








Unveiling Emotions: NLP-Based Mood Classification and Well-Being Tracking for Enhanced Mental Health Awareness

Asha Rani Mishra¹, Amrita Rai², Durgesh Nandan^{3*}, Ujwala Kshirsagar⁴, Mahesh Kumar Singh⁵

¹ CSE Department, Galgotias College of Engineering and Technology, Greater Noida 201310, India

² ECE Department, LLOYD Institute of Engineering and Technology, Greater Noida 201310, India

³ School of CS&AI, S R University, Warangal 506371, India

⁴ E&TC Department, Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University), Pune 412115, India

⁵ ECE Department, Aditya University, Surrampalem 533437, India

Corresponding Author Email: durgeshnandano51@gmail.com

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/mmep.120228>

ABSTRACT

Received: 8 July 2024

Revised: 13 September 2024

Accepted: 20 September 2024

Available online: 28 February 2025

Keywords:

sentiment analysis, Bidirectional Encoder Representations from Transformer (BERT), decoder, machine learning, polynomial decay, mean F1 score

Even while mental health is vital to overall welfare, in today's busy environment, it is often disregarded. These days, people are searching for ways to maintain a balance in their mental well-being as they become more aware of the importance of mental health. Raising awareness of mental health issues lowers stigma, encourages early intervention, and enhances general wellbeing. It promotes comprehension and assistance, augmenting the standard of living. Natural Language Processing (NLP) and transformer models, like as GPT-4 and Bidirectional Encoder Representations from Transformer (BERT), can use sentiment analysis and ailment tracking to identify mental health problems early on. By providing scalable support solutions and language translation, these technologies enhance access to care. They also support clinical decision-making and research. This study proposes a methodology that uses ideas from Natural Language Processing to categorize people's emotions into six groups. The accuracy of the model, which was developed utilizing transformer models refined on an emotion's dataset, was over 94% in each of the six categories it generated. The model was trained on a sizable corpus of English data using the transformers architecture. It made use of a batch size of 256 and the optimizer "Adam," with a weight decay value of 0.01 and a learning rate value of $1e-4$. Several epochs were used in the training phase, and the accuracy of the model was improved iteratively on both the training and test datasets throughout each epoch. Significant advancement was made by the suggested model, which attained excellent accuracy on training and validation datasets. Strong generalization skills are demonstrated by the validation accuracy, which stayed constantly high. This study highlights the possibility of NLP-based models to support self-reflection and provide insights into people's moods, both of which are important for managing mental health.

1. INTRODUCTION

Depression is a significant contributor to disability and suicide is a fourth leading causes of death among individuals aged 15-29. Demographic changes have led to a 13% increase in mental health conditions, with 20% of children and adolescents worldwide experiencing a mental health condition. These conditions can greatly impact various areas of life, such as school and work performance, relationships, and community participation. Depression and anxiety are two of the most prevalent mental health conditions and they are costing the global economy \$1 trillion annually. Despite these significant costs, the global median of government health expenditure allocated to mental health is less than 2%. The World Health Organization estimates that the rate of age-adjusted suicide in India is 21.1 per 100,000 people, while the burden of mental health conditions in the country is 2443

disability-adjusted life years (DALYs) per 100,000 people. Personal qualities including the capacity to control one's thoughts, emotions, behaviors, and interactions with others are among the aspects that affect mental health. Additionally, particular psychological and behavioral qualities, genetic elements, and environmental, social, cultural, economic, and political considerations all play a part. According to data, 1 in every 8 people worldwide suffers from a mental condition. Types of mental disorders anxiety disorder: According to estimates, 301 million people worldwide, including 58 million children and teenagers, suffered from an anxiety illness in 2019. Excessive fear, concern, and accompanying behavioral abnormalities are hallmarks of anxiety disorders. Significant anxiety or functional impairment might result from these symptoms. Depression: According to a study, in 2019, depression affected 280 million individuals globally, including 23 million adolescents and teens. This illness is

distinct from ordinary mood swings and transient emotional reactions to daily challenges. An individual may feel a persistently sad mood, a lack of interest in activities, and other symptoms over a protracted period of at least two weeks, nearly every day, when they are experiencing a depressive episode. Bipolar Disorder: 40 million people worldwide suffered from bipolar illness in 2019. Bipolar individuals fluctuate between periods of manic symptoms and depressive bouts. During a depressive episode, the person experiences a low mood (feels hopeless, agitated, or empty) or loses interest in activities for most of the day, virtually every day. Post-Traumatic Stress Disorder (PTSD): The incidence of PTSD and other psychological conditions is prevalent in areas affected by conflict. PTSD can occur as a result of exposure to traumatic or highly distressing situations or events. Schizophrenia: Approximately 24 million individuals worldwide, or 1 in 300, are affected by schizophrenia. Schizophrenia patients may experience problems with cognitive functioning for an extended period of time. Individual who may be susceptible to mental illness: At any given time, a wide range of structural, familial, community, and individual factors may combine to either promote or undermine mental health. Even though most people are resilient, there is a greater chance of harm for those who are subjected to adverse circumstances like poverty, violence, disability, and inequality. Loneliness: Because we are social beings by nature, we need surroundings where we may interact in safety and security. For general mental and physical well-being, it is essential to maintain healthy social ties. Loneliness can result from a breakdown in social ties. Through history, loneliness has been acknowledged as a common human condition. It can also play a role in the onset of a number of psychiatric conditions, including Alzheimer's disease, alcoholism, depression, child abuse, sleep disorders, and personality disorders. A terrible and universal emotion with evolutionary roots is loneliness. It acts as a reminder of the risks linked with solitude and its possibilities. Loneliness is defined as the lack of essential social connections and the absence of emotional intimacy in current relationships. It is a common experience, with a high percentage of the population reporting feelings of loneliness at some point in their lives, particularly among adolescents and young children (80% of those under 18 years old) and older adults (40% of those over 65 years old). Studies indicate that loneliness is observed more commonly among younger age groups, contrary to the popular perception that it is more common among the elderly. Prolonged feelings of loneliness may be detrimental to an individual's mental health. Research indicates that being alone raises the likelihood of mental health problems such as stress, worry, depression, low self-esteem, and difficulty sleeping. It is evident that social contact is essential for maintaining emotional well-being, even though the relationship between loneliness and mental health is complex. Certain people or groups may be more susceptible to loneliness, according to research. People who belong to minority groups and live in areas without similar communities, who have limited mobility or financial resources and are excluded from social activities, who are estranged from their family, who lack friends and support, who are stigmatized or discriminated against because of a disability, long-term health condition, or mental health issue, who are stigmatized or discriminated against because of their gender, race, or sexual orientation, and who have a history of sexual or physical abuse are some examples of these. In the realm of mental health, machine learning (ML) has

