizophrenia, and bipolar disorder, while also outlining the broader public health implications of untreated mental illness.

III SYSTEM ANALYSIS

System analysis is the process of examining a system to identify its components, how they interact, and how they can be improved. It involves understanding user requirements and evaluating current processes and technologies. The goal is to create efficient and effective solutions to meet organizational needs. Analysts gather data, identify problems, and design improvements. This step is crucial for successful system development and implementation.

3.2.SYSTEM REQUIREMENTS SPECIFICATIONS

The System Requirements and Specifications section outlines the essential hardware, software, and functional components necessary for the successful development and operation of the system. It serves as a blueprint for developers, ensuring clarity and alignment with project goals. This section defines both the user and system-level expectations. It also highlights any constraints or limitations that may affect the design. Overall, it forms the foundation for implementation and future enhancements.

3.2.1. HARDWARE REQUIREMENTS

- Processor : Pentium –IV
- RAM : 4 GB (min)
- Hard Disk : 20 GB
- Key Board : Standard Windows Keyboard
- Mouse : Two or Three Button Mouse
- Monitor : SVGA

3.2.2. SOFTWARE REQUIREMENTS

- Operating system : Windows 7 Ultimate.
- Coding Language : Python.
- Front-End : Python.
- Back-End : Django-ORM
- Designing : HTML , CSS , JavaScript.
- Data Base : MySQL (WAMP Server).

IV. SYSTEM DESING

4.1. UML DIAGRAMS SYSTEM DESIGN UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group. The goal is for UML to become a common language for creating models of object-oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML. The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems. The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems. The UML is a very important part of developing objects- oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

GOALS
The Primary goals in the design of the UML are as follows

- Provide users a ready-to-use, expressive visual modeling Language so that they can develop and exchange meaningful models.
- Provide extendibility and specialization mechanisms to extend the core concepts.
- Be independent of particular programming languages and development process.
- Provide a formal basis for understanding the modeling language.
- Encourage the growth of OO tools market.
- Support higher level development concepts such as collaborations, frameworks, patterns and components.
- Integrate best practices.

4.1.1. USE-CASE DIAGRAM A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

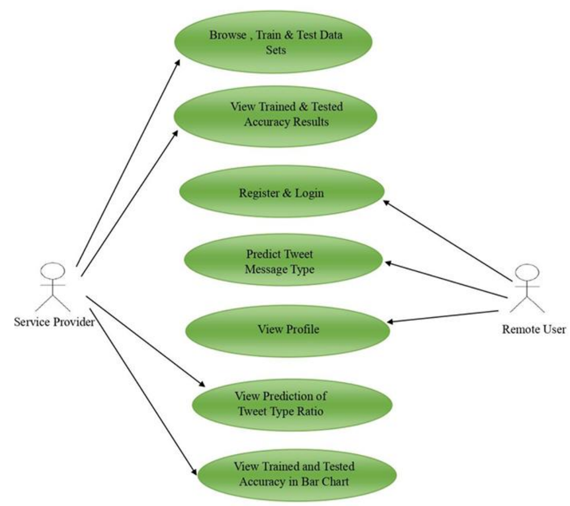

Fig.4.1.1. *Use-Case Diagram*

V SYSTEM TESTING

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirements.

PHASES OF SYSTEM TESTING

A video tutorial about this test level. System testing examines every component of an application to make sure that they work as a complete and unified whole. A QA team typically conducts system testing after it checks individual modules with functional or user-story testing and then each component through integration testing. If a software build achieves the desired results in system testing, it gets a final check via acceptance testing before it goes to production, where users consume the software. An app-dev team logs all defects, and establishes what kinds and amount of defects are tolerable.

5.1. UNIT TESTING

`Unit Testing is a software testing technique where individual components or functions of a system are tested in isolation to ensure they work as expected. Each unit is tested independently to verify its correctness, reliability, and performance. It helps identify bugs early in the development cycle, making debugging easier and faster. Unit tests are typically automated and written by developers during the coding phase. This process improves code quality and facilitates future code changes or integration.

