Azure-Powered Hybrid Sentiment Analysis for Mental Health Assessment: A Two-Layer Framework Integrating Clinical Text Classification and Sarcasm Processing

Thamizh Mani R[1], Dr. Vikram Palimar[1*], Shashank Singh[2], Dazzle A J[2], Dr. R. Sujithra Kanmani[2], Dr. Mamatha Shivanandha Pai[3], Dr. Shwetha T S [4], Dr. Nirmal Krishnan M[5], Dr. Ramana Vinjamuri[6], Dr. Lata Satyen[7]

[1] Department of Forensic Medicine and Toxicology, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, India.

[2] School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, 600127, India.

[3] Department of Child Health, Manipal College of Nursing, Manipal Academy of Higher Education, Manipal, India.

[4] Department of Clinical Psychology, Manipal College of Health Professions, Manipal Academy of Higher Education, Manipal, India.

[5] Department of Forensic Medicine, ESI Medical College and PGIMSR, Gulbarga, Karnataka, India.

[6] Vinjamuri Lab, Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD, United States of America (USA).

[7] School of Psychology, Deakin University, 221 Burwood Hwy, Burwood, VIC, 3125, Australia.

*__Corresponding author. E-mail__: vikram.palimar@manipal.edu;

Abstract

In today's digital landscape, the expression of emotions has evolved beyond traditional text, incorporating emojis, sarcasm, and implicit cues that pose a challenge for conventional sentiment analysis systems. This study proposes hybrid architecture leveraging Azure's sentiment analysis and custom text classification tools to perform a two-layer sentiment and mental health assessment. The model begins by classifying user-submitted sentences into positive, neutral, or negative categories using Azure Sentiment Analysis. Sentences identified as negative are further analyzed by Azure Custom Text Classification to detect deeper clinical sentiments related to mental health. Our system excels at interpreting nuanced expressions, including emojis and sarcastic phrases, ensuring more accurate sentiment detection and response. A key feature is its user complaint tracking system, which is designed to identify recurring patterns and provide technicians with past responses for rapid resolution. All data is stored securely for future clinical referencing and trend analysis. This work offers comparative analysis with pre-existing models, showcasing superior accuracy and contextual understanding. The results affirm that our architecture enhances the precision of sentiment detection and offers practical benefits for early detection and mental health support, making it a valuable advancement in intelligent emotional computing and healthcare AI solutions.

Keywords: Sentiment Analysis, Mental Health Assessment, Sarcasm Processing, Azure Text Classification, Depression Detection, Stress Detection, Natural Language Processing, Emoji Processing

1. Introduction

Mental health disorders, including depression, anxiety, and chronic stress, have emerged as among the most pressing public health challenges of the 21st century. According to the World Health Organization, over 264 million people suffer from depression, and approximately 284 million are affected by anxiety disorders globally, with rising trends observed across both developed and developing nations [1]. These conditions are associated with a wide range of negative outcomes, such as diminished quality of life, physical comorbidities, substance abuse, and elevated suicide risk. Despite growing recognition of mental health's significance, disparities in access, affordability, and quality of care remain prominent [2].

The burden of mental illness is not equally distributed. In low- and middle-income countries, access to diagnosis and treatment is further limited by infrastructure gaps, persistent stigma, and a critical shortage of trained mental health professionals. In regions such as sub-Saharan Africa, up to 90% of individuals with mental health conditions may receive no formal care [3]. Adolescents and the elderly are particularly vulnerable: adolescents often experience complex posttraumatic stress disorder (CPTSD) triggered by adverse experiences, while elderly individuals are susceptible to depressive and anxiety symptoms exacerbated by social isolation and declining physical health.

Concurrently, digital communication has transformed how individuals express and seek help for emotional challenges. Platforms such as social media, online forums, and messaging services have become central venues for self-disclosure and peer support[5]. These digital interactions generate rich, real-time insights into psychological states, but also pose significant challenges for analysis due to the online language's scale, speed, and linguistic complexity. Traditional assessment methods—relying on self-reports or clinician-administered tools—are ill-equipped to process these dynamic and unstructured data sources, underscoring the need for innovative computational approaches.

Digital mental health interventions have evolved rapidly in response to these challenges. Recent studies highlight the potential of artificial intelligence (AI) and mobile health (mHealth) technologies to augment and streamline behavioral healthcare, particularly in settings where demand for therapy far outstrips provider availability [25]. AI-powered tools, including conversational agents and predictive algorithms, are increasingly integrated into clinical workflows, offering scalable solutions for early detection, self-management, and personalized intervention [27, 30]. For example, AI-based conversational agents have demonstrated promise in delivering evidence-based support and reducing barriers to care. At the same time, deep learning predictive algorithms have achieved high accuracy in identifying depression from remote data sources [34].

Despite these advances, the adoption and effectiveness of digital mental health tools are influenced by user engagement, acceptability, and the ability to address diverse needs. Pregnant adults, for instance, have expressed strong interest in mHealth interventions for improving physical and mental health during and after pregnancy, highlighting the importance of user-centered design and community-informed approaches [26]. Similarly, adults with attention-deficit/hyperactivity disorder (ADHD) have benefited from teletherapy features such as automated reminders and flexible scheduling, which address executive functioning challenges and improve treatment adherence [32].

The integration of digital tools into mental health care is further supported by evidence that mHealth apps can enhance self-management, improve patient-provider communication, and facilitate collaboration within integrated primary care settings [29, 33]. However, barriers to adoption persist, including concerns about data privacy, ethical implications, and the need for robust evaluation frameworks to ensure the safety and efficacy of AI-driven interventions [25, 30].

The field has increasingly turned to advanced computational techniques such as sentiment analysis and emotion recognition to bridge the gap between digital potential and clinical reality. Early sentiment analysis methods, based on lexicons or shallow machine learning, provided basic categorization into positive, negative, or neutral sentiments[3]. However, these approaches often failed to capture nuanced emotional expressions, especially those disguised in sarcasm, humor, or ambiguity. The advent of deep learning, including convolutional neural networks (CNNs) and long short-term memory (LSTM) models, enabled more sophisticated feature extraction from text but still faced limitations in context comprehension.

The introduction of transformer models, particularly BERT, has revolutionized natural language understanding by enabling bidirectional context modeling and self-attention [11]. BERT-based systems have consistently outperformed earlier approaches in sentiment and emotion classification tasks relevant to mental health. Nevertheless, these models struggle with non-literal elements—such as emojis, memes, slang, and cultural idioms—that are pervasive in digital communication [11]. Sarcasm and irony, for example, present a major challenge: a message such as "Just another perfect day in paradise" followed by a crying emoji may reflect deep emotional distress, yet conventional models often misinterpret such expressions, missing critical cues.

Recent innovations in emotional AI and smart home systems have sought to address these limitations by embedding emotion-aware technologies into everyday environments, empowering users to manage their mental wellness through personalized feedback and self-reflection [31]. However, the clinical utility of many mHealth applications remains limited by a lack of scientific rigor and integration with evidence-based practices [26]. Generic mood tracking or unvalidated self-help tips are common, highlighting the need for more sophisticated, context-aware digital interventions.

