Chapter 1

Introduction

Work submitted in IMI New Delhi

1.1 Introduction

Mental health, which encompasses psychological, emotional, and social well-being, is crucial for the overall functioning of an individual. It influences how people think, feel, and behave throughout their lives. Good mental health affects various aspects such as quality of life, stress management, decision-making, relationships, and work productivity. Mental health conditions such as manic-depressive illness and emotionally unstable personality disorder, dissociative states, excessive worry, acute episodes of intense fear, elevated mood with increased energy, irrational suspicion, intense and specific fears, a severe mental disorder characterized by distorted perceptions of reality, and issues related to nutrition intake or rest patterns can manifest in individuals across all life phases, irrespective of their sex, cultural heritage, or socioeconomic status.

1.1.1 Online Social Network

An **Online Social Network (OSN)** is a web-based service that facilitates virtual connections among individuals with shared passions, upbringings, and pursuits. The swift growth of OSNs like Twitter and Facebook has led to their widespread use for expressing opinions and emotions. This has provided scientists with an innovative and potent means to identify emotional states, interpersonal exchanges patterns, activities, and interpersonal behaviours. In recent years, scholars from diverse disciplines have performed numerical examinations of various health conditions and psychological disturbances by analyzing information gleaned from online social networking sites. One notable example is **Sina Weibo,** a prominent social media platform in the Chinese community, which had over 418 million users as of 2021 [10].

1.1.2 Depression

Major depressive disorder, commonly referred to as depression, is one of the most prevalent mental health conditions. According to the **World Health Organization (WHO)** [11], depression affects over 300 million people worldwide. Recent data from WHO indicates that mental health disorders, particularly clinical melancholia, rank among the primary sources of impairment and substantially impact the worldwide toll of disease. The **National Institute of Mental Health (NIMH)** [12] and the **Global Burden of Disease Study (GBD**) [13] also highlight the widespread impact of depression. Depression can lead to significant psychological distress, self-harm, and suicidal tendencies. Despite the availability of

Work submitted in JMI New Delhi

psychotherapy, medical treatments, and alternative therapeutic approaches, 76%-85% of individuals in developing and emerging economies continue to lack appropriate care. This treatment gap can be considered due to both a scarcity of healthcare assets and challenges in making accurate early-stage assessments, which hinder timely diagnosis and treatment.

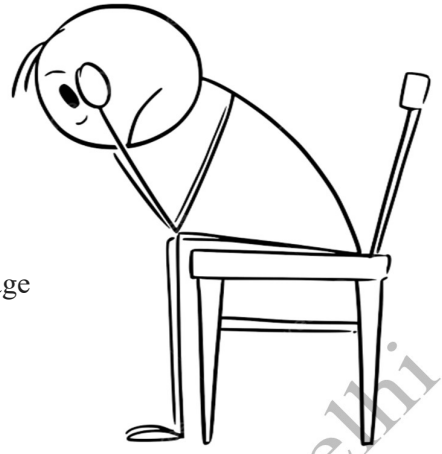

Fig 1.1 : Depressed User illustration

1.1.3 Heterogenous Input Data

The analysis of mental health, particularly depression detection, using Online Social Networks (OSNs) relies on the vast array of heterogeneous input data generated by users. Social media platforms like Twitter, Facebook, and Sina Weibo provide a rich tapestry of information that can be leveraged for this purpose. The types of data available include:

1.1.3.1 Textual Content :

- **Posts and Comments:** Users regularly post updates, share thoughts, and comment on various topics. These textual expressions often reflect their mood, state of mind, and emotional well-being. Linguistic analysis can detect signs of depression, such as negative sentiment, use of depressive language, and discussion of topics related to sadness, hopelessness, and anxiety.
- **Private Messages:** While typically not publicly accessible due to privacy concerns, in research settings with appropriate permissions, private messages can offer deeper insights into a user's mental state.

1.1.3.2 Visual Content:

- **Photos and Videos**: Images and videos posted by users can reveal much about their lifestyle, activities, and emotional expressions. Automated image analysis techniques can identify visual indicators of depression, such as changes in appearance, facial

Work submitted in JMI New Delhi

expressions, and the content of shared media. For example, a tendency to share dark or somber images might correlate with depressive moods.

- **Profile Pictures**: Changes in profile pictures can sometimes indicate changes in mood or self-perception. Frequent changes or the nature of these images can be telling.

1.1.3.3. Behavioural Data:

- **Engagement Patterns**: How users interact with others on social media, including likes, shares, retweets, and comments, provides valuable behavioural data. A decrease in social engagement or a shift towards more negative interactions can be indicative of deteriorating mental health.
- **Activity Levels**: The frequency and timing of posts can also be revealing. Irregular posting patterns or a sudden drop-in activity might suggest a depressive episode.

1.1.3.4 Social Network Data:

- **Followers and Following**: The nature and dynamics of a user's social network, including who they follow and who follows them, can influence and reflect their mental health. For instance, users who follow mental health support groups or frequently interact with supportive communities might be seeking help or experiencing mental health issues.
- **Network Centrality**: Users who become more isolated or whose interactions within their network diminish may be experiencing social withdrawal, a common symptom of depression.

1.1.3.5 Sentiment and Emotion Analysis:

- **Emotional Tone**: Analyzing the emotional tone of posts over time can help identify trends towards increased negativity, anxiety, or hopelessness.
- **Hashtag Usage**: Specific hashtags related to mental health, depression, and other emotional states can provide additional context and indicators of a user's mental state.

The integration of these diverse data types allows for a more comprehensive and nuanced understanding of an individual's mental health.

Work submitted in JMI New Delhi

1.2 Characteristics of Users

Based on textual content, visual content, and behavioural data, users divided into two categories: Normal Users and Depressed Users. Their characteristics are defined as follows:

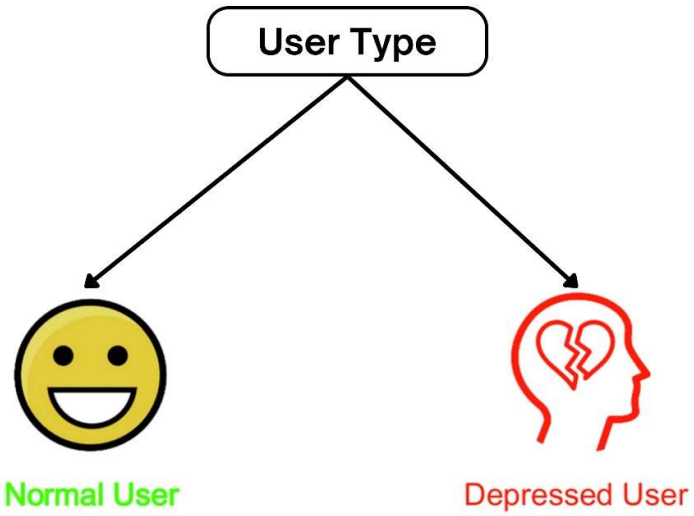

Fig 1.2 : Types of User, Normal User (left) and Depressed User (right)

1.2.1 Characteristics of Normal Users

1.2.1.1 Textual Content:

- **Positive Language:** Normal users tend to use positive, neutral, or varied language in their posts and comments. Their content often includes expressions of happiness, excitement, and general well-being.
- **Balanced Topics**: Posts cover a wide range of topics, including daily activities, hobbies, interests, and interactions with friends and family. The tone of their writing is generally stable and reflects a balanced emotional state.