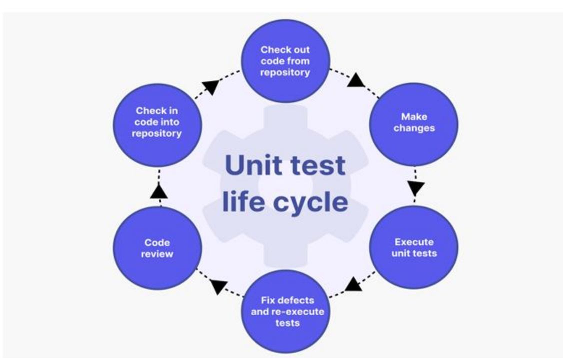

Fig.5.1. *Unit Testing*

5.2. STRUCTURAL TESTING

Structural Testing are designed to It is not possible to effectively test software without running it. Structural testing, also known as white-box testing, is required to detect and fix bugs and errors emerging during the pre-production stage of the software development process. At this stage, unit testing based on the software structure is performed using regression testing. In most cases, it is an automated process working within the test automation framework to speed up the development process at this stage. Developers and QA engineers have full access to the software's structure and data flows (data flows testing), so they could track any changes (mutation testing) in the system's behavior by comparing the tests' outcomes with the results of previous iterations (control flow testing).

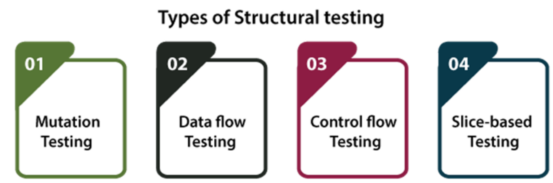

Fig.5.2. *Structural Testing*

5.3. BLACK BOX TESTING

The final stage of testing focuses on the software's reactions to various activities rather than on the mechanisms behind these reactions. In other words, blackbox testing, also known as black-box testing, presupposes running numerous tests, mostly manual, to see the product from the user's point of view. QA engineers usually have some specific information about a business or other purposes of the software ('the black box') to run usability tests, for example, and react to bugs as regular users of the product will do. Black box testing also may include automation (regression tests) to eliminate human error if repetitive activities are required.

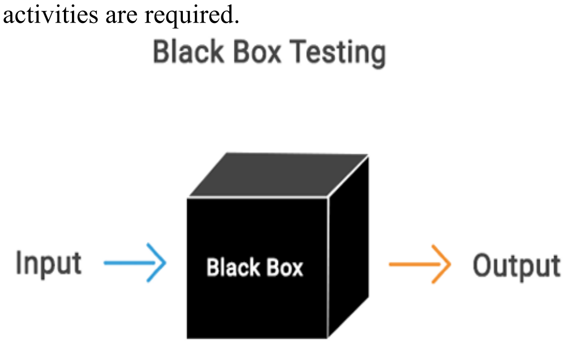

Fig.5.3. *Black Box Testing*

VI. IMPLEMENTATION AND RESULTS

6.1. ALGORITHMS

Decision Tree: A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes.

Navie Bayes: The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature.

OUTPUT SCREENS

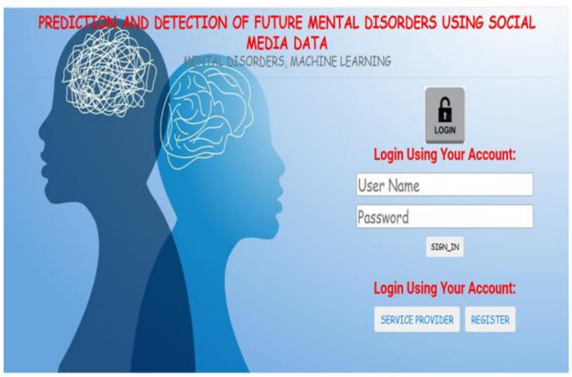

Fig.6.5.1. *Home page*

VII CONCLUSION

7.1. CONCLUSION

In this work, we showed that representations based on fine grained emotions can capture more specific topics and issues that are expressed in social media documents by users that unfortunately experience depression or anorexia. That is, the automatically extracted sub-emotions present useful information that helps the detection of these two mental disorders. On the one hand, the BOSE representation obtained better results than the proposed baselines, including some deep learning approaches, and also improved the results of only using broad emotions as features. On the other hand, the inclusion of a dynamic analysis over the sub emotions, called _-BOSE, improved the detection of users that presents signs of anorexia and depression, showing the usefulness of considering the changes of sub-emotions over time. It is worth mentioning the simplicity and interpretability of both representations, then creating a more straightforward analysis of the results.