become a viable method for sentiment analysis since it allows for the analysis of massive amounts of data and offers individualized interventions. A variety of machine learning approaches are being applied to comprehend and evaluate people's emotional and mental states. Sentiment analysis techniques powered by machine learning algorithms can be used to examine the emotional content of text data, such as posts on social media, conversations in online forums, and medical notes. Mental health practitioners can identify people who may be more vulnerable to mental health problems, track the effectiveness of treatment, and obtain broad insights into people's emotional wellbeing by automatically identifying sentiment and emotional patterns in these textual data. By means of predictive modeling, machine learning algorithms can utilize past data to forecast future outcomes related to mental health. These algorithms discover potential mental health risk factors in individuals by examining patterns and correlations in the data. Early detection makes it possible to provide timely assistance and interventions, which helps stop mental health issues from getting worse. A subset of machine learning called natural language processing (NLP) is used to analyze large amounts of text, including social media posts or clinical notes, in order to find linguistic signals that might point to mental health issues. From these unstructured textual sources, NLP algorithms can extract insightful patterns and insights that advance our knowledge of mental health issues and make early intervention possible. Additionally, visual data such as brain scans and facial expressions are analyzed using machine learning algorithms to identify patterns that may indicate mental health problems. ML models look at visual clues to help clinicians diagnose and plan treatments by assisting in the evaluation of a patient's emotional state. Sentiment analysis tools backed by machine learning are essential for providing tailored treatment plans. Machine learning algorithms have the capability to facilitate the identification of personalized and optimal treatment solutions for individuals by combining data from several sources, such as genetic information, medical history, environmental factors, and sentiment analysis results. This individualized approach may lead to better treatment results and an improvement in mental health. Even though ML presents a lot of potential for Sentiment Analysis in the context of mental health, it is crucial to solve any potential obstacles. These difficulties include making sure that data is secure and private, dealing with biases in the training set, and carrying out thorough validation tests to guarantee the security and effectiveness of machine learning algorithms. Additionally, promoting equitable access to ML-powered mental health tools and avoiding over-reliance on technology are critical considerations in the responsible deployment of ML in mental health care.

Traditional mental health assessment techniques place a great deal of reliance on self-reported symptoms, which are not always accurate indicators of a person's emotional condition and time-consuming. This can make it difficult for people to appropriately self-report, which increases the risk of misdiagnosis or delayed diagnosis. As a result, there is an increasing demand for novel methods of mental health assessment that offer more accurate and tailored suggestions. The application of machine learning (ML) algorithms is one viable remedy. These are capable of analysing a wide range of data sources, including social media activity, behavioural data, and physiological measurements, to produce more thorough and precise mental health assessments.

In conclusion, there is a clear need for innovative and effective approaches to mental health assessment and care that can address the complex and varied nature of mental health issues. To address these challenges, proposed method uses advanced NLP techniques to analyse text-based responses from individuals and generate personalized recommendations based on their unique emotional state. By analyzing the language used by individuals in response to specific questions, proposed model aims to identify patterns and indicators of mental health issues, which can be used to provide tailored recommendations that are specific to an individual's emotional needs.

The research objectives based on the content provided are as follows:

1. Develop a self-companion model capable of assessing the mood of the user and tracking their emotional state over time.
 2. Categorize the user's mood into six different categories based on their responses to five verbally or written questions.
 3. Process the user's responses and analyze the emotional content of each sentence to compute an accumulated score by taking the average.
 4. Generate a final score for a specific emotion by aggregating the accumulated scores.
 5. Implement a threshold mechanism to label sentences as neutral if the accumulated score does not surpass a certain threshold.
 6. Pre-train the model using an emotions dataset sourced from Kaggle, comprising sentences spoken in natural language.
 7. Enhance the performance of the model through techniques such as layer normalization, residual connections, and label smoothing.
 8. Evaluate the effectiveness and accuracy of the developed self-companion model in assessing user mood and tracking emotional states.
 9. Explore the applicability and potential benefits of the model in providing personalized support and guidance to individuals.
 10. Contribute to the advancement in the field of NLP models by showcasing the significance of techniques like layer normalization, residual connections, and label smoothing in improving model performance and accuracy.
- By addressing these research objectives, the paper aims to contribute to the field of mental health and NLP by developing an effective self-companion model for mood assessment and tracking emotional states.

2. LITERATURE SURVEY

Mental health in machine learning is being addressed through various techniques that target different types of mental illness and conditions. One such technique is computer vision, which is being used to analyze mental images, MRIs, and X-rays to detect diseases such as schizophrenia. As per the World Health Organization, schizophrenia is marked by substantial impairments in perception of reality and behavioral changes, which may include persistent delusions, hallucinations, feelings of influence, control, or passivity, disorganized thinking, and more. In the study conducted by Greenstein et al. [1] and Ahmed et al. [2], multivariate machine learning techniques, specifically Random Forest (RF), were suggested using structural magnetic resonance imaging (MRI) of the brain to categorize people with schizophrenia and healthy

controls. The researchers hypothesized that brain measurements could differentiate between the two groups and individuals with a higher probability of being classified as patients would exhibit more severe illness, developmental delays, and a higher genetic risk. 99 healthy controls that were matched for age, sex, and ethnicity, and 98 patients with childhood onset schizophrenia (COS) were classified using 74 anatomic brain MRI sub regions. The researchers also looked into the relationship between copy number variations (CNVs) linked to schizophrenia with the likelihood of illness based on brain measurements, symptoms, premorbid development, and the presence of CNVs. In order to discriminate between the COS (Childhood-Onset Schizophrenia) and healthy control groups, the study showed that the use of multivariate machine learning techniques—specifically Random Forest (RF)—achieved a classification accuracy of 73.7%. Furthermore, those who had a higher chance of being labeled as patients based on brain measures showed worse functioning and less developmental delays. It's interesting to note that possessing copy number variations (CNVs) was linked to a lower risk of being diagnosed with schizophrenia. Notably, the left temporal lobes, bilateral dorsolateral prefrontal areas, and left medial parietal lobes were crucial brain regions for accurate classification. It is generally known that decreased brain connection plays a role in the pathophysiology of schizophrenia.