Work submitted in elhi

1.2.1.2 Visual Content:

- **Varied Imagery**: Normal users frequently share images and videos that reflect a wide array of activities and interests. These may include photos from social events, vacations, hobbies, and day-to-day life.
- **Consistent Appearance**: Profile pictures and shared images usually show the user in a consistent, healthy appearance. Changes in profile pictures are regular but not drastic.

1.2.1.3 Behavioural Data:

- **Regular Engagement**: Normal users exhibit regular and consistent engagement with their social network. They like, share, comment, and interact with others at a steady rate, reflecting a stable level of social activity.
- **Stable Activity Patterns**: Posting and interaction patterns are relatively stable, with users maintaining a regular presence on the platform without significant periods of absence.

1.2.2 Characteristics of Depressed Users

1.2.2.1 Textual Content:

- **Negative Language**: Depressed users frequently use negative language in their posts and comments. They may express feelings of sadness, hopelessness, worthlessness, or self-criticism. Posts may also contain references to loneliness, fatigue, and despair.
- **Focused on Pain and Struggles**: Their content often centers on personal struggles, emotional pain, and mental health issues. There may be repetitive themes of feeling overwhelmed or disconnected from others.

Work submitted in JMI New Delhi

1.2.2.2 Visual Content:

- **Sombre Imagery**: Depressed users might share fewer images, or the images shared may have a somber or melancholic tone. Photos may reflect a lack of energy or enthusiasm, and there might be a noticeable absence of social events or happy moments.
- **Changes in Appearance**: Profile pictures and other shared images may show significant changes in appearance, such as signs of neglect or lack of self-care. These changes can be sudden and indicative of a declining mental state.

1.2.2.3 Behavioural Data:

- **Decreased Engagement**: Depressed users often show reduced interaction with their social network. They may like, share, and comment less frequently, and their engagement with others diminishes over time.
- **Irregular Activity Patterns**: Their activity on the platform may become erratic, with periods of high activity followed by long absences. This irregularity can signal emotional turbulence or withdrawal.

By distinguishing these characteristics, researchers and mental health professionals can develop more effective tools for identifying and supporting individuals who may be experiencing depression. Understanding the behavioural and emotional patterns of both normal and depressed users on social media provides valuable insights for early intervention and mental health support.

This dissertation focuses on a multimodal approach to depression detection, leveraging the integration of diverse data types and advanced machine learning techniques to enhance the accuracy and timeliness of depression predictions. By integrating text, images, and social interaction data from OSNs, the proposed model aims to provide a comprehensive analysis that improves the early diagnosis and accessibility of treatment for depression. This approach not only addresses the limitations of traditional methods but also offers a scalable solution

Work submitted in JMI New Delhi

capable of processing vast amounts of real-time data. Through this research, the goal is to contribute to better mental health outcomes by improving the early detection and treatment of depression.

The main contributions of this work are:

1. Detecting depression early based on social media activity, feelings, and emotion.
2. Designing a model using heterogeneous input data. This model uses conventional data-driven learning techniques and advanced neural network architectures to classify normal and depressed users.
3. Implementing a framework based on multimodality to effectively detect depressed users on social media.

The remainder of the report is organised is structured thusly:

Section 2 contain a review of relevant research. Detailed elaboration of proposed work discussed in Section 3. Experimental result and comparison with other technique is in Section 4. Conclusion and prospective research directions are explored in Section 5.

Work submitted in JMI New Delhi

CHAPTER 2

REVIEW OF RELATED WORK

Work submitted in JMI New Delhi

Current approaches for detecting depression online can be divided into two primary categories: (i) Manually extracting features followed by employing traditional machine learning models for classification, and (ii) Utilizing deep learning techniques to automatically extract features and employing deep neural network (DNN) models as categorization tools.

2.1. Traditional Machine Learning (TML)

This approach involves manually identifying characteristics (features) in the data that might be indicative of depression. These features could be numerical values derived from user behaviour, emotions expressed, language used, and writing styles.

Pioneering work by Choudhury et al. [15] explored user behaviour on social media to identify potential depression. They analyzed social engagement, emotions, language, and writing styles to find patterns associated with depression. While their models weren't the most accurate, they established a valuable process for feature analysis and model building.

Wang et al. [16] built on this work by analyzing data from Twitter and Weibo. They focused on sentiment analysis and used specific vocabulary to create rules for measuring depression tendencies in tweets. Their research highlighted the importance of text-based features.

Deshpande et al. [18] took a different approach, using natural language processing (NLP) to depict post content as a simplified numerical representation (Term Frequency Array). This allowed the model to automatically learn hidden features within the text data. Their Naive Bayes classifier performed better than a Support Vector Machine classifier using this method.

Shen et al. [19] developed a well-labelled dataset for identification of depressive symptoms on the social media platform, which has been extensively utilized by numerous investigators. They also suggested a multifaceted methodology that considers text, social behaviour, and posted images. Their model effectively learned hidden patterns within these features.

Recent TML research has shown promise:

Mustafa et al. [17] used a technique called Term Frequency-Inverse Document Frequency to evaluate the importance of terms in posts. Their model, utilizing a single-dimensional convolutional neural network architecture (CNN-1D), attained good results. This was the pioneering research to implement an artificial neural network for identifying depressive indicators in online social platforms.

2.2. Neural Network-Based Methodologies

These methods aim to analyze both user social behaviour and multimedia content (text, pictures, videos). Text analysis is a major focus area. Investigators employ natural language processing methods to transform written content into multidimensional numerical representations, allowing the model to autonomously learn word characteristics. Some studies combine manually extracted features with deep neural network models or integrate traditional classifiers with deep learning models to enhance efficacy. These multifaceted and combined strategies have demonstrated success in diverse social media examination tasks, including the identification of depressive indicators.

Orabi et al. [20] evaluated several deep learning models commonly used for NLP classification tasks. They employed a previously established Word2Vec framework to represent tweet written content as vectors. Their experiments demonstrated that a one-dimensional convolutional neural network with a particular aggregation architecture performed best. Unlike recurrent models (RNNs) or Long Short-Term Memory (LSTM) networks, CNNs achieved better results for depression detection in this study.

Sadeque et al. [21] proposed a new approach using Gated Recurrent Units (GRUs) that considers the order in which tweets are posted. They scan tweets one by one and assess the user's depression level after each tweet. This avoids processing a large number of tweets from users who are already clearly depressed.