7.2. FUTURE SCOPE

The future scope of this system is vast, with potential for integration into mental healthcare platforms for early intervention and support. As AI and natural language processing technologies evolve, prediction models can become more accurate and context-aware. With better data privacy frameworks, such systems can be ethically deployed at scale, aiding mental health professionals in proactive diagnosis. Integration with wearable devices and real-time monitoring can further enhance predictive capabilities. Ultimately, this approach could revolutionize how mental health issues are detected, managed, and prevented in digital society.

REFERENCE

[1] R. Kessler, E. Bromet, P. Jonge, V. Shahly, and Marsha., “The burden of depressive illness,” Public Health Perspectives on Depressive Disorders, 2017.

[2] W. H. Organisation, “Mental health: Fact sheet,” https://www.euro.who.int/en/health topics/noncommunicablediseases/ mental-health, 2019.

[3] M. Renteria-Rodriguez, “Salud mental en mexico,” NOTA-INCyTUNU´MERO 007, 2018.

[4] S. Guntuku, D. Yaden, M. Kern, L. Ungar, and J. Eichstaedt, “Detecting depression and mental illness on social media: an integrative review,”Current Opinion in Behavioral Sciences, 2017.
[5] J. Pestian, H. Nasrallah, P. Matykiewicz, A. Bennett, and A. Leenaars, “Suicide note classification using natural language processing: A content analysislin heidelberg,” Biomed Inform Insights, 2010.
[6] P. Chikersal, D. Belgrave, G. Doherty, A. Enrique, J. E. Palacios, D. Richards, and A. Thieme, “Understanding client support strategies to improve clinical outcomes in an online mental health intervention,” In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020.
[7] M. Kosinski, D. Stillwell, and T. Graepel, “Private traits and attributes are predictable from digital records of human behavior,” Proceedings of the national academy of sciences, 2013.
[8] M. Kosinski, Y. Bachrach, P. Kohli, D. Stillwell, and T. Graepel, “Manifestations of user personality in website choice and behaviour on online social networks,” Machine learning, 2014.
[9] D. Preot¸iuc-Pietro, S. Volkova, V. Lampos, Y. Bachrach, and N. Aletras, “Studying user income through language, behaviour and affect in social media.,” PloS one 10.9, 2015.
[10] T. Correa, A. Willard Hinsley, and H. G. De Zuniga, “Who interacts on the web?: The intersection of users’ personality and social media use.,” Computers in human behavior 26.2, 2010.
[11] S. Volkova and Y. Bachrach, “Inferring perceived demographics from user emotional tone and user environment emotional contrast,” Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2016.
[12] D. Ram´ırez-Cifuentes and A. Freire, “Upf’s participation at the clef erisk 2018: Early risk prediction on the internet,” Proceedings of the 9th International Conference of the CLEF Association, CLEF 2018, Avignon, France, 2018.
[13] H. Schwartz, J. Eichstaedt, M. Kern, G. Park, M. Sap, D. Stillwell, M. Kosinski, and L. Ungar, “Towards assessing changes in degree of depression through facebook,” In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality, 2014.
[14] G. Coopersmith, M. Dredze, and C. Harman, “Quantifying mental health signals in twitter,” Workshop on Computational Linguistics and Clinical Psychology, 2014.
[15] C. Xuetong, D. Martin, W. Thomas, and E. Suzanne, “What about mood swings? identifying depression on twitter with temporal measures of emotions,” Companion Proceedings of the The Web Conference2018, International World Wide Web Conferences Steering Committee, pp. 1653–1660, 2018.