Policy Research Journal
ISSN (E): 3006-7030 ISSN (P) : 3006-7022 Volume 3, Issue 10, 2025

IMPACT OF SOCIAL MEDIA ON MENTAL HEALTH USING DEEP LEARNING

Fariha Siddiqui[1], Naeem Aslam[2], Sadia Latif[*3], Hira Saleem[4]

[1,2,4]*Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan*
[*3]*Department of Computer Science, Bahauddin Zakeriya University, Multan, Pakistan*

[1]farihasiddique50@gmail.com; [2]naeem.aslam@nfciet.edu.pk; [*3]sadialatifbzu@gmail.com;
[4]hira.saleem@nfciet.edu.pk

DOI:https://doi.org/10.5281/zenodo.17248612

Keywords
Mental Health, LSTM, CNN-Transformer, social media and AI

Article History
Received: 16 July 2025
Accepted: 16 September 2025
Published: 02 October 2025

Copyright @Author
Corresponding Author: *
Sadia Latif

Abstract
The study evaluates how deep learning can be used to forecast mental health problems using information on social media taking into account the specificity of speech, behaviors, and other multimodal data inherent to the online media platforms. The complete dataset collected at Kaggle with diverse demographic and social media usage underwent thorough preprocessing, involvement of cleaning, normalization and tokenization along with the state-of-the-art feature extraction tools such as word embeddings and transformers-based representations. The paper presented several deep learning architectures Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, as well as transformer models such as BERT, which have been trained and evaluated on various measures, such as their accuracy, precision, recall, F1-score, and AUC-ROC. Quantitative by their designs, transformer-based and hybrid CNN-Transformer models were outperforming other models consistently at an accuracy rate of over 91, with a precision and recall of over 90 and an F1-score of over 0.90, which is indicative of a stronger capacity to detect contextual and lingo-based cues concerning mental conditions. In comparison, CNN and LSTM models acted more mediocrely, boasting accuracy of 85-87 percent, lower precision and recall figures of 83-87, which underline the importance of modelling context of unstructured social media text. The results also show that the deep learning models will soon provide a powerful 10-15 percent improvement on the predictive accuracy compared to the traditional machine learning baselines, which were Support Vector Machines and Logistic Regression, with an average prediction accuracy of 75-82 percent. Confusion matrices showed that more complex models significantly reduce false positives and false negatives that result in more credible models that are critical in clinical and social health applications. These findings indicate that deep learning-based automated mental health monitoring and intervention systems can be practiced in real-time, particularly multimodal data and longitudinal behavior integration to boost the power of prediction. The thesis offers a fresh view of scalable and decipherable AI models to be used in different contexts of the social media that will become the foundations of the future research of privacy-sensitive algorithms and as an extension of the analyses to the

fusion of the audio-visual modality as a continuation of the later improvements of the mental health surveillance.

INTRODUCTION

Social media has been an element of the modern daily life and has redefined how individuals communicate, relate to each other and perceive information. By 2024, 4.6 billion people across the globe use social media sites, and these consumers spend more than six years on social media, which can be considered an average lifetime cumulative time. The social media dominant today is of that kind, and the media are: Tik Tok; Instagram; Facebook; twitter and snapchat on which individuals can create - make or share - virtual communities even talk in them. Social media has come to establish a new social norms and behavioral pattern and they can provide young people with distinctive opportunities to develop cognitively but it also carries severe consequences as far as mental health is concerned [1]. Its accessibility as well as its impact on the use of social media is a basis on which it can be used in mental health. Teenagers (13-17) are the most vulnerable because 91 percent of them reported checking in more than three hours daily on average on social media and even the young adults (18-29) outdid it with more time on the five leading social media. This proliferation has resulted in a growing worry regarding the psychological effect of a prolonged exposure to the social media scene. Anxiety, depression, loneliness, and stress are the mental health issues that are becoming more and more linked with the usage of digital media, particularly when these use trends become problematic or excessive. Recent research indicates that the adverse outcomes of overuse of social media have negative correlations with the mental health. As an example, individuals spending more than three hours a day on their social media are approximately twice as prone to be vulnerable to mental problems compared to the light users. Spending over three hours a day social media among adolescents is perceived to have led to a 66 percent increase in incidences of depression, development of signs of anxiety and the feeling of being isolated. Such figures may be compared to the clinical portion of the evidence that depressed and suicidal adolescents have a higher chance to support the problematic SMU behaviors through emotional over-attachment and distress about failing to navigate the cyber world [2].

It is too high-end dumping of quality and rubbish in social web. At the positive side, social media has established social connections, information sources and mental health awareness and peer support. However, it may also expose them to harmful stimuli (such as cyber bullying and unrealistic comparison) and even challenge them with badly new content that continues to cause anxious reactions and other depressive symptoms. The filtered aspect of the social media profiles of people can also promote peer comparison pressures and inadequacy feelings, the other account feels that the lives of other users are more satisfying or happy than their own. This impact leads to rumination and the further worsening of mood disorders. Interestingly, the sensation of loneliness and social isolation that was improved thanks to social media. Despite the supposed connectivity, many users have been said to experience loneliness after accessing the social media sites- another irony that comes with the superficial communication replacing meaningful communication. Despite being convenient, digital communication may inhibit face-to-face social skills and empathy, and emotional control, particularly in developmental stages, including adolescence, when the identity and interpersonal skills are of significant concern. Besides, the variation in the effects of social media on the mental health among the genders and ages have been documented. The probability that young women and girls experience adverse emotional effects of social media use is statistically higher than that of boys with some researchers showing a 20% greater likelihood of adversely affected effects. The older generation is less prone to it, both in how many people use social media and how many are affected psychologically because only 15 percent mention stress related to social media[3].