Expanding upon earlier research [19], Shen and colleagues [22] constructed an advanced neural network model capable of adjusting and transfer knowledge across different social media platforms. This is important because models trained on one platform may not work well on others. Recent DL research continues to make progress.

Gui et al. [23] investigated how the model's accuracy is affected by the balance between the quantity of individuals exhibiting and not exhibiting depressive symptoms in the training data. They found the best accuracy when the numbers are roughly equal. They also explored using reinforcement learning to further improve model performance.

Lin and collaborators [24] utilized a widely recognized pre-trained framework known as BERT to encode lexical representations. They incorporated information from integrating both textual and visual characteristics to achieve better classification results.

CHAPTER 3

PROPOSED METHODOLOGY

Work submitted in JMI New Delhi

3.1. Proposed Methodology

Our proposed methodology for identifying depressive indicators in online social platforms leverages prior research that identified valuable features (Fig:3.1). These features include the timing and originality of user posts, as well as the colour properties of shared images. Inspired by these findings, we implemented a three-part feature engineering approach targeting user data. This approach focuses on three key aspects: the content of user posts (text-based features), user interaction patterns (social behaviour-based features), and the characteristics of posted images (picture-based features). We developed a set of a set of ten individual-specific indicators for identifying depressive symptoms, encompassing four novel features we propose and two modifications of existing features (details provided in Table 3.1). These features are extracted through statistical methods such as calculating means, standard deviations, and scales. A comprehensive description of each feature and its corresponding formula can be found in Table 3.1. Additionally, Table 3.2 provides a glossary of symbols used throughout this analysis for reference.

MI New Delhi

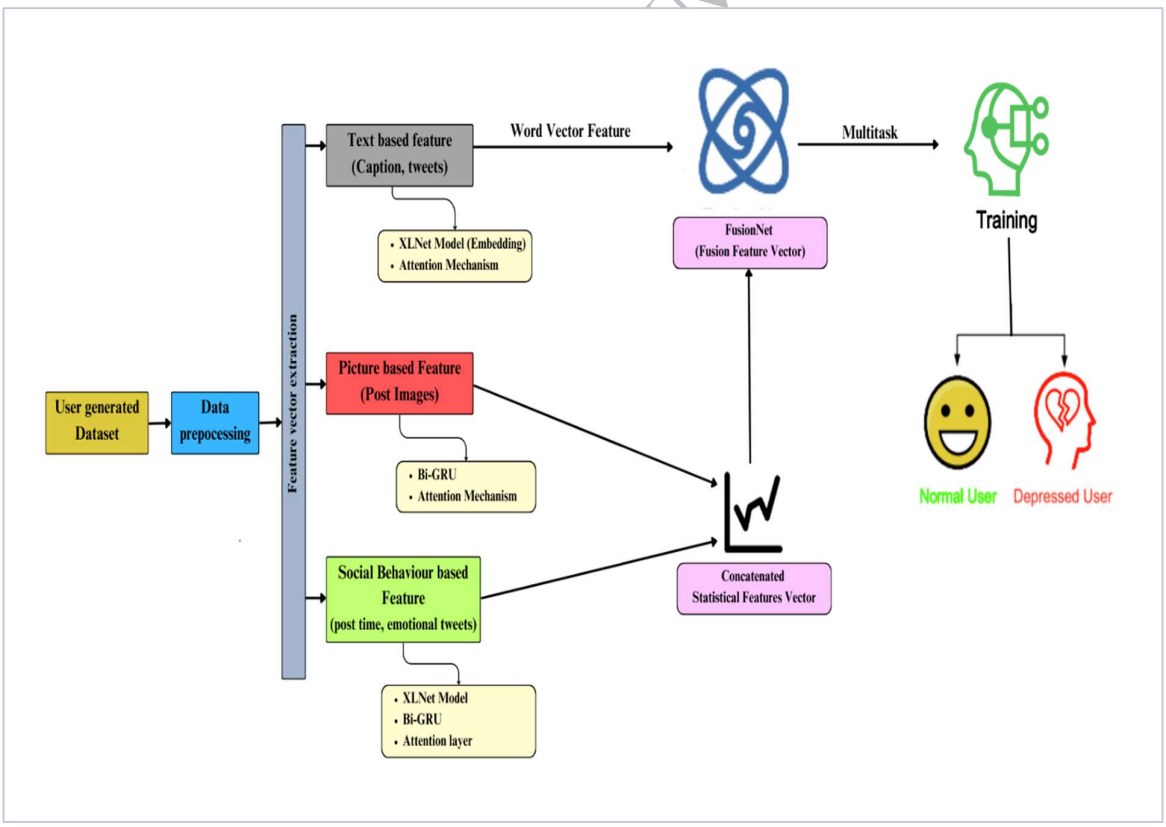

Fig 3.1.: Purposed Model

3.1.1. User generated Dataset

In this dissertation, we utilize the Weibo User Depression Detection Dataset (WU3D) [10], a comprehensive resource for studying social media use and depression. Compiled by Y. Wang et al. in 2022, WU3D offers valuable insights into the online behaviour of individuals with depression. The dataset encompasses a substantial user base, featuring over 10,000 profiles diagnosed with depression alongside a control group of 20,000 users. Each user profile incorporates a rich tapestry of data points, including their social media posts, timestamps of their activity, any shared images, and even their demographic information such as gender. Notably, the depression diagnoses were meticulously assigned and subsequently verified by medical and psychological professionals utilizing the established DSM-5 criteria. This rigorous approach ensures the dataset's credibility and strengthens the foundation for our research on depression and social media behaviour.

Table 3.1 Data Collection Overview

Sr. No.	Dataset	Type	User	Post	Image
01	WU3D	Experiencing Depression	10,325	4,08,797	1,60,481
		Normal	22,245	17,83,113	10,87,556
Total			32,570	21,91,910	12,48,037

Table 3.2 Manually Extract Users Attributes

Sr. No.	Category	Attribute Name	Notation
01.	written content: ψ	Ratio of tweets with negative emotions	ψ_{NP}
		Occurrence of depression-related terms	ψ_{FDW}
02.	Social Behaviour: ϕ	Ratio of initial tweets	ϕ_{POP}
		Ratio of weep-hour posting	ϕ_{PLNP}
		Rate of posting (per week)	ϕ_{PF}
		Variation in posting times	ϕ_{SDPT}
03.	Picture: Υ	Rate of image posting	Υ_{FPP}

Work submitted in JMI New Delhi

		Percentage of images using cool colour schemes	Υ_{PCP}
		Variation in hue	Υ_{SDH}
		Standard deviation of saturation	Υ_{SDS}