Jo et al. [3] and Nandan et al. [4] identified schizophrenia using machine learning techniques and network analysis. The researchers looked at the brain connections between 24 healthy individuals and 48 patients with schizophrenia. They created graphs and calculated global and nodal network features by using probabilistic brain tractography. The two groups' network features were compared afterward and utilized for classification. To classify the subjects as either schizophrenia patients or healthy controls, machine learning models such as Support Vector Machine, Random Forest, Naive Bayes, and Gradient Boosting were applied. The results demonstrated a positive level of proficiency in accurately discriminating between the two groups.

Chekroud et al. [5] and Rai et al. [6] employed machine learning to analyze data from the National Survey on Drug Use and Health (NSDUH), an annual survey carried out by SAMHSA, to address the conditions of depression and anxiety. The study offers statistically accurate information on drug misuse and mental disorders in the civilian, non-institutionalized population of the United States. They combined participant-level information from public use files between 2008 and 2014 (N=391,753), excluding those who were younger than 18. The participant's self-reported (binary) response to the question "Did there ever arise a time during the last 12 months when you needed mental health treatment or counseling for yourself but didn't get it?" was the main outcome. Out of 20,785 participants, 6,271 (30%) indicated that they did not get the treatment or counseling that they needed. They were then asked, "Which of these statements explains why you did not get the mental health treatment or counseling you needed?" with 14 specific options and one option for "other" reasons. They used a tree-based machine learning algorithm (Extreme Gradient Boosting (XGBoost)) and developed an open-source software library for deriving individual participant-level measures of variable importance from XGBoost ensembles by breaking down the (directional) impact of each predictor variable for a single participant.

Ahmed et al. [2] and Singh et al. [7] conducted a study to

diagnose various levels of mental disorders, including anxiety and depression, by combining a standard psychological assessment with machine learning algorithms. The researchers assessed the performance of five AI algorithms: CNN, SVM, LDA, KNN, and Linear Regression, using two datasets related to anxiety and depression. When the results were compared using metrics like accuracy, recall, and precision, the CNN-based model yielded the highest accuracy rates, 96% for anxiety and 96.8% for depression. This study highlights the possibility of accurately identifying mental diseases by combining machine learning methods with psychological evaluations. These results offer encouraging perspectives for improving mental health diagnosis and carrying out suitable interventions.

The INTREPID system, a cutting-edge monitoring tool created especially for patients with anxiety disorders during therapy sessions, was described by Katsis et al. [8] and Nandan [9]. The gadget collects physiological data on blood volume, pulse, respiration, heart rate, and galvanic skin response using non-invasive technologies. The subject's affective state is correctly determined when these signals are analysed. This emotional state is divided into predefined categories, including calm, neutral, shocked, nervous, and extremely nervous. To achieve correct categorization, the INTREPID system employs four separate classification algorithms: SVM, ANN, Random Forests, and a Neuro-Fuzzy System. Utilising these sophisticated algorithms on the physiological data that was gathered, the system was able to achieve an astounding 84.3% total classification accuracy. The study by Katsis et al. demonstrates how well the INTREPID system monitors and classifies the affective states of patients with anxiety disorders during treatment sessions. By combining advanced classification algorithms with non-invasive technology, the approach shows great promise for improving our knowledge of and ability to treat anxiety disorders. Hilbert et al. [10] and Ojha et al. [11] employed machine learning techniques in a sample of 32 participants to differentiate between complex and healthy concerns as well as between major depressive illness without generalized anxiety disorder and generalized anxiety disorders. They used multimodal behavioral data from people with significant depressive illnesses, healthy people, and people with generalized anxiety disorders. When a binary Support Vector Machine was used, it was discovered that it was difficult to predict generalized anxiety disorders from clinical questionnaire data alone. However, the classification accuracy for cases and disorders was 90.1% and 67.46%, respectively, when other variables including cortisol levels and grey matter volume were taken into account.

In order to develop depression detection software that utilises many data sources obtained from cellphones, Xu et al. [12] and Rai et al. [16] conducted research. The focus of the study was on feature extraction from GPS modalities, audio files, text messages, and social media data. In addition to sophisticated feature extraction and creation approaches for the analysis of text, audio, and GPS data, special emphasis was paid to text and audio analysis. The study's main goal was to collect information and improve the analysis of features taken from speech and text data. In order to determine which machine learning algorithm performed the best, the researchers assessed the effectiveness of several algorithms as shown in Table 1.

In order to ascertain whether pattern recognition may be utilised to distinguish bipolar disorder patients from healthy controls, Rocha-Rego et al. [13] and Mishra et al. [14, 15]

conducted study. Participants in the study included two groups of individuals with bipolar disease in remission. The researchers employed a classification technique based on Gaussian processes and structural magnetic resonance imaging (MRI) data of grey and white matter. They assessed the algorithm's precision in recognising white matter and grey matter patterns linked to bipolar illness. The study's conclusions showed that, in the first study population, the algorithm identified grey matter patterns with an accuracy of 73%, while in the second cohort, it reached an accuracy of 72%. In a similar vein, the algorithm's accuracy for white matter patterns was 69% in the first population and 78% in the second. These results indicate the potential of using pattern recognition techniques and structural MRI data to differentiate bipolar disorder patients from healthy individuals. The study contributes to the ongoing research on developing objective methods for diagnosing and understanding bipolar disorder.