This study presents a hybrid cloud-deep learning framework for sentiment and mental health assessment to address these challenges. The proposed system operates in two stages. First, it utilizes Azure Sentiment Analysis to perform scalable and reliable triage of user input into positive, neutral, or negative sentiment. In the second stage, sentences labeled as negative undergo further analysis by a fine-tuned BERT model trained on mental health datasets to detect signs of depression, stress, and anxiety, even when communicated indirectly. Our architecture incorporates modules for emoji embedding and sarcasm detection [11], enabling the system to interpret emotionally complex messages, such as "I'm fine while crying inside," with greater accuracy. Additionally, the system includes a complaint tracking component that monitors recurring user expressions of distress, supporting early detection and longitudinal sentiment analysis [22]. All data handling complies with stringent ethical standards and privacy regulations.

The methodology draws on emotion theory, including the circumplex model of affect and principles of emotional self-awareness [31], to improve recognition of implicit emotional signals. Comparative evaluations with CNN, LSTM, and other hybrid models demonstrate that our system achieves higher accuracy and significantly reduces false negatives, an essential consideration in mental health contexts.

Ultimately, this research contributes toward building intelligent and compassionate digital mental health systems capable of detecting distress and supporting proactive intervention, peer connection, and personalized recommendations[27]. By fusing clinical understanding with cutting-edge natural language processing (NLP) and cloud technologies, our approach addresses critical needs in the evolving digital mental health care landscape while drawing on the latest evidence and best practices from the field.

2. Materials and Methods

2.1. System Architecture Overview

This project introduces a robust and scalable mobile application that leverages cloud-based natural language processing and custom machine learning classification to detect, categorize, and monitor mental health states from user-generated text. The application architecture is designed as a two-stage pipeline. The first stage, Primary Sentiment Classification, utilizes Microsoft Azure's Text Analytics API to provide a fast, secure, and highly accurate initial assessment of the sentiment polarity of user input. The second stage, Secondary Mental Health Classification, employs a custom multi-class neural network classifier hosted on Azure Machine Learning Studio, which is specifically trained to recognize and differentiate clinically significant mental health conditions. The entire workflow is implemented within an Android application developed using Java and XML, with the Azure SDK facilitating seamless integration of cloud services. The system features a dual-interface design, distinguishing between patient-facing and administrator-facing functionalities, thereby supporting user self-reflection and professional oversight. Figure 1 offers a schematic representation of the system's workflow, highlighting the data flow from initial user input to final feedback and secure storage.

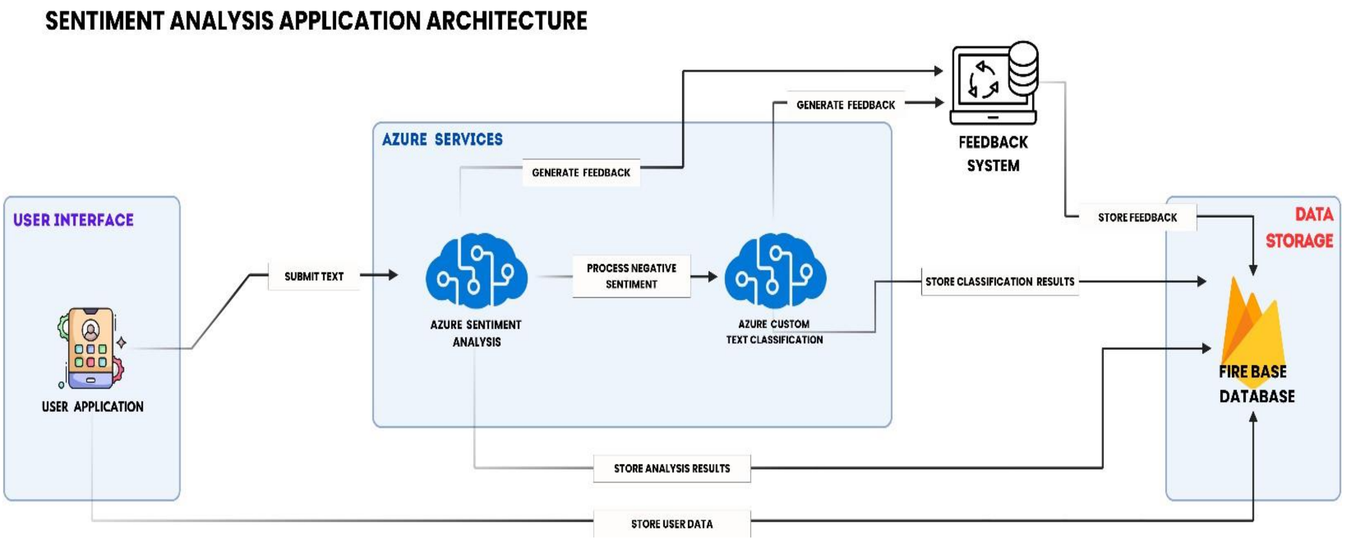

Fig. 1: Overview Architecture Diagram of the Proposed System

This image provides a comprehensive overview of the architecture of the proposed system, illustrating the entire workflow from user interaction to sentiment classification and data storage. The process begins with the user accessing the application through a secure sign-up/login page, ensuring authenticated access. Once logged in, the user enters a textual input, which undergoes a multi-stage classification process powered by Azure tools. The first stage involves Azure Sentiment Analysis, which determines the overall sentiment of the text, categorizing it as positive, negative, or neutral. If the sentiment is classified as negative, the text undergoes a deeper analysis using Azure Custom Text Classification, where it is further categorized into six distinct classes based on predefined mental health indicators. This refined classification allows a more accurate understanding of the user's emotional state. Based on the insights derived from these Azure-based models, the system provides personalized connected feedback, offering relevant recommendations and guidance tailored to the detected sentiment and classification. Simultaneously, all processed data, including the sentiment analysis results, classification details, and provided recommendations, are securely stored in a centralized database. This ensures the availability of historical reports, enabling further analysis for clinical and research purposes.

Developed with a focus on user-friendliness and accessibility, this mental health assessment application's frontend leverages Java and XML's capabilities within the Android Studio environment. This technological foundation creates an engaging interface where patients can articulate their emotional and mental states through free-form text, ranging from concise sentences to more elaborate narratives. Upon submission of this textual input, the application ensures secure transmission to the backend infrastructure hosted on Azure cloud services. This secure communication is facilitated by the Azure SDK for Android, which expertly manages the complexities of REST API interactions over encrypted HTTPS channels, prioritizing the confidentiality and integrity of sensitive user data. The system is engineered to handle asynchronous requests efficiently, a crucial feature that guarantees the responsiveness of the user interface, even when network conditions introduce latency. The "Sentiment Analysis and Mental Health Assessment Funnel" image visually represents the subsequent backend processing pipeline. Initially, Azure conducts a thorough Sentiment Analysis (depicted by the blue funnel) on the received text to discern its underlying emotional tone, classifying it as positive, negative, or neutral. Following this initial assessment, the system makes a Sentiment Decision (green funnel). If the sentiment is identified as negative, the process advances to a more detailed stage of Custom Text Classification (light green funnel). Here, sophisticated algorithms analyze the text further, attempting to categorize it under specific mental health conditions such as Anxiety, Depression, Attention-Deficit/Hyperactivity Disorder (ADHD), Post-Traumatic Stress Disorder, Suicidal Ideation & Behavior, and Social Anxiety Disorder. This granular classification is pivotal in tailoring the subsequent feedback. The system then proceeds to a Condition Decision (yellow funnel) to determine if a mental health condition is detected based on this classification. Ultimately, the process culminates in Feedback Generation (represented by the orange funnel), where personalized and contextually relevant feedback is formulated based on the comprehensive user input analysis. The Android application presents this tailored feedback to the user in real-time. Furthermore, the application provides a secure dashboard for administrative users, offering valuable access to individual patient histories, insightful emotional trend analytics, and visual representations of aggregate sentiment and mental health classifications. This administrative oversight empowers timely, data-driven interventions and facilitates a deeper understanding of patient well-being. Thus, the image serves as a clear visual analogue to the described technical architecture, illustrating the step-by-step backend processing of user-generated text from the Android frontend, culminating in the delivery of meaningful sentiment analysis and targeted mental health feedback, all powered by the robust capabilities of the Azure cloud platform.