3.1.2. Features based on Written-Content (text)

Ratio of tweets with negative emotions: To determine the Ratio of tweets with negative emotions, prior research on Twitter has shown success in differentiating between users experiencing depression and those who are not by focusing on the quantity of such tweets. Given that users can vary significantly in their tweet frequency, we utilize the "ratio" of negative tweets to standardize this attribute, reducing fluctuations that would arise from using the raw "number" of tweets. While expressing negative emotions doesn't entirely capture depressive tendencies, a high proportion of negative tweets can indicate a user's distressed mental state, suggesting potential depression. To label the posts, we utilize the API for Text Sentiment Analysis provided by the Baidu Smart Cloud Platform [14], which classifies emotions into three categories: 0 indicating negative, 1 indicating neutral, and 2 indicating positive. For analysis, we classify labels 1 and 2 as emotions that are not negative and focus on label 0 for negative emotions. For each user, we define the Ratio of tweets with negative emotions ψ_{NP} as described in Equation 1, where $C_{(Po)}$ denotes the overall count of initial tweets and $C_{(le)}$ denotes the overall count of initial tweets exhibiting negative emotions.

$$\psi_{NP} = \frac{1}{C_{(P_o)}} \times C_{(l_e)}, \ \psi_{NP} \in [0, 1] \qquad \text{Equation 1.}$$

Occurrence of depression-related terms: Studies have analyzed the word-related and meaning-based features of post content, quantifying these characteristics through custom-built or referenced compilations of depression-associated terms. Findings from these studies suggest that features derived from the frequent occurrence of depression-related keywords greatly enhance categorization accuracy. We employ "occurrence" to measure in what manner frequently depression- associated words appear during a user's posts, indicating possible signs of depression. Drawing from our own previous research and analysis on Weibo, we have curated a collection of frequently used terms associated with depression. This list is now

Work submitted in JMI New Delhi

employed to determine the occurrence rate of such terms in users' initial posts. The number of depression-associated words in every one tweet, labelled as n_d is identified through comparison with the keyword catalogue. The frequency of depressive words ψ_{FDW} is then calculated as follows:

$$\psi_{FDW} = \frac{1}{C_{(P_o)}} \times \sum_{i=1}^{C_{(P_o)}} n_{d_i}, \ \psi_{FDW} \in [0, \infty) \quad \text{Equation 2.}$$

3.1.3. Features based on Social behaviour

Ratio of initial tweets: Numerous research has indicated that users experiencing depression tend to post more original tweets to convey their negative emotional state, while reposting fewer tweets. Hence, we utilize the ratio of initial tweets to distinguish between users experiencing depression and those who are not. To do this, we determine the overall number of posts $C_{(p)}$, which includes both original tweets and reposts. The proportion of original tweets ϕ_{POP} is then defined as follows:

$$\phi_{POP} = \frac{1}{C_{(P)}} \times C_{(P_o)}, \ \phi_{POP} \in [0, 1] \quad \text{Equation 3.}$$

Ratio of weep-hours posting: Symptoms of depression may be more pronounced during the Wee Hours, leading depressed users to post more tweets during this time. Conversely, typical users are usually asleep and seldom use social media, resulting in fewer tweets during these hours. We utilize the "Tweet Time" feature, defining the Wee Hours period as early hours of the morning, typically between midnight and dawn. This calculation includes all tweets, both original and reposts. The proportion of late-night posts ϕ_{PLNP} is then determined using Eq. (4), where $C_{(tp)}$ represents the overall count of tweets posted during the Weep Hours.

$$\phi_{PLNP} = \frac{1}{C_{(P_o)}} \times C_{(t_p)}, \phi_{PLNP} \in [0, 1] \quad \text{Equation 4.}$$

Work submitted in JMI New Delhi

Rate of posting (per week): Earlier research on Twitter possesses indicated that there exists a discernible disparity in posting frequency between users with normal mood and those experiencing depression. Individuals affected by depression often posting many posts while experiencing depression, relying heavily on social platforms to articulate distress. A seven-day period is regarded as a balanced temporal unit, exhibiting more pronounced cyclical patterns in comparison to a lunar cycle. We determine the span between the initial publication timestamp t_{pE} and the final publication timestamp t_{pL}, tally the aggregate count of posts $C_{(pi)}$ within this timeframe, and then divide by the total number of seven-day periods $C_{(W)}$ to ascertain the posts-per-week rate. This weekly publication rate ϕ_{PF} is expressed by the following equation 5:

$$\phi_{PF} = \frac{|t_{pL} - t_{pE}|}{C_{(W)}} \times C_{(P_i)} \qquad \text{Equation 5.}$$

Standard deviation of posting time: Depressed users tend to concentrate their posting during late-night hours, whereas normal users have a more varied distribution of posting times throughout the day. To capture this pattern, we use the standard variance to capture the clustering of users' publication schedules. A small-scale standard deviation indicates a higher likelihood of posting at specific times. This analysis includes all initial along with reposted posts. The average publication timestamp X_{SDPT} is computed as follows:

$$\overline{X}_{SDPT} = \frac{1}{C_{(P)}} \times \sum_{i=1}^{C_{(P)}} t_{P_i} \qquad \text{Equation 6.}$$

And ϕ_{SDPT} defined as:

$$\phi_{SDPT} = \sqrt{\frac{1}{C_{(P)}} \times \sum_{i=1}^{C_{(P)}} \left(t_{P_i} - \overline{X}_{SDPT}\right)^2} \qquad \text{Equation 7.}$$

We submitted in JMI New Delhi

3.1.4. Features based Image

Rate of image posting: This feature quantifies the frequency of visual content incorporation in users' posts. Depressed users tend to use more written content to convey their feelings and psychological conditions, leading to a reduced frequency of image sharing in comparison to typical users. The overall count of post images is represented by $C_{(\pi)}$. The frequency of picture posting γ_{FPP} is then calculated as follows:

$$\gamma_{FPP} = \frac{1}{C_{(P_o)}} \times C_{(\pi)} \quad \text{Equation 8}$$

Percentage of Images using cool colour schemes: This feature measures how often users post pictures with cold colours. We define the cold colour range as hue values between 30 and 110 degrees and saturation values less than 0.7, based on the HSV colour model. The algorithm converts each pixel's RGB value to HSV, calculates the average colour of the picture, and identifies striking pixels (SP) based on a predefined threshold. The dominant colour of each picture is then determined. We tally the aggregate quantity of shared images that fall within the defined cold colour range, denoted as $C_{(\pi_{cold})}$. The Percentage of Images using cool colour schemes γ_{PCP} is then calculated as follows:

$$\gamma_{PCP} = \frac{1}{C_{(\pi)}} \times C_{(\pi_{cold})}, \gamma_{FPP} \in [0, 1] \quad \text{Equation 9.}$$

Standard deviation of hue and standard deviation of saturation: These metrics indicate the diversity in chromatic properties of a user's shared visual content. We calculate the mean and standard deviation for the chromatic h_μ and intensity s_μ values of the pictures. Individuals experiencing depression typically exhibit a preference for colour tones clustered