Spathis et al. [16] and Rai et al. [17] conducted a study where they proposed an innovative end-to-end machine learning model for predicting future sequences of mood based on past self-reported mood data. The study utilized real-world data collected through mobile devices, allowing for mood prediction in real-world settings. The model used multi-task encoder-decoder recurrent neural networks with machine translation and video frame prediction capabilities. The algorithm was able to learn trends from several users thanks to this strategy, which improved predictions for users with scant self-reported data. This model was able to forecast a series of future emotions as opposed to simply one, unlike conventional time series forecasting techniques. 33,000 user-weeks of data were used in the experiments by the researchers. The findings showed that three weeks was the ideal amount of self-reported moods for precise prediction. Furthermore, the study discovered that "valence and arousal" were the two characteristics of mood that multi-task learning models predicted better than conventional or separate models. The performance of the model was also shown to be significantly influenced by variables including mood swings, personality features, and the day of the week, according to the researchers. Overall, the study shows how machine learning methods, particularly the suggested model, can be used to forecast future mood sequences based on historical self-reported data. This has repercussions for comprehending and managing mood-related difficulties in practical contexts.

In their study, Sau and Bhakta [18] and Goel et al. [19] developed a predictive algorithm that uses machine learning to detect depression and anxiety in senior citizens. The model was created to investigate the sociodemographic and health-related traits of senior patients. To evaluate the performance of the model, ten classifiers were evaluated on a dataset of 510 elderly patients using a ten-fold cross-validation approach. The Random Forest (RF) classifier produced predictions with the best accuracy, 89%. 110 elder patients were used as a second dataset on which to test the model's external validity. According to the findings, the RF model's predicted accuracy was 91% and its false positive rate was 10%, which was on par with the gold standard diagnostic instrument.

Vaswani et al. [20] and Srivastava et al. [21] introduced the transformer model architecture for tasks which uses Natural Language Processing. Modern NLP models are used in this study article, such as the BERT model from Google and the GPT family of models from OpenAI. By utilizing only attention processes to interpret input sequences, the transformer model presents a distinctive neural network

architecture that differs from conventional recurrent or convolutional layers. The transformer design, which uses self-attention processes to control input sequences, was presented by the researchers in their work. The transformer consists of numerous layers of feed-forward and self-attention neural networks with an encoder and a decoder. The encoder converts an input sequence into hidden representations, which the decoder utilizes to create the output sequence. Self-attention enables the model to focus on specific input sequence segments, which enables it to identify correlations and long-term associations between various input segments. The transformer model outperforms state-of-the-art models across a range of natural language processing tasks, such as language modeling, machine translation, and text summarization, according to the researchers. The transformer, in particular, proved to be a top model in the NLP business by excelling at the WMT 2014 English-to-German and English-to-French translation tasks [15]. The use of self-attention mechanisms, which enable the model to selectively focus on different portions of the input sequence without depending on sequential processing, is the transformer model's primary novelty. This makes the model less computationally intensive

and enables it to process longer input sequences than earlier models [22]. Table 2 highlights previous studies collectively demonstrate the potential of machine learning techniques in mental health research, highlighting both the promise and challenges of applying these methods to complex psychological and neurological conditions.

Table 1. Comparison on the basis of mean F1 score for various machine learning algorithms

Machine Learning Algorithms	Mean F1 Score in Case of Audio	Mean F1 Score in Case of Text
	Features	Features
Gaussian process	0.48	0.71
Logistic Regression	0.48	0.69
Neural Networks	0.42	0.68
Random Forest	0.44	0.73
Support Vector Machine (SVM)	0.40	0.72
XGBoost	0.50	0.69
K-Nearest Neighbors (KNN)	0.67	0.49

Table 2. Existing research studies on mental health

References	Highlights	Benefits	Limitations
Greenstein et al. [1]	Achieved 73.7% accuracy in differentiating COS patients from healthy controls using structural MRI data and RF.	Large sample size, well-matched control group, use of multivariate machine learning techniques, focus on brain regions crucial for classification.	Moderate classification accuracy, reliance on structural MRI which might not capture all relevant brain abnormalities.
Ahmed et al. [2]	CNN achieved highest accuracy (96% for anxiety, 96.8% for depression) in diagnosing mental disorders.	High accuracy, comparison of multiple algorithms, use of standard psychological assessments.	Potential for overfitting in CNN, limited external validation, may not generalize to all populations.
Jo et al. [3]	Effective discrimination between schizophrenia patients and healthy controls using network analysis and machine learning.	Comprehensive analysis of brain connectivity, use of multiple machine learning models for robust results.	Smaller sample size, potential overfitting due to multiple models, limited generalizability.
Chekroud et al. [5]	Identified key predictors of unmet mental health needs using XGBoost on large NSDUH dataset.	Large, diverse sample size, detailed analysis of unmet mental health needs, development of participant-level variable importance measures.	Self-reported data may be biased; binary outcome may oversimplify complex mental health needs.
Katsis et al. [8]	INTREPID system achieved 84.3% accuracy in classifying affective states of patients with anxiety disorders.	Non-invasive data collection, real-time monitoring capability, use of multiple classification algorithms.	Moderate accuracy, reliance on physiological signals which may not capture all aspects of anxiety, potential technical challenges in real-world deployment.
Hilbert et al. [10]	Achieved 90.1% accuracy in differentiating between complex and healthy concerns using multimodal data.	Use of multimodal data (clinical, cortisol, grey matter volume), high accuracy for cases.	Moderate accuracy for disorders, small sample size, challenges in integrating diverse data types.
Xu et al. [12]	Effective use of multimodal smartphone data for depression detection.	Innovative use of diverse data sources (GPS, audio, text, social media), robust feature extraction methods.	Privacy concerns with smartphone data, variable data quality, potential biases in self-reported data.
Rocha-Rego et al. [13]	Achieved 72-73% accuracy in identifying grey matter patterns linked to bipolar disorder.	Use of structural MRI data, focus on grey and white matter patterns, potential for objective diagnosis.	Moderate accuracy, small sample size, potential for overfitting.
Spathis et al. [17]	Developed an end-to-end model for predicting future mood sequences using past self-reported data.	Use of real-world data, ability to predict mood sequences, multi-task learning approach.	Reliance on self-reported data, variability in data quality, potential challenges in real-world application.
Sau and Bhakta [19]	Achieved 91% accuracy in detecting depression and anxiety in senior citizens using RF	High accuracy, external validation on a second dataset, focus on sociodemographic and health-related traits.	Potential overfitting, limited generalizability to other age groups, reliance on a single classifier.
Vaswani et al. [21]	Introduced the transformer model, which outperformed state-of-the-art NLP models.	Novel self-attention mechanism, high performance in various NLP tasks, broad applicability.	High computational cost, complexity in training and implementation, initial application focused on NLP tasks rather than mental health.