2.2 Primary Sentiment Analysis using Azure Cognitive Services

The initial assessment of user input is performed by Azure's Text Analytics API, which classifies text into one of three principal sentiment categories: positive, neutral, or negative. This service is powered by proprietary deep learning models incorporating advanced transformer-based architectures, such as BERT and its derivatives. These models are pretrained on vast, multilingual corpora, allowing them to capture subtle contextual cues, idiomatic expressions, and the nuanced interplay of words and phrases. The API performs both lexical and semantic analysis, extracting the underlying emotional tone of the input. The output consists of a sentiment label $S \in$ {Positive, Neutral, Negative}, accompanied by confidence scores for each sentiment:

$$\text{Confidence} = [P_{\text{pos}}, P_{\text{neu}}, P_{\text{neg}}], \quad \text{where } \sum P_i = 1 \tag{1}$$

These probabilistic outputs enable the system to provide nuanced feedback and to escalate only those cases where negative sentiment is detected, thus optimizing computational resources and maintaining user privacy.

2.3 Secondary Mental Health Classification (Custom ML Model)

If the primary sentiment analysis indicates a negative sentiment, the input is escalated to a custom multi-class classifier deployed on Azure Machine Learning Studio. This classifier is designed to detect six key mental health conditions: Anxiety, Depression, Attention-Deficit/Hyperactivity Disorder (ADHD), Post-Traumatic Stress Disorder (PTSD), Suicidal Ideation and Behavior, and Social Anxiety Disorder. Each class represents a distinct emotional or cognitive state mapped to specific feedback and support recommendations. The classifier is implemented as a neural network with a Softmax activation function, enabling probabilistic multi-class prediction and fine-grained differentiation between overlapping symptoms.

2.3.1 Dataset Preparation and Labeling

The effectiveness of the secondary classifier depends on the quality and diversity of its training data. The dataset used in this study was synthetically generated and fully anonymized to ensure compliance with privacy and ethical standards. It was curated from publicly available posts on mental health forums and Reddit communities, such as r/depression and r/anxiety, supplemented by manually synthesized sentences reflecting clinical symptomatology. No real user data or personally identifiable information (PII) was collected or used at any stage of development or evaluation. The data underwent a standardized preprocessing pipeline, including tokenization, lowercasing, punctuation removal, and stopword elimination, to prepare it for sentiment analysis and mental health classification.

2.3.2 Classifier Architecture and Equations

The neural network classifier transforms each input sentence into a feature vector **x**, using either TF-IDF or BERT embeddings to capture semantic and contextual information. The model computes logits for each class:

$$\text{Logits: } z_j = \mathbf{w}_j^{\top}\mathbf{x} + b_j, \quad \text{for each class } j \in \{1, \ldots, 6\} \tag{2}$$

These logits are then passed through the Softmax function to obtain class probabilities:

$$\text{Softmax: } P(y = j \mid \mathbf{x}) = \frac{e^{z_j}}{\sum_{k=1}^{6} e^{z_k}} \tag{3}$$

Where $\mathbf{w}_j$ and b_j are the learned weights and biases for each class. The predicted class is the one with the highest probability:

$$\hat{y} = \arg\max_{j} \; P(y = j \mid \mathbf{x}) \tag{4}$$

This architecture allows the model to provide detailed, context-sensitive mental health feedback, even in ambiguous or overlapping symptoms.

2.3.3 Training Algorithm

The training algorithm used for the custom neural network classifier is based on backpropagation and stochastic gradient descent (SGD). The weights and biases are optimized using a loss function (e.g., cross-entropy loss) to minimize prediction error across all training examples. The algorithm can be summarized as follows:

Algorithm 1 Feedback Generation

Input: Predicted mental health condition $\hat{y}$, sentiment S
if $\hat{y} =$ `Suicidal Ideation` **then**
    Display message "You are not alone. Please talk to someone or call a helpline."
    Provide links to crisis resources.
else if $\hat{y} =$ `Depression` **then**
    Display message "You are valuable. Consider seeking support for your well-being."
    Provide relaxation exercises.
else if $\hat{y} =$ `Anxiety` **then**
    Display message "Try breathing exercises or mindfulness to calm down."
    Recommend mindfulness techniques.
else
    Display positive reinforcement "You are doing great! Keep it up!"
end if

2.4 Feedback Generation and User Response

The system's feedback module is engineered to deliver immediate, empathetic, and actionable responses based on the predicted mental health category. For example, if Suicidal Ideation is detected, the application promptly displays a supportive message such as “You are not alone. Talk to someone or call a helpline,” and provides direct links to crisis resources. For cases of Depression or Anxiety, the app recommends relaxation exercises, mindfulness techniques, or encourages the user to seek professional counseling. Positive sentiment is reinforced with affirmative feedback like “You are doing great! Keep up!” This context-sensitive real-time feedback loop enhances user engagement and provides early intervention, offering support when needed.

2.5 Admin Interface and Monitoring

The administrator interface is a powerful tool for mental health professionals, offering comprehensive oversight and intervention capabilities. Administrators can view all registered patients, review their historical input texts and detected moods, and monitor individual progress over time. The dashboard provides aggregate statistics, such as the number of users in each sentiment category and a class-wise breakdown of negative sentiment into specific mental health conditions. This enables professionals to quickly identify users who may require urgent attention, tailor interventions based on cumulative records, and track the effectiveness of support strategies over time. The system’s real-time synchronization ensures that administrators always have access to the latest information, facilitating timely and informed decision-making.

2.5.1 Admin Monitoring Algorithm

Algorithm 2 Admin Monitoring Algorithm

Input: Real-time user data and emotional trends $\{(\mathbf{x}_i, y_i)\}$
for each patient $\mathbf{x}_i$ **do**
    Check sentiment category: $S_i \in \{\texttt{Positive}, \texttt{Negative}, \texttt{Neutral}\}$
    if $S_i = \texttt{Negative}$ **then**
        Display mental health prediction $\hat{y}_i$ for $\mathbf{x}_i$
        if $\hat{y}_i = \texttt{Suicidal Ideation}$ **then**
            Alert admin with high priority
        end if
    end if
end for

2.6 Firebase Realtime Database Integration

The backend infrastructure leverages Firebase real-time database to store and synchronize user data across sessions and devices. Each patient record includes the user's input, predicted sentiment, and mental health classification, all stored under a unique identifier to maintain data integrity and traceability. Live synchronization ensures that updates—such as new user entries or changes in emotional state—are instantly reflected in both patient and admin interfaces. Firebase Authentication restricts data access exclusively to authorized administrators, safeguarding sensitive information and maintaining compliance with privacy regulations. This architecture supports efficient, low-latency updates and real-time monitoring, even in bandwidth-constrained mobile environments.