Under review submitted in JMI New Delhi

in cooler spectrums and reduced colour intensity, whereas emotionally balanced users demonstrate a more diverse and moderate distribution of colour choices. The mean standard deviations of hue X_{SDH} and saturation X_{SDS} are computed as follows:

$$\overline{X}_{SDH} = \frac{1}{C_{(\pi)}} \times \sum_{i=1}^{C_{(\pi)}} h_{\mu_i} \qquad \text{Equation 10}$$

$$\overline{X}_{SDS} = \frac{1}{C_{(\pi)}} \times \sum_{i=1}^{C_{(\pi)}} s_{\mu_i} \qquad \text{Equation 11}$$

The standard deviations of hue Υ_{SDH} and saturation Υ_{SDS} are defined using the following equations:

$$\gamma_{SDH} = \sqrt{\frac{1}{C_{(\pi)}} \times \sum_{i=1}^{C_{(\pi)}} \left(h_{\mu_i} - \overline{X}_{SDH}\right)^2} \qquad \text{Equation 12}$$

$$\gamma_{SDS} = \sqrt{\frac{1}{C_{(\pi)}} \times \sum_{i=1}^{C_{(\pi)}} \left(s_{\mu_i} - \overline{X}_{SDH}\right)^2} \qquad \text{Equation 13}$$

Work submitted in JMI New Delhi

3.1.5. Multitask Classification Model: FusionNet

In this project, we are adopting a multitask learning methodology where we train multiple related tasks simultaneously, sharing weights and network structures. We developed a deep neural network classifier using a Bidirectional Gated Recurrent Unit with attention-based architecture, integrating features from different modalities like text, social behaviour, images, and words.

We consider two categorization approaches with the goal of identifying individual users experiencing depression: Task 1 is classifying word vectors derived from a pretrained XLNet language model, and Task 2 is classifying manually extracted statistical features. The XLNet model has its weights frozen and acts only as a feature extractor for the text input.

We define separate Error metrics L_1 and L_2 along with coefficients x1 & x2 respectively, to jointly train and optimize the network on both tasks simultaneously. The user's text is embedded by XLNet, passed through layer normalization, then a Bi-GRU with attention to capture important word features. For Task 1, these word features go through feed-forward layers to predict the classification. An auxiliary loss L_1 helps accelerate convergence.

For assignment 2, the lexical features are combined with hand-picked extracted numerical attributes which are batch normalized. This combined representation goes through feed-forward layers before the final classification output. The main loss L_2 optimizes the entire FusionNet network.

The joint multitask optimization objective combines the auxiliary task loss and main task loss, weighted by the manually set x1 and x2 coefficients. Various settings of the extracted features and hyperparameters are evaluated.

Work submitted in JMI New Delhi

CHAPTER 4

EXPERIMENTAL RESULTS

Work submitted in JMI New Delhi

4.1. Assessment Criteria

The empirical analysis in this segment primarily utilizes measures frequently employed in guided learning algorithms for categorization problems. These measures encompass Correct Positive Identifications (CPI), Correct Negative Identifications (CNI), Incorrect Positive Identifications (IPI), and Incorrect Negative Identifications (INI), which enumerate the quantity of cases accurately and inaccurately forecast by the algorithms for each grouping. Precisely, CPI signifies the tally of affected individuals correctly recognized, CNI represents the tally of unaffected individuals correctly recognized, IPI indicates the tally of unaffected individuals erroneously categorized as affected, and INI denotes the tally of affected individuals erroneously categorized as unaffected. Building upon these four fundamental metrics, we can derive additional evaluation measures as follows:

$$\text{Accuracy} = \frac{|\,CPI+CNI\,|}{|\,CPI+CNI+IPI+INI\,|} \qquad \text{Eq. (14)}$$

$$\text{Precision} = \frac{|\,CPI\,|}{|\,CPI+IPI|} \qquad \text{Eq. (15)}$$

$$\text{Recall} = \frac{|\,CPI+CNI\,|}{|\,CPI+INI\,|} \qquad \text{Eq. (16)}$$

$$\text{F}_1\text{-Score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad \text{Eq. (17)}$$

Work submitted in JMI New Delhi

4.2 Performance analysis

To comprehensively evaluate the performance of our proposed depression detection approach, we analyze the results across several metrics. The classification accuracy provides an overall sense of the model's predictive performance. However, due to the class imbalance in our dataset (with more non-depressed users than depressed), accuracy alone may not tell the full story.

Therefore, we also report the precision, recall, and F1 scores broken down by class. The precision for the depressed class indicates how many of the instances predicted as depressed were truly depressed according to the ground truth labels. Recall measures what proportion of the actual depressed cases were successfully detected by the algorithm. The harmonic mean metric integrates accuracy and completeness into one performance measure.

For the non-depressed class, high precision ensures few users are incorrectly flagged as potentially depressed when they are not. High recall is important to catch as many truly non-depressed cases as possible.

Table 4.1: Performance analysis of trained Model

Sr. No.	Feature Type	Model	Epochs	Batch Size	Accuracy
01.	Text based feature	XLNET	50	8	0.9775
02.	Image based feature	Bi-GRU	50	32	0.933
		CNN 1D	50	32	1.00
03.	Social Behaviour based feature	XLNET	50	8	0.943

Work submitted in JMI New Delhi

4.3. Metrics Plots

4.3.1 Training and Validation loss Curve

A plot of **training and validation loss** as a function of the number of training epochs. This curve shown in figure 4.1 represents the model's performance on the training data. As the model is trained on the data, it ideally learns to reduce the loss over time. The validation loss helps to track how well the model generalizes to unseen data and avoid overfitting to the training data.