Proposed method highlights the potential of NLP-based models to support self-reflection and provide insights into people's moods. This capability is crucial for managing mental health, as it enables individuals to recognize and address emotional issues proactively.

3. PROPOSED DESIGN

The BERT model is a neural network architecture that is based on the transformer architecture, which allows it to model long-range dependencies in input sequences. The architecture consists of an encoder that processes input sequences in a bidirectional mode allowing having access both the past and future tokens of a sequence. Steps followed by BERT model for sentiment analysis is discussed next.

1. Data Pre-processing:
 - Tokenize the input text using Word Piece tokenization.
 - Add special tokens ('CLS' to indicate the start of the input and 'SEP' to separate different sentences).
 - Pad the sequences to ensure they have the same length.
2. Fine-Tuning:
 - Initialize the BERT model with pre-trained weights.
 - Add a task-specific output layer on top of the BERT encoder.
 - Train the model on labeled data for sentiment analysis.
 - Input: Tokenized and padded sequences.
 - Output: Sentiment labels (positive, negative, or neutral).
 - Update the weights of the task-specific output layer while keeping the pre-trained BERT encoder fixed.
 - Adjust the internal representation of words and their relationships to optimize the sentiment analysis task.
3. Inference:
 - Given a new, unseen input sentence:
 - Tokenize the input and add the 'CLS' and 'SEP' tokens.
 - Feed the tokenized sequence through the pre-trained BERT encoder.
 - Obtain a sequence of hidden representations from the BERT encoder.
 - Pass the representation of the 'CLS' token through the task-specific output layer.
 - Generate a probability distribution over the sentiment labels (positive, negative, or neutral).
 - Predict the sentiment label with the highest probability as the predicted sentiment for the input text.

This algorithm outlines the steps involved in using BERT for sentiment analysis. It starts with data pre-processing, followed by fine-tuning the BERT model on labeled data, and concludes with the inference process for predicting the sentiment label of new input sentences.

There are two stages to the suggested system. The initial stage involves pre-processing the data needed to construct and optimize the BERT model. Using the BERT model, attention models for mood classification are developed and trained in the second phase. Many well-known frameworks are used to complete the tasks related to natural language processing. The first framework is the popular open-source library Transformers 4.24.0, which was created especially for tasks involving natural language processing. Google's TensorFlow 2.9.2 is used because it provides a feature-rich toolkit for creating and refining machine learning models. Another well-known open-source package that makes loading and processing datasets easier is TensorFlow Datasets 2.7.0, which offers a number of functions and classes. Tokenization of

different text inputs and support for several tokenization algorithms are provided by Tokenizers 0.13.2, a fundamental framework that facilitates the preprocessing of vast amounts of text data.

These frameworks, which include TensorFlow, Transformers, TensorFlow Datasets, and Tokenizers, are essential to the creation of the suggested BERT model-based mood categorization model. They offer the tools and functions needed to efficiently handle and examine data in natural language.

3.1 Dataset description

The dataset being discussed is an English Twitter dataset made up of tweets that convey the six emotions that are common to all people: fear, pleasure, sadness, anger, and surprise. A research team under the direction of Elvis Saravia assembled this dataset, which was made available as part of the Conference on Empirical Methods in Natural Language Processing Proceedings in 2018.

Researchers collected a distinct set of English tweets encompassing eight essential emotions—angry, expectancy, disgust, fear, joy, sadness, surprise, and trust—using the Twitter API in order to compile this dataset. They meticulously analyzed these tweets to ensure their relevance and intelligibility, producing a final dataset of about 9,000 tweets that represented one of the six basic emotions stated above. Many pre-processing methods were used to guarantee the dataset's quality. For instance, a Sentiment Analysis tool was used to examine the general sentiment expressed by the tweets, and a Part-of-Speech (PoS) tagger was used to determine the grammatical structure of the tweets. Moreover, duplicate tweets were removed from the dataset and non-English tweets were located and deleted. These pre-processing methods improved the dataset's dependability and quality. All things considered, this dataset is excellent for researching the connection between emotions and natural language. The annotations are comprehensive and accurate, and the dataset has been refined through meticulous pre-processing. Different categories of training data are shown in Figure 1.

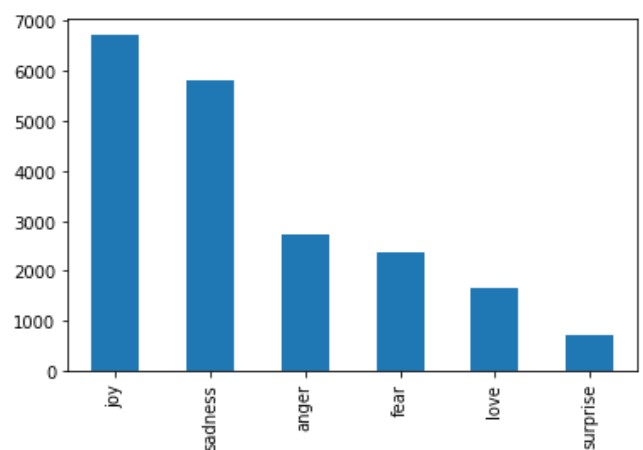


Figure 1. Training data categorization

Here are a few examples of labeled text data for emotion classification:

```
{
  "label": 1,
  "text": "I am so excited to go on vacation next week!"
}
```

```

}
{
"label": 2,
"text": "I just got a promotion at work and I am feeling really
proud of myself."
}
{
"label": 3,
"text": "The sight of blood makes me feel queasy and
anxious."
}
{
"label": 4,
"text": "I just got the news of my outcome, and I'm devastated
and unhappy."
}
{
"label": 5,
"text": "My best friend was involved in an automobile
accident, and I just received a call about it. This has me really
worried and nervous."
}

```

3.2 Data preparation

1. Tokenization: Tokenization is the initial stage of data preparation, in which unprocessed text input is divided into discrete words or tokens. The Tokenizers framework's Byte Pair Encoding (BPE) algorithm is used in the suggested methodology. By breaking down words into sub word units, BPE lowers the number of model parameters and allows the model to handle terms that are not in the lexicon.