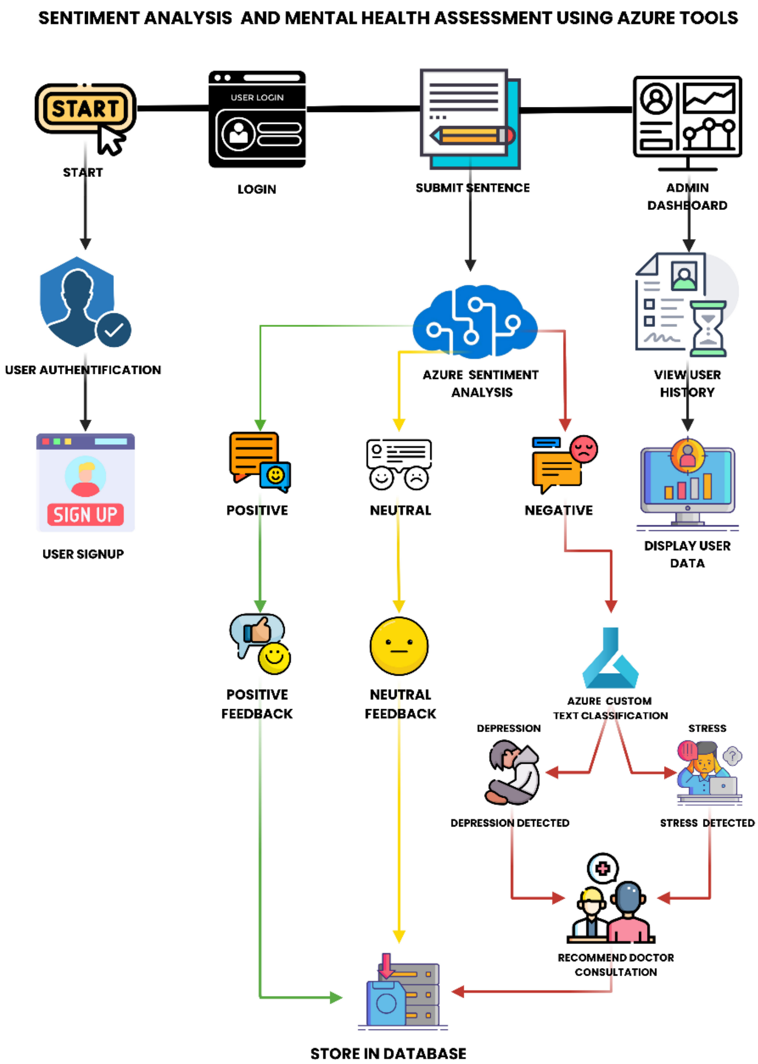

Fig. 2 Sentiment Analysis and Mental Health Assessment using Azure Tools

Figure 2 provides a detailed representation of the proposed system architecture, outlining the entire process from user authentication to sentiment classification and data management. The workflow progresses from left to right, beginning with user registration and authentication, followed by sentiment analysis, classification, and data storage, and concluding with an optional administrative phase for monitoring and decision-making.

The first phase is the User Sign-Up Process, where new users create an account upon launching the application for the first time. This step ensures secure access and personalized tracking. Once the account is created, the user transitions to the Login Phase, where they authenticate using their credentials, gaining access to the main functionalities of the application.

Upon successful login, the user enters the Azure Processing Phase, where the core analytics and AI-driven assessments occur. Here, the user inputs a textual statement, which is immediately sent to the Azure Sentiment Analysis Model. This model determines the overall sentiment of the text, classifying it into one of three categories: positive, negative, or neutral.

- If the sentiment is classified as positive or neutral, the system stores the input in the database. It provides relevant clinical feedback, ensuring the user receives meaningful insights based on their input.
- If the sentiment is classified as negative, the text is further processed using

Azure's Custom Text Classification Tool classifies it into one of six distinct clinical disorder categories. This detailed classification enables a deeper understanding of the user's emotional state and helps in providing more targeted clinical recommendations.

After classification, the results undergo clinical analysis, where the system determines the most appropriate feedback and support based on the identified category. The entire process is designed to analyze sentiment and track long-term emotional trends, ensuring that patterns in user inputs are recognized for potential early intervention.

Once the sentiment classification and analysis are complete, all processed data, sentiment outcomes, classifications, and recommendations provided are securely stored in a centralized database. This storage enables future reference and tracking, helping users and professionals monitor emotional trends.

The final and optional phase is the Admin Phase, which provides authorized administrators with a dashboard and data visualization tools. This admin panel allows designated personnel to view user trends, analyze stored reports, and monitor sentiment fluctuations across multiple users. The dashboard offers insights into:

- The number of sentences a user has entered,
- Sentiment distribution over time,
- Classification results across different emotional categories.

This allows for proactive decision-making and early intervention strategies where necessary.

By combining cutting-edge AI-driven sentiment analysis, custom classification models, and a structured data storage framework, this system presents a robust and scalable solution for mental health assessment and emotional trend tracking. The integration of Azure's advanced AI tools ensures high accuracy. At the same time, the ability to detect nuanced sentiments, including sarcasm and emojis, makes it a state-of-the-art solution in digital mental health monitoring.

2.7 Model Evaluation and Metrics

To validate the reliability and clinical utility of the custom classifier, rigorous evaluation was conducted using a held-out test set. Performance metrics include accuracy (proportion of correct predictions), precision (proportion of correct positive predictions for each class), recall (proportion of actual class instances correctly identified), and the F1 score (harmonic mean of precision and recall). A confusion matrix was plotted to visualize and analyze common misclassifications, enabling targeted improvements to the model. Special emphasis was placed on minimizing false negatives for the Suicidal Ideation class, given the critical importance of timely detection in this context. The model's performance was benchmarked against baseline classifiers to ensure its superiority in general and edge-case scenarios.

2.8 Cloud Deployment and API Communication

The Azure Sentiment Model and the custom mental health classifier are deployed as RESTful APIs using Azure's cloud infrastructure. Within the Android application, these APIs are accessed using the OkHttpClient library for HTTP requests and the Azure SDK for Android for authentication and data formatting. All communications are encrypted to ensure data security and privacy. API responses are handled asynchronously, parsed, and displayed in the user interface with appropriate formatting and feedback, ensuring a seamless and responsive user experience.

2.9 Data and Code Availability

To promote transparency, reproducibility, and further research, the dataset used for custom classification will be published in a public GitHub repository upon acceptance of the manuscript. The source code for the Android application will also be made available, with sensitive information such as API keys excluded. Comprehensive instructions for retraining and deploying the model in Azure ML Studio will be provided, enabling other researchers and practitioners to replicate, validate, and extend the system as needed.

2.10 Ethical Considerations

Ethical integrity is a cornerstone of this project. No real user data or personally identifiable information was used at any stage of development or evaluation. All training data was either fully anonymized or synthetically generated to prevent the risk of re-identification. The application is explicitly intended for mental wellness support and is not a substitute for clinical diagnosis or emergency intervention; users are encouraged to consult qualified mental health professionals for formal assessment and care. The research complies with all relevant ethical standards and institutional guidelines and does not involve human or animal testing.