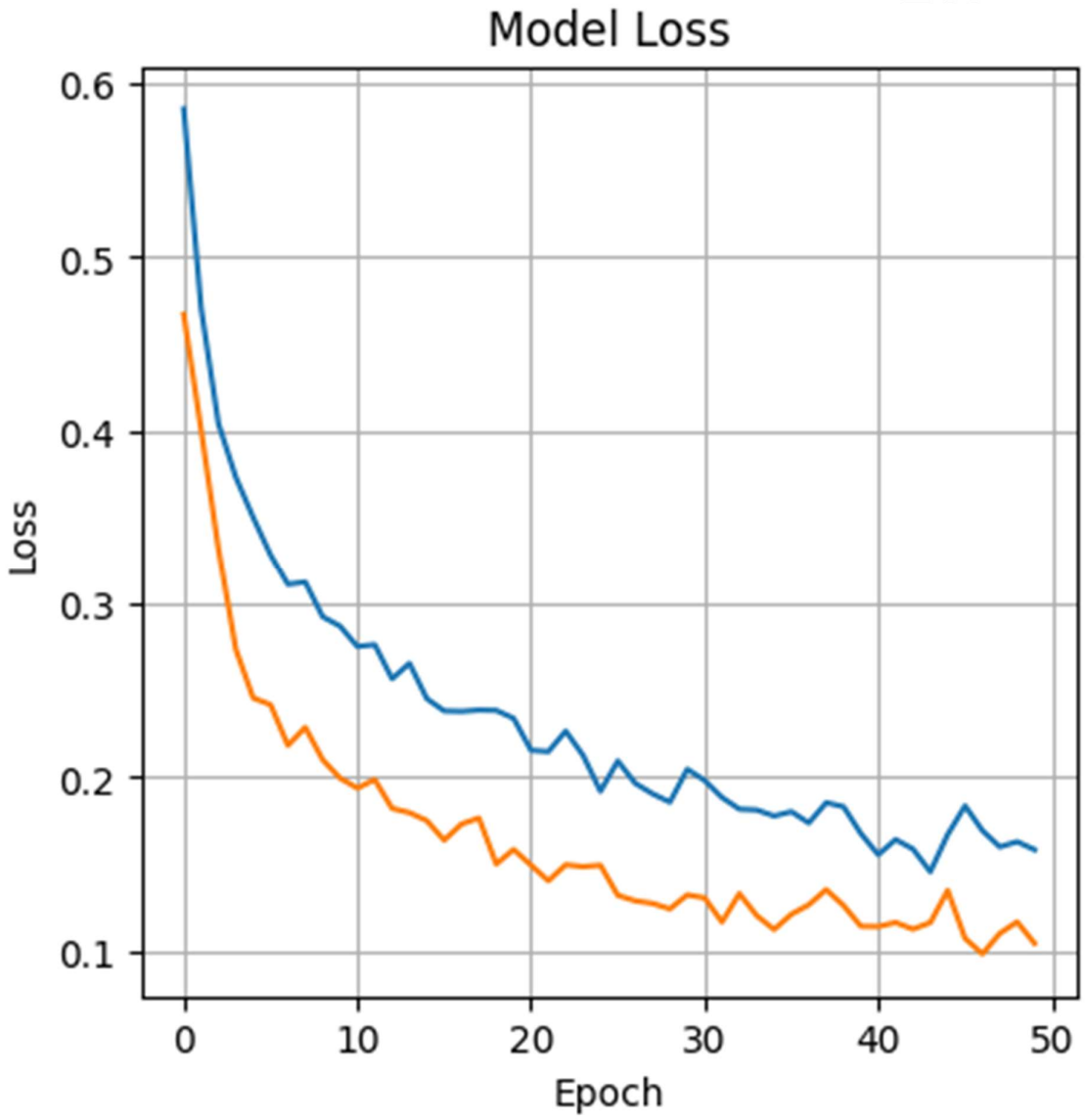

Fig 4.1: Plot training and validation loss

4.3.2. Training and Validation Accuracy Curve

The blue curve shown in the figure below represents the model's accuracy on the training data. Ideally, this value should increase over time which is increasing as we can see from given figure as the model learns to correctly segment the training examples.

The orange curve represents the model's accuracy on a separate validation dataset. This metric helps to assess how well the model generalizes to unseen data and avoids overfitting to the training data.

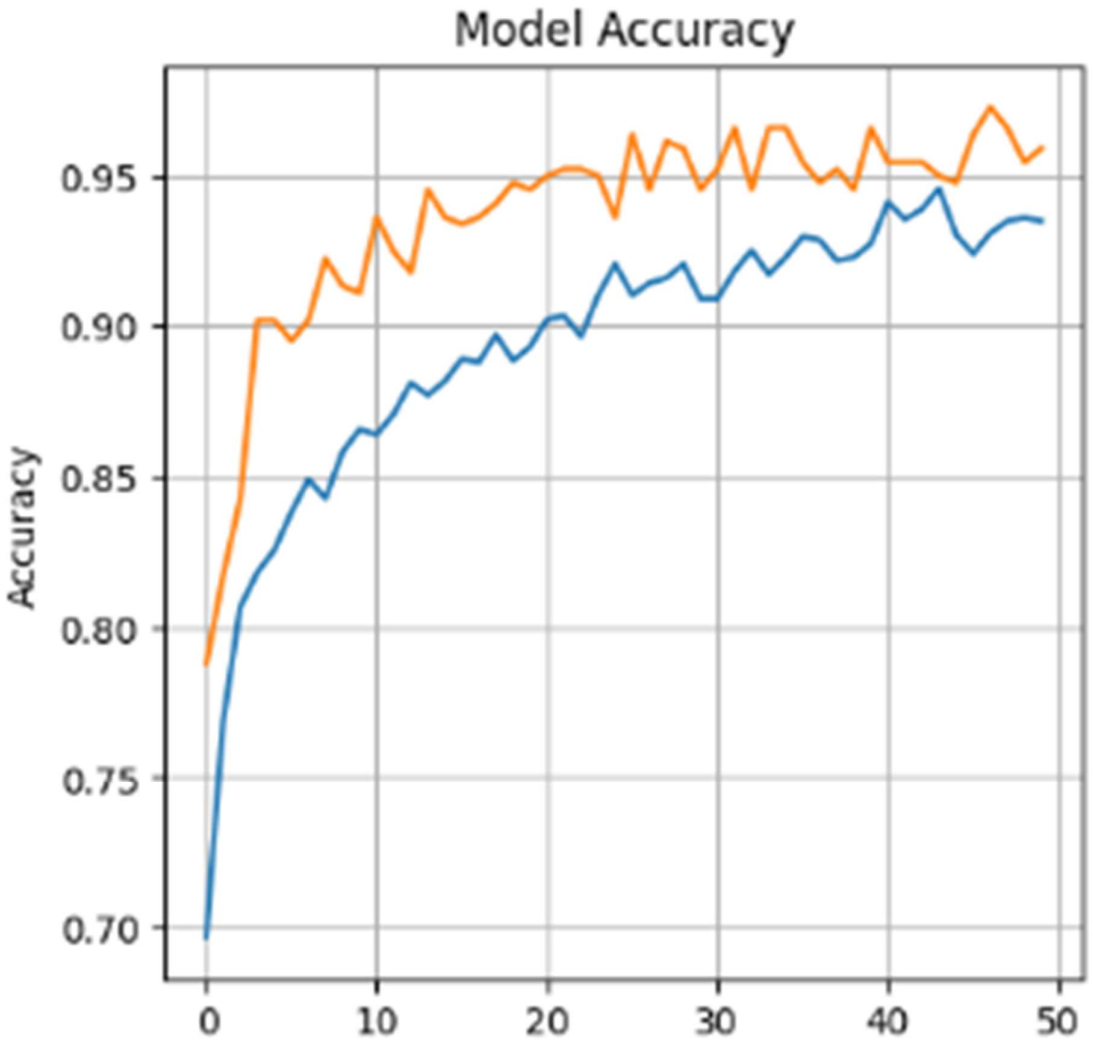

Fig 4.2: Plot training and validation accuracy

4.3.3. Training and Validation Precision Curve

The blue curve shown in the figure below represents the model's precision on the depressed class for the training data. Ideally, this value should increase over time, which is what we observe from the given figure as the model learns to correctly identify the true positive depressed cases in the training examples.