2. Encoding: Following the tokenization process, the tokens are encoded into numerical values that the model can comprehend. The basic BERT model, which includes a preset vocabulary, is used in the encoding process. Based on this vocabulary, a distinct numerical value is assigned to each token, resulting in a sequence of numerical values that represent the tokenized text data.

3. Batching: To maximize the training process, batching entails dividing the encoded input into groups. With a batch size of 8, or eight sequences of encoded text data per batch, the tokenized and encoded data is transformed into TensorFlow datasets, which make it easier to feed the data into the model during the training phase.

4. Extra Processing: To incorporate more information, the input data is processed once more during the batching process. This includes attention masks and token type IDs. The attention mask is a binary mask that indicates which tokens are actual words and which are padding tokens. Token type IDs are used in tasks that require segmenting input sequences into multiple parts, such as the next sentence prediction task.

5. Training and Validation Datasets: Training and validation datasets are created from the prepared data. The model is trained using the training dataset, while its performance is evaluated using the validation dataset. To ensure consistent evaluation on the same data distribution in each epoch, the validation dataset is not shuffled.

By following these steps, the raw text data is transformed into a suitable format for training and evaluation of trained models, specifically in the context of the proposed methodology.

3.3 Base model description

The base model used in this research is the base-BERT-uncased, which was pre-trained using the transformers architecture in a self-supervised manner. The two main objectives of the pre-training phase were Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). By taking into account the surrounding context, MLM assists the model in predicting masked words in a phrase. This helps the model develop bidirectional representations and comprehend the contextual links between words. In order to help the model learn to recognize sentence relationships and comprehend context, NSP concentrates on predicting the relationship between two sentences. The underlying model was trained using a distributed training method called pod configuration. A total of 16 TPU chips, or four cloud TPUs, were used to process the data in parallel. The pre-training objectives and distributed training process improve the model's understanding of natural language and contextual linkages, making it applicable to a variety of downstream tasks.

3.4 BERT as classifier model

The research utilized the BERT architecture as the foundation for their machine learning model. The BERT base model from the Hugging Face Transformers Library was downloaded and fine-tuned for sentiment analysis using a dataset of tweets that included the fundamental human emotions of fear, joy, love, sadness, and surprise.

During the fine-tuning phase, a batch size of 8 and 3 epochs were chosen. The number of training steps was determined based on the size of the training dataset and the number of epochs. To optimize the training process, a polynomial decay learning rate scheduler was employed with an initial learning rate of $5e-5$, gradually reducing the learning rate over time.

The 'Adam' optimizer was used during training phase which combines the advantages of AdaGrad and RMSProp algorithms. Early stopping was implemented to prevent over fitting, monitoring the model's performance on a validation set and stopping training when the model no longer improved on the validation set.

Following training, an independent test set of Twitter tweets was used to assess the model's performance. For every one of the six emotions, F1 score, precision, recall, and accuracy were computed. Furthermore, a confusion matrix was created in order to evaluate the model's effectiveness in each of the emotion categories. This approach allowed the researchers to fine-tune BERT for sentiment analysis specifically on the emotions of fear, joy, love, sadness, and surprise, and evaluate its effectiveness using comprehensive metrics and a confusion matrix.

This work suggests a methodology that uses NLP to classify individual feelings into six groups. Transformer models were refined on a dataset of emotions. Using the transformers architecture, a sizable corpus of English text data was used to train the model. It made use of the "Adam" estimator with an initialization rate of $1e-4$ and an average decay value of 0.01 with a batch size of 256. In order to iteratively increase the accuracy of the model on both the initial training and test datasets, multiple epochs were employed throughout the training phase.

4. RESULTS AND DISCUSSIONS

The model was trained for a total of one million steps, where each step corresponds to a single update of the model's parameters based on a batch of 256 samples. The sequence length during training was limited to 128 tokens for 90% of the steps and 512 tokens for the remaining 10%. This allowed the model to handle both shorter and longer input sequences effectively.

During training, the Adam optimizer was used with a learning rate of 1e-4, beta1 and beta2 values of 0.9 and 0.999, and a weight decay of 0.01. A warm-up period of 10,000 steps was employed, during which the learning rate increased linearly. After the warm-up, the learning rate gradually decreased. The model's architecture, based on transformers, enabled it to learn bidirectional representations of sentences, which is crucial for many downstream tasks in natural language processing. By using a larger batch size of 256, the training process could be more efficient, leveraging the computational power of multiple TPUs. The inclusion of sequence length variations during training allowed the model to handle different input lengths commonly encountered in real-world data.

Overall, the model was pre-trained on a substantial amount of English data using self-supervised learning with the transformer's architecture. During fine-tuning, a batch size of 256 and Adam optimization with specific hyperparameters were employed as shown in Table 3.

Table 3. Configuration of hyperparameters

Hyperparameter Used	Value
Optimizer	Adam
Class of Learning Rate	Polynomial Decay
Starting Learning Rate	5e-05
Decay Steps	6000
Ending Learning Rate	0.0
Power	1.0
Cycle	False
Decay	0.0
Beta 1	0.9
Beta 2	0.999
Epsilon	1e-07
Amsgrad	False

These settings contribute to the model's ability to learn contextual representations and handle various input lengths effectively. The summary of machine learning model is shown in Figure 2.

```
Model: "tf_bert_for_sequence_classification"
```

Layer (type)	Output Shape	Param #
bert (TFBertMainLayer)	multiple	109482240
dropout_37 (Dropout)	multiple	0
classifier (Dense)	multiple	4614

```

Total params: 109,486,854
Trainable params: 109,486,854
Non-trainable params: 0

```

Figure 2. Model parameters summary

The training process involves iteratively training a model on dataset to achieve accurate classification of data into appropriate categories. The model undergoes multiple epochs,

which represent complete iterations over the training dataset.