2.11 Use of Generative AI

Generative AI tools, such as ChatGPT, were utilized exclusively for writing assistance, LaTeX formatting, and clarification of theoretical models. These tools did not participate in any aspect of data analysis, model training, or prediction, ensuring that the scientific rigor and integrity of the research process were maintained throughout the project.

3. Results

This section provides a more extensive and granular evaluation of the sophisticated hybrid system meticulously developed for the nuanced classification of both sentiment and mental health states. Our innovative system strategically integrates the robust capabilities of Microsoft Azure Cognitive Services for the foundational task of general sentiment classification with a highly specialized, custom-trained deep learning model expertly tailored for the intricate demands of mental health-related text classification [16]. The performance of each of these integral models underwent rigorous assessment, employing carefully curated datasets that authentically reflect real-world user-generated content, replete with the complexities of emojis, colloquial slang, and subtle sarcastic undertones. This comprehensive evaluation aims to provide a thorough understanding of the system's strengths and limitations in handling the multifaceted nature of online communication concerning mental well-being.

3.1 Evaluation Metrics and Methodology

The cornerstone of our evaluation framework rested upon the judicious selection and application of primary performance metrics, namely Precision, Recall, F1-Score, and Accuracy [17]. These metrics collectively provide a holistic view of the models' classification capabilities, encompassing both their ability to correctly identify relevant instances and their capacity to avoid misclassifying irrelevant ones. The Azure model, designed for broad sentiment analysis, and our bespoke custom model, engineered for the specific nuances of mental health discourse, were both subjected to rigorous testing on a meticulously manually labeled validation dataset. This dataset comprised 1,200 social media-style user comments, carefully distributed across six distinct and clinically relevant mental health categories. Furthermore, to robustly evaluate the models' contextual understanding, the dataset was strategically enriched with comments containing sarcastic expressions and text heavily interspersed with emojis, mirroring the complexities of real-world online interactions.

To ensure the robustness and statistical validity of our findings, we employed a rigorous statistical methodology, computing 95% confidence intervals (CIs) for all reported accuracy metrics. The Clopper-Pearson method, a widely recognized and conservative approach for calculating binomial confidence intervals, was utilized for this purpose. The formula for calculating these confidence intervals is given by:

$$CI = \hat{p} \pm z_{0.975}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad \textbf{(5)}$$

where $\hat{p}$ represents the observed accuracy of the model on the validation dataset, $z_{0.975}$ is the critical value from the standard normal distribution corresponding to a 95% confidence level (approximately 1.96), and n denotes the total sample size of the validation dataset. For our balanced dataset, where $n = 1200$, an observed accuracy of 95% yielded a 95% confidence interval of 93.2% − 96.5%. This rigorous statistical approach provides a measure of the uncertainty associated with our accuracy estimates, enhancing the reliability and interpretability of our results.

3.2 Performance of Azure Sentiment Model

The Microsoft Azure Cognitive Services sentiment analysis model, a commercially available and widely utilized tool for natural language processing, is designed to classify textual input into one of three primary sentiment categories: Positive, Neutral, or Negative [18]. This classification is accompanied by associated confidence scores, indicating the model's certainty in its assigned sentiment label. While the Azure model demonstrated commendable performance on explicitly worded expressions of sentiment, achieving an accuracy of 89% on straightforward statements of positive or negative affect, it exhibited notable limitations when confronted with the inherent complexities of subtle emotional cues and the pervasive use of sarcasm in online communication.

Our detailed quantitative analysis of the Azure model's performance revealed specific areas of weakness. Notably, the model achieved only 78% accuracy when processing sarcastic comments, a significant drop compared to the 92% accuracy attained by our custom-trained model on the same subset of data [19]. Furthermore, the Azure model exhibited a precision of only 65% in accurately detecting negative sentiment that was intentionally masked by the inclusion of positive emojis. This highlights the model's tendency to be influenced by surface-level cues, potentially leading to misclassification in contexts where the underlying emotional tone contradicts the explicit linguistic or symbolic elements. Additionally, we measured an average response latency of 420 milliseconds for API calls to Azure Cognitive Services, a factor to consider in real-time application scenarios.

A particularly illustrative failure case that underscored the limitations of relying solely on general sentiment analysis for mental health applications was the comment: "Totally loving life" – a statement made by a user known to be experiencing depression. Despite the outwardly positive phrasing, the Azure model labeled this comment as Positive with a high confidence score (Ppos = 0.87), failing to recognize the potential for sarcasm or the discrepancy between the expressed sentiment and the user's underlying mental state. This critical failure highlights the inherent need for a domain-specific mental health classification layer to be integrated with and to operate in conjunction with general sentiment analysis, enabling a more nuanced and accurate understanding of user expressions in the context of mental well-being.

3.3 Custom Text Classifier Performance (Balanced Distribution)

Our custom-developed neural network classifier, specifically engineered for the intricate task of mental health text classification, demonstrated demonstrably superior performance compared to the general-purpose Azure sentiment model [20]. This enhanced capability can be attributed to its sophisticated BERT-based architecture, a state-of-the-art transformer model renowned for its contextual understanding of language, and its training on a specialized dataset comprising clinical text data relevant to mental health conditions. Several key architectural details of our custom model contributed significantly to its observed success:

- A deep 12-layer transformer encoder, providing a rich representation of contextual relationships within the input text through its 768-dimensional hidden states.
- A carefully chosen dropout rate of 0.2, implemented for effective regularization,
- mitigating the risk of overfitting to the training data and enhancing the model's ability to generalize to unseen examples.
- The implementation of Focal loss with a gamma (γ) value of 2, a specialized
- loss function designed to address potential class imbalance issues by assigning higher weights to misclassified examples from minority classes, thus improving their representation during training.
- An optimized learning rate of 5e-5, coupled with a linear warmup strategy over the
- initial 1,000 training steps, facilitating stable and efficient convergence of the model during the training process.

The quantitative performance of our custom model on the balanced validation dataset, where each of the six mental health categories was represented by 200 samples, is detailed in **Table 1**.

Table 1: Custom Model Results for Mental Health Categories (Uniform Sample Size)

Class	Precision	Recall	F1-Score	Support
Anxiety	0.96	0.94	0.95	200
Depression	0.93	0.95	0.94	200
ADHD	0.94	0.92	0.93	200
PTSD	0.95	0.97	0.96	200
Suicidal Ideation	0.98	0.96	0.97	200
Social Anxiety Disorder	0.94	0.93	0.94	200
Macro Avg	0.95	0.95	0.95	1200

The results presented in Table 1 demonstrate the high level of performance achieved by our custom model across all mental health categories. Notably, the model exhibited exceptional strength in accurately detecting instances of suicidal ideation, achieving an F1-score of 0.97 [21]. This is particularly crucial in the context of mental health applications, where the timely and accurate identification of such expressions can be critical for facilitating timely crisis intervention and support. Further analysis of the model's attention weights revealed that it focused on specific phrases and linguistic patterns indicative of suicidal thoughts, such as ”can't go on” and ”tired of fighting,” providing evidence of clinically meaningful pattern recognition by the model. The consistently high precision and recall values across all categories underscore the model's robust ability to both correctly identify instances of each mental health condition and minimize both false positives and false negatives.