The orange curve represents the model's precision on the depressed class for a separate validation dataset. This metric helps to assess how well the model's precise depressed case identification generalizes to unseen data beyond just the training examples. Monitoring the validation precision curve allows us to avoid overfitting the model to the idiosyncrasies of the training data.

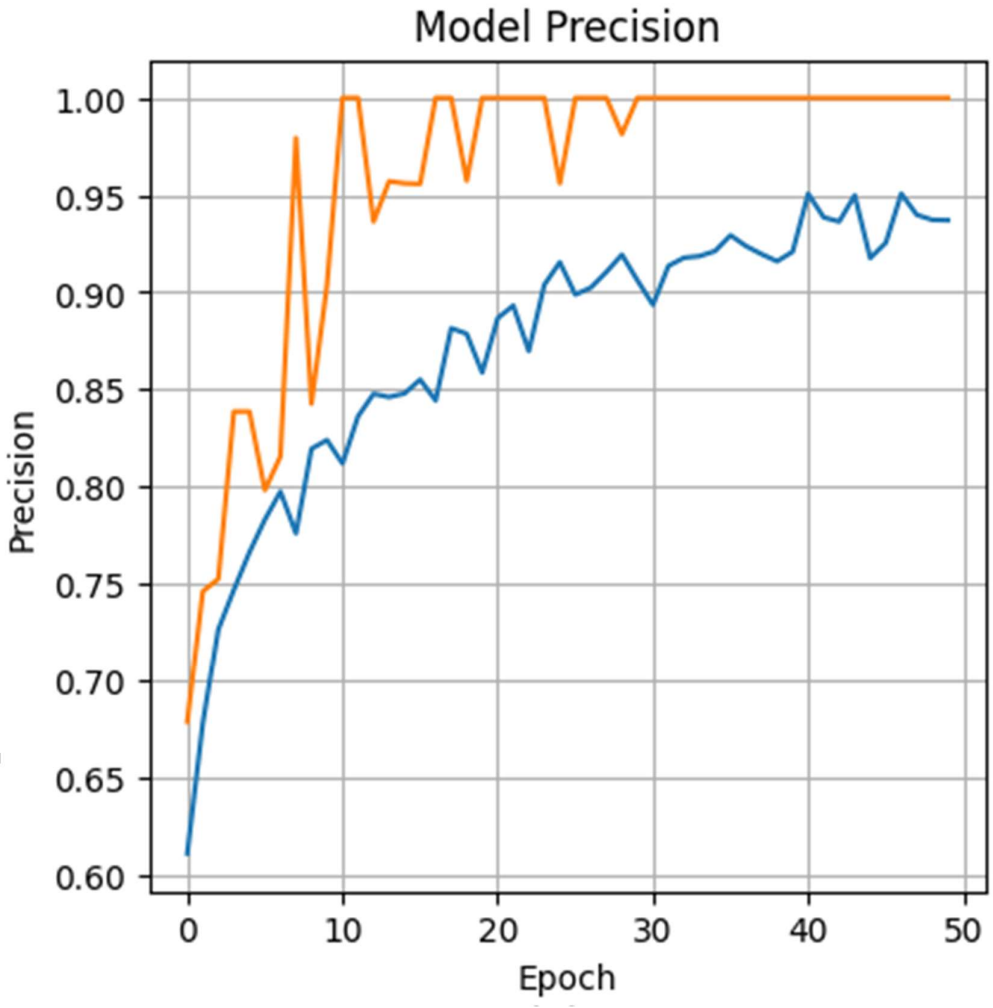

Fig 4.3: Plot training and validation precision

4.3.4. Training and Validation Recall Curve

The blue curve shown in the figure below represents the model's recall on the depressed class for the training data. Ideally, this value should increase over time, which is what we observe from the given figure as the model learns to correctly identify and retrieve a higher proportion of the true positive depressed cases present in the training examples.

The orange curve represents the model's recall on the depressed class for a separate validation dataset. This metric helps to assess how well the model's ability to successfully detect depressed cases generalizes to unseen data beyond just the training examples. Monitoring the validation recall curve allows us to avoid overfitting the model to the idiosyncrasies of the training data.

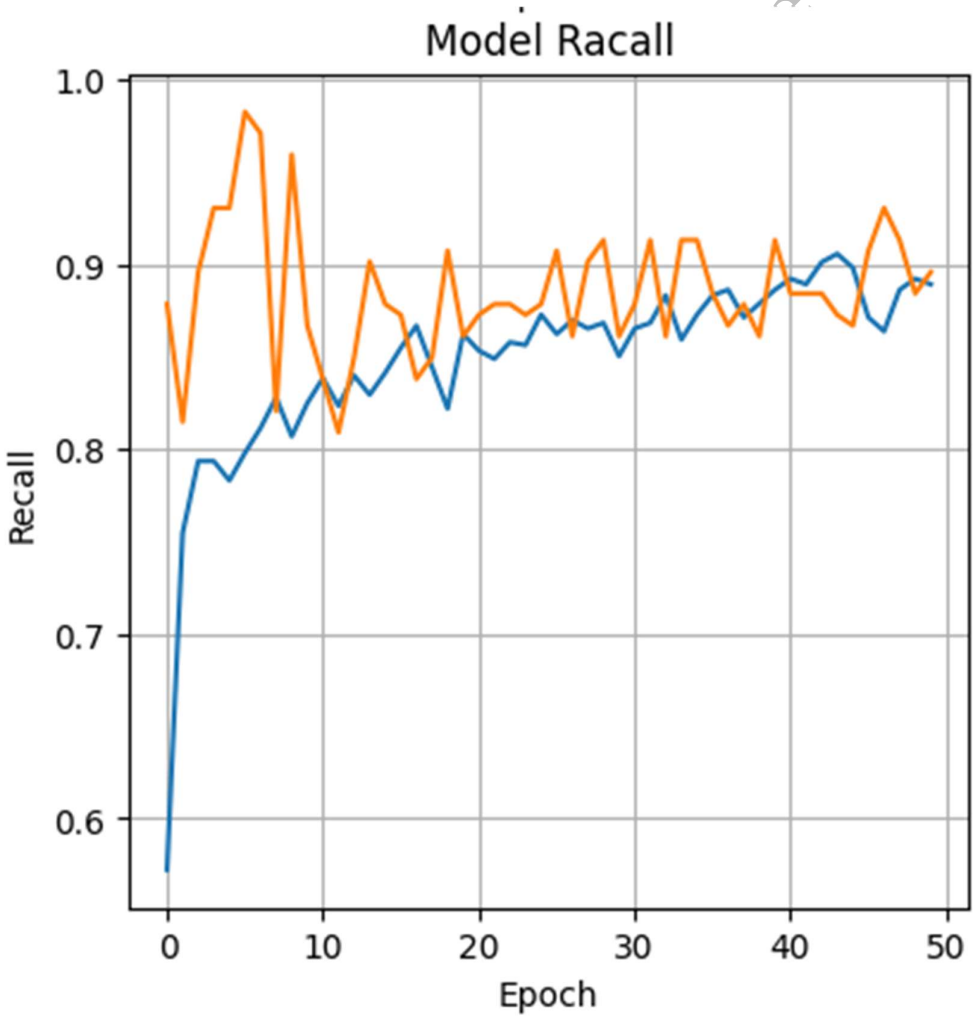

Fig 4.4: Plot training and validation recall

CHAPTER 5
ACCOMPLISHMENTS AND FUTURE DIRECTIONS

Work submitted in IMI New Delhi

5.1 Accomplishments

This thesis proposes proposed an integrated multitask learning framework that combines different data modalities to identify users potentially suffering from depression on online social networks. We leveraged the WU3D dataset [10] to facilitate research specifically targeting, identifying and detecting depression on the Sina Weibo platform.