For epoch 0, the training process started with a train accuracy of 0.8637 and an initial train loss of 0.3768. This suggests that although the model initially made some classification mistakes, it was able to attain 86.37% accuracy on the training set. With a validation accuracy of 0.9345, the validation loss was 0.1393. These outcomes show how well the model performed on the validation dataset, demonstrating its potent generalization capability.

The model showed notable progress in epoch 1, yielding a train accuracy of 0.9426 and a train loss of 0.1185. This implies that the model improved its accuracy and produced fewer classification errors. With a validation accuracy of 0.9380, the validation loss was 0.1309. On the validation data, the model performed well even though the validation accuracy was somewhat lower than it was in the prior epoch. Epoch 2 saw an even greater improvement in the model's performance, with a train accuracy of 0.9583 and a train loss of 0.0785. The validation loss was 0.1144 with a validation accuracy of 0.9385. These findings demonstrate that neither the accuracy nor the generalization ability of the model has declined.

The model achieved a train accuracy of 0.9701 and a train loss of 0.0621 in epoch 3, indicating lower classification mistakes and higher accuracy. With a validation accuracy of 0.9385, the validation loss did not change from the previous epoch. This implies that the accuracy and generalization capacity of the model have stabilized. Table 4 displays the results for the first four epochs of the training phase. Figure 3 shows accuracy by epoch.

Table 4. Training statistics

Epoch No.	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
0	0.3768	0.8637	0.1393	0.9345
1	0.1185	0.9426	0.1309	0.9380
2	0.0785	0.9583	0.1144	0.9385
3	0.0621	0.9701	0.1144	0.9385

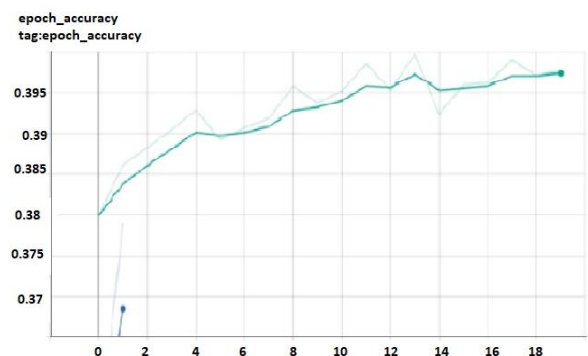


Figure 3. Epoch wise accuracy as depicted by tensor board

Table 5. Training results

Parameters	Results
Train Loss (Categorical Cross Entropy)	6.21%
Train Accuracy	97.01%
Validation Loss	11.44%
Validation Accuracy	93.85%
Number of Epochs	3

As demonstrated in Table 5, the model's overall classification accuracy was high, exceeding 97% on the

training set. Throughout the training procedure, the validation accuracy stayed high, demonstrating the model's good generalization to new data. A popular option for machine learning models, the float32 precision was used to train the model. Adam was the optimization algorithm utilized, which is a well-liked optimizer for changing the weights of neural networks during training. In order to avoid over fitting, the learning rate was gradually decreased over time using the Polynomial Decay method of scheduling. Figure 4 shows the adjustments for learning rate. A high-performing machine learning model was effectively developed during the training procedure with remarkable precision as shown in Figure 5.

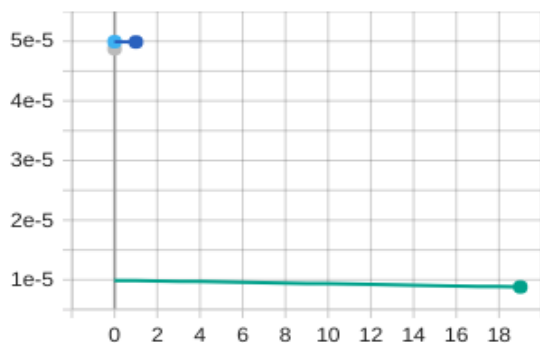


Figure 4. Learning rate adjustment

	precision	recall	f1-score	support
0	0.96	0.92	0.94	813
1	0.87	0.94	0.91	712
2	0.95	0.96	0.95	2028
3	0.87	0.81	0.84	492
4	0.96	0.97	0.97	1739
5	0.94	0.71	0.81	216
accuracy			0.94	6000
macro avg	0.92	0.89	0.90	6000
weighted avg	0.94	0.94	0.93	6000

Figure 5. Confusion matrix of the BERT model

5. CONCLUSIONS

In addition to highlighting the importance of mental health, this research study presents a model that uses NLP to categorize people's moods into six different groups. The capacity of the model to examine user inputs and obtain a thorough comprehension of their mood has significant consequences for the management of mental health. The results highlight how critical it is to address mental health issues in the fast-paced world of today, where emotional difficulties are exacerbated by factors like isolation. Using NLP and machine learning techniques, the model provides a useful tool for people to express their feelings, think back on their day, and get individualized insights on their mental health.

This study's machine learning model, which is based on BERT, performed remarkably well on a Twitter dataset that included six basic emotions. The model successfully categorized messages into the different emotion groups, opening up a variety of uses, such as sentiment analysis for social media marketing campaigns or tools for monitoring mental health. The model demonstrated its effectiveness in identifying and classifying users' emotions, with accuracy

surpassing 94% in all six mood categories. This opens the door for useful applications in mental health monitoring and support. It's important to understand that the model cannot take the place of expert mental health care, though. When dealing with serious mental health issues, it is still imperative to get the proper assistance from licensed healthcare providers, even though it can provide insightful information and support.

NLP and transformer models have significant potential in the field of mental health. By providing tools for early detection and scalable support, these technologies can enhance mental health care and improve overall well-being. The proposed methodology for emotion classification using transformer models demonstrates promising results, with high accuracy and strong generalization capabilities, underscoring the value of NLP in mental health. It elaborates the value of NLP in managing mental health and encourages further exploration and application of transformer models in mental health domains.

In summary, this research paper exemplifies the potential of NLP-based models to contribute to mental health management by providing insights into individuals' moods and facilitating self-reflection. By addressing mental health concerns and promoting well-being, we can strive towards a society that values and supports emotional health alongside physical health.