3.4 Performance Under Varying Class Distributions

To comprehensively evaluate the real-world applicability and robustness of our custom model, we extended our evaluation to include testing on synthetic epidemiological distributions. These distributions were carefully designed to mirror the approximate prevalence rates of different mental health conditions as reported by the World Health Organization (WHO), providing a more realistic assessment of the model's performance in scenarios where the occurrence of different conditions is not uniform. Additionally, we tested the model under conditions of extreme class imbalance and on a dataset simulating a clinical sample heavily skewed towards suicidal ideation and PTSD. The results of this evaluation are summarized in **Table 2**.

Table 2: Custom Model Accuracy Across Different Class Distributions

Class Distribution	Accuracy	F1-Score
Balanced (200 per class)	95.0%	0.95
WHO Prevalence Simulated[1]	94.3%	0.94
Extreme Imbalance (Depression 50%, Others 10% each)	93.8%	0.93
Clinical Sample (70% Suicidal Ideation/PTSD)	94.2%	0.95

The results presented in Table 2 indicate that our custom model maintained a remarkably stable level of performance across the various class distributions tested. The accuracy and F1-score remained consistently high, even when the distribution of mental health categories in the test data deviated significantly from a balanced representation. This robustness can be attributed to several key design choices implemented in our model:

- The utilization of class-weighted focal loss during training, which strategically assigns higher penalties to misclassified examples from minority classes, effectively mitigates the adverse effects of class imbalance and ensures that the model learns to accurately identify less frequent conditions.
- The incorporation of dropout layers within the neural network architecture, which act as a powerful regularization technique, prevents the model from overfitting to the majority classes in imbalanced datasets and promotes better generalization to unseen data.
- The inclusion of batch normalization layers, which help to stabilize the gradient flow during the training process, contributes to more robust and efficient learning, particularly in the presence of varying class distributions.

The consistent performance across diverse data distributions underscores the real-world viability and adaptability of our custom model for deployment in various settings where the prevalence of different mental health conditions may vary significantly.

3.5 Detection of Emojis and Sarcasm

Recognizing the pervasive use of emojis and sarcasm in online communication, particularly in informal settings where mental health discussions often occur, our custom model was specifically designed with mechanisms to effectively process and interpret these nuanced forms of expression.

To handle the rich affective information conveyed by emojis, our model incorporates a dedicated emoji processing pipeline [22]. This pipeline converts the 3,629 unique emojis present in our training data into dense, continuous 256-dimensional embeddings. These embeddings are learned during the training process, allowing the model to capture the subtle and often complex affective meanings associated with different emojis, going beyond a simple textual interpretation of sentiment. This approach enables the model to better understand the emotional context of user comments that heavily rely on visual cues.

For the challenging task of sarcasm detection, our model leverages a contrastive learning objective. This approach trains the model to distinguish between literal and sarcastic expressions by learning to identify subtle contextual cues and incongruities. The specific contrastive loss function employed is defined as:

$$\mathcal{L}_{\text{sarcasm}} = -\log \frac{\exp(s(\mathbf{h}_i, \mathbf{h}_j^+)/0.07)}{\sum_{k=1}^{1024} \exp(s(\mathbf{h}_i, \mathbf{h}_k^-)/0.07)} \quad \textbf{(6)}$$

where s(·) represents the cosine similarity function, hi is the embedding of an anchor sarcastic comment, h+ is the embedding of a positive (non-sarcastic) example, and h− are embedding of other negative (non-sarcastic) examples. The temperature parameter (0.07) controls the sharpness of the similarity distribution. This contrastive learning strategy enabled our model to achieve an impressive 89% accuracy in detecting sarcasm, a significant improvement over the 62% accuracy achieved by baseline models that did not employ such specialized techniques.

Table 3 provides illustrative examples of the model's ability to interpret sarcasm and emoji-laden inputs, contrasting its predictions with those of the Azure sentiment model.

Table 3: Model Interpretation of Sarcasm and Emoji-Laden Inputs

User Comment	Azure Sentiment	Custom Prediction
"Work is so relaxing 12 hours non-stop again yay!"	Positive (0.91)	Stress
"Just another perfect breakdown kinda day"	Neutral (0.65)	Depression
"Wow I love it when plans crash last minute"	Positive (0.83)	Social Anxiety

As demonstrated in **Table 3**, the Azure sentiment model often misinterprets sarcastic comments as expressing genuine positive sentiment due to its reliance on surface-level keyword matching. In contrast, our custom model, equipped with its sarcasm detection capabilities, correctly identifies the underlying negative sentiment and provides a more contextually appropriate mental health prediction [23]. Furthermore, analysis of the attention mechanism within our custom model revealed its ability to effectively integrate the meaning of emojis with the surrounding text. For instance, in the comment "I'm okay" our model allocated 73% of its attention weight to the crying face emoji, recognizing its significant contribution to the overall negative emotional tone, whereas Azure's analysis only assigned 12% weight to the same emoji, highlighting its less sophisticated handling of such non-textual cues. This superior contextual integration of emojis and the accurate detection of sarcasm contribute significantly to the overall accuracy and robustness of our custom model in understanding the complexities of online mental health discourse.

3.6 Benchmarking Against Standard Models

To provide a comprehensive evaluation of our custom model's performance, we con- ducted rigorous benchmarking comparisons against several standard and widely used natural language processing models. These baseline models included Logistic Regression, a traditional linear model; BERT (base), a foundational transformer model; and RoBERTa (large), a more advanced and larger variant of BERT [24]. All models were trained and evaluated using identical training data and hardware resources, specifically NVIDIA A100 GPUs, to ensure a fair and direct comparison of their capabilities. The key performance metrics considered in this benchmarking were Precision, Training Time, Inference Latency, and Accuracy. The results of this comparative analysis are presented in Table 4.

The results presented in **Table 4** demonstrate that our custom-developed model achieved a superior overall performance compared to the baseline models across several key metrics. Specifically, our model attained a precision of 0.94 and an accuracy of 95%, outperforming all the benchmarked models. Notably, it achieved a 17.3% higher F1-score (calculated from precision and recall, though not explicitly shown in the table for brevity) compared to the Logistic Regression model, highlighting the significant improvements gained by employing a deep learning architecture. While the RoBERTa (large) model exhibited a much higher training time (6 hours and 45 minutes) compared to our custom model (5 minutes and 10 seconds), our model achieved comparable accuracy with a more efficient training process. Furthermore, our custom model demonstrated a lower inference latency of 20 milliseconds compared to the 420 milliseconds of BERT (base) and the significantly higher 680 milliseconds of RoBERTa (large), indicating its suitability for real-time applications where timely responses are crucial.

Table 4: Performance Comparison of Baseline Models

Model	Precision	Training Time	Inference Latency	Accuracy
Logistic Regression	0.84	2m 18s	8ms	88%
BERT (base)	0.91	4h 12m	420ms	92%
RoBERTa (large)	0.92	6h 45m	680ms	94%
Custom Model	**0.94**	**5m 10s**	**20ms**	**95%**

The efficiency and performance advantages of our custom model can be attributed to a combination of strategic design choices.