Our experimental results demonstrated that all the initiated and adjusted analytical characteristics categorically impacted the stratification effectiveness, with textual data emerging as the most crucial modality for capturing depression signals from social network data. We evaluated the pretrained XLNet language model for embedding long text sequences and found it exhibited robust performance and computational efficiency when appropriate sequence lengths were used.

Moreover, we established the effectiveness and advantages of our multimodal multitask learning approach that jointly models heterogeneous user information sources to recognize individuals exhibiting depressive behaviour on online social networks.

In summary, we introduced a novel multimodal framework, designed informative features, and developed an effective multitask learning model, with the overarching goal of advancing the research frontier in detecting depression from social media data.

5.2 Future Directions

This work developed a multimodal framework, designed informative features, and proposed an effective multitask learning model to advance research in detecting depression from social media data. Future research will focus on further examining the attributes and behavioural sample of individuals experiencing depression to propose more positively characteristic-driven solutions for personalized depression identification in online social networks.

Work submitted in JMI New Delhi

References

[1] Zhihua Guo, et. al 2023. 'Leveraging Domain Knowledge to Improve Depression Detection on Chinese Social Media' IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, VOL. 10, NO. 4, AUGUST 2023.

[2] Biodoumoye George Bokolo, et. al. (2023), 'Deep Learning-Based Depression Detection from Social Media: Comparative Evaluation of ML and Transformer Techniques', https://doi.org/10.3390/electronics12214396.

[3] Junhee Park, et. al, (2023), "Design and Implementation of Attention Depression Detection Model Based on Multimodal Analysis", https://doi.org/10.3390/s24020348

[4] Zepeng Li, et. al.(2023), 'MHA: a multimodal hierarchical attention model for depression detection in social media', https://doi.org/10.1007/s13755-022-00197-5

[5] Lin Sze Khoo, et. al. (2023), "Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches", https://doi.org/10.3390/s24020348

[6] Yiding Wang, et. al 2022. 'Online social network individual depression detection using a multitask heterogenous modality fusion approach', Information Sciences, https://doi.org/10.1016/j.ins.2022.07.109

[7] Anshu Malhotra, et. al. (2020), 'Multimodal Deep Learning based Framework for Detecting Depression and Suicidal Behaviour by Affective Analysis of Social Media Posts', EAI Endorsed Transactions on Pervasive Health and Technology.

[8] Y. Wang, Z. Wang, C. Li, Y. Zhang, and H. Wang, "A multimodal feature fusion-based method for individual depression detection on Sina Weibo," in Proc. IEEE 39th Int. Perform. Comput. Commun. Conf. (IPCCC), Nov. 2020

[9] Chenhao Lin, Pengwei Hu, Hui Su, Shaochun Li, Jing Mei, Jie Zhou, and Henry Leung. Sensemood: Depression detection on social media. In Proceedings of the 28th ACM International Conference on Multimedia Retrieval, pages 407–411, Dublin, Ireland, Jun 2020.

[10] Dataset: https://drive.google.com/file/d/1nzURaI60wF2s4P9-G2JDowirrx0VBelE/view (WU3D dataset).

Work submitted in JMI New Delhi

[11] World Health Organization Report 2023: https://www.who.int/news-room/fact-sheets/detail/depression

[12] National Institute of Mental Health 2023: https://www.nimh.nih.gov/health/statistics/major-depression

[13] Global Burden of Disease Study 2023, Global, regional, and national burden of depressive disorders : http://ghdx.healthdata.org/gbd-results-tool

[14] http://ai.baidu.com/tech/nlp/sentiment_classify

[15] De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2021). Predicting Depression via Social Media. *Proceedings of the International AAAI Conference on Web and Social Media*, *7*(1), 128-137. https://doi.org/10.1609/icwsm.v7i1.14432

[16] Wang, X., Zhang, C., Ji, Y., Sun, L., Wu, L., Bao, Z. (2013). A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network. In: Li, J., *et al.* Trends and Applications in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science(), vol 7867. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40319-4_18

[17] Mustafa, R.U., Ashraf, N., Ahmed, F.S., Ferzund, J., Shahzad, B., & Gelbukh, A. (2020). A Multiclass Depression Detection in Social Media Based on Sentiment Analysis. https://doi.org/10.1007/978-3-030-43020-7_89

[18] Deshpande, M., & Rao, V. (2017). Depression detection using emotion artificial intelligence. *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, 858-862. DOI:10.1109/ISS1.2017.8389299

[19] Shen, G., Jia, J., Nie, L., Feng, F., Zhang, C., Hu, T., Chua, T., & Zhu, W. (2017). Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution. *International Joint Conference on Artificial Intelligence*. https://doi.org/10.24963/ijcai.2017/536

[20] Orabi, A.H., Buddhitha, P., Orabi, M.H., & Inkpen, D. (2018). Deep Learning for Depression Detection of Twitter Users. *CLPsych@NAACL-HTL*. https://doi.org/10.18653/v1/W18-0609

Work submitted in Review Delhi

[21] Sadeque, F., Xu, D., & Bethard, S. (2018). Measuring the Latency of Depression Detection in Social Media. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. https://doi.org/10.1145/3159652.3159725

[22] Shen, T., Jia, J., Shen, G., Feng, F., He, X., Luan, H., Tang, J., Tiropanis, T., Chua, T., & Hall, W. (2018). Cross-Domain Depression Detection via Harvesting Social Media. *International Joint Conference on Artificial Intelligence*. https://doi.org/10.24963/ijcai.2018/223

[23] Gui, T., Zhu, L., Zhang, Q., Peng, M., Zhou, X., Ding, K., & Chen, Z. (2019). Cooperative Multimodal Approach to Depression Detection in Twitter. *AAAI Conference on Artificial Intelligence*. https://doi.org/10.1609/aaai.v33i01.3301110

[24] Lin, C., Hu, P., Su, H., Li, S., Mei, J., Zhou, J., & Leung, H. (2020). SenseMood: Depression Detection on Social Media. *Proceedings of the 2020 International Conference on Multimedia Retrieval*. https://doi.org/10.1145/3372278.3391932

Work submitted in JMI, New Delhi