REFERENCES

- [1] Greenstein, D., Malley, J.D., Weisinger, B., Clasen, L., Gogtay, N. (2012). Using multivariate machine learning methods and structural MRI to classify childhood onset schizophrenia and healthy controls. *Frontiers in Psychiatry*, 3: 53. <https://doi.org/10.3389/fpsyt.2012.00053>
- [2] Ahmed, A., Sultana, R., Ullas, M.T.R., Begom, M., Rahi, M.M.I., Alam, M.A. (2020). A machine learning approach to detect depression and anxiety using supervised learning. In 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Gold Coast, Australia, pp. 1-6. <https://doi.org/10.1109/CSDE50874.2020.9411642>
- [3] Jo, Y.T., Joo, S.W., Shon, S.H., Kim, H., Kim, Y., Lee, J. (2020). Diagnosing schizophrenia with network analysis and a machine learning method. *International Journal of Methods in Psychiatric Research*, 29(1): e1818. <https://doi.org/10.1002/mpr.181>
- [4] Nandan, D., Singh, M.K., Kumar, S., Yadav, H.K. (2022). Speaker identification based on physical variation of speech signal. *Traitement du Signal*, 39(2): 711-716. <https://doi.org/10.18280/ts.390235>
- [5] Chekroud, A.M., Zotti, R.J., Shehzad, Z., Gueorguieva, R., Johnson, M.K., Trivedi, M.H., Cannon, T.D., Krystal, J.H., Corlett, P.R. (2016). Cross-trial prediction of treatment outcome in depression: A machine learning approach. *The Lancet Psychiatry*, 3(3): 243-250.
- [6] Rai, A., Sharma, D., Rai, S., Singh, A., Singh, K.K. (2021). IoT-aided robotics development and applications with AI. In *Emergence of Cyber Physical System and IoT in Smart Automation and Robotics: Computer Engineering in Automation*, pp. 1-14. https://doi.org/10.1007/978-3-030-66222-6_1
- [7] Singh, M.K., Kumar, S., Nandan, D. (2023). Faulty voice diagnosis of automotive gearbox based on acoustic

- feature extraction and classification technique. *Journal of Engineering Research*, 11(2): 100051. <https://doi.org/10.1016/j.jer.2023.100051>
- [8] Katsis, C.D., Katertsidis, N.S., Fotiadis, D.I. (2011). An integrated system based on physiological signals for the assessment of affective states in patients with anxiety disorders. *Biomedical Signal Processing and Control*, 6(3): 261-268. <https://doi.org/10.1016/j.bspc.2010.12.001>
- [9] Nandan, D. (2020). An efficient antilogarithmic converter by using correction scheme for DSP processor. *Traitement du Signal*, 37(1): 77-83. <https://doi.org/10.18280/ts.370110>
- [10] Hilbert, K., Lueken, U., Muehlhan, M., Beesdo-Baum, K. (2017). Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: A multimodal machine learning study. *Brain and Behavior*, 7(3): e00633. <https://doi.org/10.1002/brb3.633>
- [11] Ojha, M.K., Rai, A., Prakash, A., Tiwari, P., Gupta, D. (2022). Cuckoo search constrained gamma masking for MRI image detail enhancement. *Traitement du Signal*, 39(4): 1387-1397. <https://doi.org/10.18280/ts.390433>
- [12] Xu, A.J., Flannery, M.A., Gao, Y., Wu, Y. (2019). Machine learning for mental health detection. <https://digitalcommons.wpi.edu/mqp-all/6732/>.
- [13] Rocha-Rego, V., Jogia, J., Marquand, A.F., Mourao-Miranda, J., Simmons, A., Frangou, S. (2014). Examination of the predictive value of structural magnetic resonance scans in bipolar disorder: A pattern classification approach. *Psychological Medicine*, 44(3): 519-532. <https://doi.org/10.1017/S0033291713001013>
- [14] Mishra, A.R., Panchal, V.K., Kumar, P. (2020). Similarity search based on text embedding model for detection of near duplicates. *International Journal of Grid and Distributed Computing*, 13(2): 1871-1881.
- [15] Mishra, A.R. (2019). Impact of feature representation on supervised classifiers—A comparative analysis. *Global Sci-Tech*, 11(2): 69-74. <http://doi.org/10.5958/2455-7110.2019.00010.7>
- [16] Spathis, D., Servia-Rodriguez, S., Farrahi, K., Mascolo, C., Rentfrow, J. (2019). Sequence multi-task learning to forecast mental wellbeing from sparse self-reported data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, USA, pp. 2886-2894. <https://doi.org/10.1145/3292500.3330730>
- [17] Rai, A., Kundu, K., Dev, R., Nayak, S., Keshari, J.P., Nandan, D. (2023). Special vehicle like ambulance recognition and security system using mobility accredit system. *Revue d'Intelligence Artificielle*, 37(2): 509-515. <https://doi.org/10.18280/ria.370228>
- [18] Sau, A., Bhakta, I. (2017). Predicting anxiety and depression in elderly patients using machine learning technology. *Healthcare Technology Letters*, 4(6): 238-243. <https://doi.org/10.1049/htl.2016.009>
- [19] Goel, K., Nagar, M., Tiwari, M., Mishra, A.R., Chauhan, S.S. (2024). Enhanced ultrasonic cane for visually impaired people. In *2024 2nd International Conference on Disruptive Technologies (ICDT)*, Greater Noida, India, pp. 500-505. <https://doi.org/10.1109/ICDT61202.2024.10489694>
- [20] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach California, USA, pp. 6000-6010.
- [21] Srivastava, P., Mishra, A.R., Chauhan, S.S. (2024). Deep Learning-based lung cancer detection enhancing early diagnosis and treatment outcomes. In *2024 2nd International Conference on Disruptive Technologies (ICDT)*, Greater Noida, India, pp. 1092-1096. <https://doi.org/10.1109/ICDT61202.2024.10489161>
- [22] Babu, N.V., Kanaga, E.G.M. (2022). Sentiment analysis in social media data for depression detection using artificial intelligence: A review. *SN Computer Science*, 3(1): 74. <https://doi.org/10.1007/s42979-021-00958-1>