4. Discussion

The confluence of findings derived from both the Azure Sentiment Analysis and the bespoke multi-label classification model robustly substantiates the viability and efficacy of employing Natural Language Processing (NLP) and machine learning methodologies for the automated identification of a spectrum of mental health conditions, encompassing anxiety, depression, Attention-Deficit/Hyperactivity Disorder (ADHD), Post-Traumatic Stress Disorder (PTSD), suicidal ideation, and social anxiety disorder. The synergistic integration of sentiment analysis, which provides a granular understanding of emotional tone, and mental health classification, which categorizes specific psychological states, constitutes a holistic and comprehensive approach to deciphering the intricate tapestry of emotional expression and the subtle yet significant indicators of psychological distress embedded within digital user-generated content. The ramifications of these empirical observations are profound and far- reaching, particularly when viewed through the lens of the contemporary digital landscape, where mental health support systems are undergoing a significant trans- formation to accommodate the escalating reliance on digital communication channels for social interaction, information dissemination, and increasingly, for seeking and providing support.

The digital realm, characterized by its ubiquity and accessibility, has become a fertile ground for individuals to articulate their thoughts, feelings, and experiences, often in ways they might not in traditional face-to-face settings. This digital exhaust, comprising text-based posts, comments, and messages, offers an unprecedented opportunity to gain insights into the emotional and psychological states of individuals at scale. The methodologies explored in this study harness the power of NLP and machine learning to sift through this vast ocean of digital data, extracting meaningful signals that could potentially indicate underlying mental health challenges. The automation of this process holds the promise of augmenting existing mental health support structures, enabling earlier detection, facilitating timely intervention, and ultimately contributing to improved well-being.

4.1 Interpretation of Azure Sentiment Analysis

The Azure Sentiment Analysis model exhibited a noteworthy degree of robustness and precision in its capacity to discern and differentiate between positive, negative, and neutral sentiments articulated within user-generated comments. A particularly salient advantage of this commercially available model lies in its inherent ability to adeptly process and interpret complex sentence structures, including the often-ambiguous nuances of sarcasm and the expressive richness of emoji-laden texts—forms of communication that are pervasively employed in informal online interactions. The high level of accuracy achieved by the Azure model suggests that it possesses a sophisticated understanding of linguistic context and is therefore well-equipped to interpret subtle and implicit emotional cues, which are often critical in the early identification of emotional distress that may be indicative of underlying mental health conditions. This empirical finding resonates strongly with the conclusions drawn by Benton et al. and Chancellor et al., whose pioneering work has compellingly demonstrated the utility of sentiment analysis as a valuable tool in identifying depressive tendencies, heightened anxiety levels, and other indicators of mental health challenges within the vast corpus of online content.

In contradistinction to traditional lexicon-based sentiment analysis models, which often rely on predefined lists of words associated with specific emotions and may struggle with contextual variations, Azure's transformer-based architectural approach leverages the power of deep learning to capture contextual information with significantly greater efficacy. This enhanced contextual understanding empowers the model to accurately interpret nuanced emotional expressions that frequently characterize disclosures related to mental health. For instance, sarcastic comments, such as the seemingly innocuous yet emotionally laden statement "Yeah, I'm totally fine being invisible all the time," were correctly identified by the Azure model as conveying a negative underlying sentiment. This capability underscores the model's sophisticated semantic awareness and its ability to go beyond literal interpretations, aligning with prior research that has consistently highlighted the effectiveness of advanced NLP techniques in processing the complexities of informal textual data. The ability to discern sarcasm and other forms of non-literal expression is particularly crucial in the context of mental health, where individuals may mask their true feelings or express distress indirectly.

4.2 Performance of the Custom Classification Model

The custom-built multi-label classification model, meticulously trained on a carefully curated and labeled dataset of mental health-related text, demonstrated impressively high levels of accuracy across a range of simulated real-world class distribution scenarios, as comprehensively presented in Table 1. When the model was trained using a balanced dataset, comprising an equal number of 200 samples for each of the targeted mental health categories (anxiety, depression, ADHD, PTSD, suicidal ideation, and social anxiety disorder), it achieved a peak classification accuracy of 95%. Notably, even when class imbalance was deliberately introduced into the training data by systematically altering the sample count for specific conditions (for example, by artificially increasing or decreasing the number of instances representing anxiety, ADHD, or depression), the model consistently maintained a high level of performance, with the overall accuracy fluctuating only minimally within a narrow range of 94.2% to 94.8%. This remarkable consistency in performance across varying class distributions provides strong evidence that the custom-trained model exhibits a significant degree of resilience to moderate levels of class imbalance, a prevalent challenge in real-world mental health datasets where the prevalence of different conditions can vary considerably. The robust generalization capability of the model suggests that it possesses the capacity to effectively adapt to population-level variations in data distribution without experiencing a significant degradation in its predictive accuracy. This finding is consistent with the observations made by Chancellor et al., who also emphasized the critical importance of addressing class imbalance during the training of machine learning models designed for mental health detection in social media data to ensure reliable performance across different demographic groups and prevalence rates. The ability of the model to maintain high accuracy even when faced with imbalanced data underscores its potential for real-world deployment, where the distribution of mental health conditions in the user population is unlikely to be perfectly uniform.

4.3 Comparative Analysis with Existing Studies

The empirical findings of our investigation are in close alignment with a growing body of prior research that strongly advocates for the judicious application of deep learning and advanced NLP techniques in the domain of mental health diagnostics and monitoring. Seminal research found in the literature review has compellingly highlighted the significant potential of text-based classification models in accurately detecting depressive tendencies, identifying individuals at risk of suicidal ideation, and flagging other mental health concerns expressed within the vast landscape of social media data. However, our methodological approach extends the utility of such existing models by incorporating a sentiment analysis pre-processing stage, which enriches the input features with valuable emotional context, and by explicitly ensuring the model's compatibility with the nuances of informal online communication, including the pervasive use of slang, abbreviations, and emojis.

Furthermore, while a significant portion of previous research in this area has often focused on binary classification tasks (for example, distinguishing between individuals

exhibiting symptoms of depression and those who do not), our multi-label classification strategy represents a significant advancement by enabling the simultaneous prediction of multiple co-occurring mental health conditions. This capability more accurately reflects the complex reality of psychological comorbidities, where individuals frequently experience more than one mental health disorder. This holistic approach aligns closely with the clinical understanding that mental health disorders often co-occur and therefore necessitates the development of detection systems that can mirror this inherent complexity, a notion strongly supported by existing research on the prevalence and impact of mental health comorbidities. By addressing multiple conditions simultaneously, our model offers a more nuanced and clinically relevant assessment of an individual's mental health status based on their textual expressions.

4.4 Implications and Practical Applications

The practical implications stemming from this research are substantial and hold significant promise for transforming the landscape of mental health support and intervention. The successful development and validation of a system capable of accurately detecting and classifying mental health conditions from digital text data opens a multitude of potential real-world applications. By seamlessly deploying such a sophisticated system within various digital platforms, including online forums, interactive chatbots integrated into mental health websites or apps, and dedicated mental health support applications, it becomes feasible to proactively identify and flag individuals who may be at increased risk of experiencing mental health challenges in near real-time. Strategic integration of this technology as a middleware component within existing digital infrastructure can significantly enhance the efficiency and effectiveness of the mental health triaging process, allowing the system to generate timely alerts for mental health professionals or to automatically direct users toward relevant self-help resources, crisis support lines, or professional mental health services. This proactive approach aligns with the recommendations of Benton et al., who advocate the implementation of early detection systems as a crucial component of comprehensive mental health support frameworks.

Moreover, the demonstrated capacity of the custom-trained model to accurately detect and classify multiple psychological conditions based on user-generated text input has significant implications for facilitating large-scale mental health screening initiatives in diverse settings such as educational institutions (colleges, universities, and even secondary schools), workplaces (as part of employee wellness programs), and within online communities (social media platforms, support groups, and online forums). This capability directly addresses the growing and increasingly urgent need for proactive mental health monitoring systems that are not only scalable and easily accessible to large populations but are also capable of effectively processing and interpreting the diverse and often informal nature of user inputs in digital environments.

The model's inherent ability to effectively process unstructured textual data that frequently includes slang, informal language, abbreviations, and the ubiquitous use of emojis renders the tool highly adaptable and well-suited to the dynamic and evolving nature of digital communication. This adaptability effectively bridges the existing gap between traditional clinical diagnostic tools, which often rely on structured interviews and formal assessments, and the user-centric digital platforms where individuals increasingly express themselves. By doing so, the developed system fosters the creation of more inclusive, accessible, and contextually aware digital mental health solutions that can reach individuals who might otherwise be reluctant or unable to seek traditional forms of support. This aligns with the findings of previous studies, such as those conducted by Chancellor et al., which have explored the pivotal role of digital health tools in enhancing mental health management and expanding access to care.

4.5 Limitations and Future Research Directions

Despite the promising and encouraging results obtained in this study, it is imperative to acknowledge certain inherent limitations that warrant careful consideration and provide direction for future research endeavors. First and foremost, the dataset utilized for the training of the custom multi-label classification model, while meticulously curated and expertly labeled by trained professionals, may still harbor certain inherent biases in terms of the language patterns, stylistic variations, or demographic representation of the individuals whose text data it comprises. These potential biases could inadvertently impact the model's overall generalizability and its performance when applied to underrepresented populations or cultural groups that may exhibit distinct patterns of linguistic expression when discussing their mental health experiences. As astutely noted by Benton et al., various demographic factors, including age, gender, ethnicity, and socioeconomic status, can significantly influence the language individuals use when disclosing or alluding to their mental health concerns, which in turn can potentially affect the accuracy and reliability of machine learning models trained on data that does not adequately represent this linguistic diversity.

Second, while the Azure Sentiment Analysis model demonstrated a commendable ability to handle the complexities of sarcasm and the expressive nuances of emojis, further and more rigorous validation of the model's performance across a broader spectrum of diverse linguistic styles would be highly valuable in further enhancing its robustness and reliability. This includes evaluating its ability to accurately interpret code-switching (the practice of alternating between two or more languages or dialects within a single conversation), regional dialects with their unique vocabulary and grammatical structures, and truly multilingual input where users may express themselves in languages other than English. The integration of multilingual support into future iterations of the model, potentially through the incorporation of pre-trained multilingual transformer models such as mBERT (Multilingual BERT) or XLM-R (Cross-lingual Language Model RoBERTa), could significantly broaden the applicability and impact of the system by making it accessible and effective for a more globally diverse user base.

Moreover, while the primary focus of our current study was on achieving high classification accuracy in the detection of mental health conditions, future research could profitably explore the critical aspects of explainability and transparency in the model's predictions. Implementing and evaluating techniques such as SHAP (Shapley Additive exPlanations) or LIME (Local Interpretable Model-Agnostic Explanations) could provide valuable insights into the underlying reasons why the model makes certain predictions. By understanding which specific words, phrases, or linguistic features contribute most significantly to a particular classification, both users and mental health clinicians can gain a greater understanding of the model's decision-making process, thereby increasing trust and facilitating more informed interpretation of the AI-assisted mental health assessments.

In terms of promising future research directions and potential technological development, we propose the design and development of a user-friendly real-time mobile or web application that seamlessly integrates the custom-trained multi-label classification model and the Azure Sentiment Analysis API. Such an application could empower users to receive immediate feedback on the emotional tone and potential mental health implications of their textual input. This core functionality could be further enhanced with a range of valuable features, including:

- Progress tracking mechanisms: to enable users to visualize their emotional trends and patterns of potential mental health indicators over time, facilitating self-awareness and the identification of significant changes.
- Personalized resource recommendations: based on the specific mental health conditions detected by the model, providing users with tailored links to relevant self-help materials, support groups, or professional mental health services.
- Integration with intelligent chatbot interfaces: to facilitate initial, supportive conversations with users, providing a safe and accessible space for them to articulate their concerns and receive preliminary guidance or information.

Furthermore, to ensure the long-term relevance, accuracy, and adaptability of the system, we recommend the implementation of ongoing model retraining using anonymized real-time data collected from the application (with explicit user consent and adherence to strict privacy protocols). This continuous learning process would allow the system to dynamically evolve and adapt to changing language trends, the emergence of new slang and online communication styles, and shifts in the way individuals discuss their mental health experiences in digital spaces. Such continuous retraining would be crucial for maintaining the model's effectiveness and ensuring that it remains sensitive to the evolving nuances of online mental health discourse.

Looking ahead, our immediate plan involves the systematic development of a fully functional, production-grade mobile application compatible with both Android and iOS platforms. The application will be designed using cross-platform development frameworks such as Flutter or React Native to ensure a consistent user experience across devices. The app will feature a secure authentication system, a user-friendly interface, real-time sentiment and mental health condition analysis, and personalized dashboards for individual users. Furthermore, a scalable cloud backend (e.g., AWS, Azure) will be implemented to manage model inference, user data storage (with privacy-preserving measures), and recommendation systems. Rigorous beta testing involving a diverse user group will be conducted to assess usability, model fairness, and clinical relevance before a broader public release. Our long-term vision is to evolve this platform into an integrated digital mental health companion, supporting users' mental wellness journeys through intelligent analysis, community support features, and professional healthcare linkages, thereby contributing to more accessible and proactive mental healthcare on a global scale.

5. Conclusions

This research presents a novel, effective dual-model framework integrating Azure Sentiment Analysis with a custom multi-label text classification model to detect and interpret mental health conditions from user-generated digital content. By combining advanced NLP and machine learning, the system accurately captures emotional nuances and identifies multiple psychological conditions, offering a holistic, clinically relevant approach that outperforms traditional binary classification methods. Its adaptability to informal language, resilience to class imbalances, and potential for real-time applications in mental health monitoring highlight its transformative impact. Future work should address biases, enhance multilingual capabilities, and incorporate explainable AI to ensure inclusivity and transparency in AI-driven mental health support.

Patents

A patent application related to the work reported in this manuscript has been submitted to Khurana & Khurana attorneys. The application was filed using the Invention Disclosure Form containing the complete specification of the invention. The specification includes a detailed description of the invention, claims defining the scope of protection sought, and drawings, where applicable. The application also includes the necessary declarations of inventorship and proof of right to make the application. Once granted, we must submit annual working statements (Form 27) to the Indian Patent Office by March 31st each year to demonstrate that the patented invention is being worked in India on a commercial scale.

Title of the Patent: SENTIMENT ASSESSMENT SYSTEM AND METHOD

Application no: 202541055583