Value of Analysis of Mental Health impact in the Digital Era

Unparalleled infiltration of digital communication in the everyday life of human being is what dictates the demand to study the impact of social media on mental health in the digital era. These comprise mental illness like anxiety and depression, which are the major cause of disabilities in the world and that consequent it leaves behind a mark in the name of an individual, family and health system. As more and more people are exposed to social media, and more and more people use social media, people need to understand how social media exposures are sinking into the skin of an individual, and this impacts the psychological well-being in such a way as to write evidence-based interventions, policies and recommendations that are realistically implementable. The digital age and information society has produced new blood stains of mental health that can be partially justified and explained by classical epidemiological models. Environmental and behavioral risk factors that mediate the case of mental health issues are the social media and screen time. Of central importance to the public health, is the finding that such exposures and exposure pathways, to the extent to which digital content and interactions are affecting the brain functions, mood regulation, social cognition, have not only been established but elaborated upon. The social media is also characterized by medium fine-grained perception because it is a location where positive and negative information may co-exist and therefore risks of protection and buffer is needed. On one hand, it might be the setting where ant stigma campaigns, peer support groups and mental health resources may be developed, and on the other hand, cyberbullying, misinformation and social exclusion might be encouraged. The enactment of these two possibilities in sub populace of demography and clinics should be decided.

The growing symptomatology of mental health on the Web (the latter is linked to the use of social media) require novel approaches to its study that will consider more sophisticated forms of data and processes. Since deep learning and other sophisticated models are being developed at an extremely fast rate to complete a calculation, the red flags or pattern of behavior that is pursued in the social media can be traced in real time in order to give a recommendation that can be presented to the person to act. Another critical area of mental health in the digital technological environment is the behavior change and the social issues. Regarding the example of the Generation Z, more than 60 percent say they believe that social media is more detrimental than beneficial, in most cases, this means that the concept of digital well-being must be addressed as a direct challenge. The COVID-19 explained the acceleration of the digital communication, but also revealed the inequality in mental health when taking into consideration the questions of social isolation and Internet communications [5]. Lastly, given that those are the teenage years and early adulthood when the habit of social media use is the most active, and the mental disorders are more likely to set in, the digital tools that prioritize prevention and early intervention can become the game changer. It is a partnership to deliver evidence-based testable digital models of mental health between clinical psychology, psychiatry and the computational sciences and the public health. With the development of online social networks, unprecedented amounts of unstructured data are being accumulated that captures human opinion, sentiment, behavior. This digital life is a wonderful chance to analyze mental disorders, depression, anxiety and suicidal behavior, as well as other psychological problems through the mediated user by means of his social media. This multidimensional turbid and noisy information is also a tricky undertaking to derive value. The sub-discipline of artificial intelligence or rather, neural network structures coincidentally turn out to be the state-of-the-art solution in that it was demonstrated to be capable of identifying complex patterns and semantic concepts in diverse data representations with an astronomical success. One of the motivations to use deep learning to investigate how the social media influence mental health is the increasing volume and variety of data generated on these media. The data of social media is not just text but images, videos, emojis and metadata (e.g., timestamp and user interactions) that results in multimodal information that cannot be easily learned in holistic sense by the traditional machine learning methods. Convolutional Neural networks

(CNNs) and Recurrent Neural networks (RNNs) are deep learning paradigms that are highly effective at this heterogeneous data to capture the spatial and temporal relationships. The last performance of the Transformer-based models are also highly competitive baselines in this task. Transformer-based models, including BERT and Roberta, are indicatively in the state of the art to perform natural language processing tasks that are critical in understanding the sentiments, intentions, and mental states of users based on the text in the social media[6].

The second reason is that deep learning is more efficient in automatic acquisition and learning of features on raw data and does not need manual feature engineering, which is both time-consuming and potentially biased. Social media cues of mental health are normally very thin and contextualized, positioned within the language style, posting patterns, and visual presentation indicators that are hard to identify according to the traditional algorithms. Layered representations, extracted by deep learning algorithms, abstract the data bit by bit, and thus can recognize subtle emotional expressions, language usage variations, and even psychological latent features such as hopelessness or suicidal thoughts. This variability of mental health state also requires models having the capability to handle the variability of time. Deep learning models with recurrent structure such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) are most effectively structured to handle sequential data, regardless of whether it is dynamic or not. It is needed to trace mental health outcomes in social media, where changes in language or patterns of interaction can lead to clinical symptoms, therefore, detecting them early and perhaps intervening[7].

In addition to the increased accuracy and learning of features, the deep learning offers scalability and cross-language and cross-social media platform flexibility. There are also problems of data sparsity and heterogeneity with large scale mental health research at risk where pretrained language models can be fine-tuned on relatively small labeled datasets and understand broader populations and domains. Such flexibility is necessary in the development of real time, automated systems to investigate the continuous flow of social media and provide real time information. The deep learning model of social media analysis, in a clinical context, offers mental health professionals a free supplemental method of mental health, which is low-cost and non-invasive. Automated detection models based on deep learning can be used to identify people with risky behaviors or deteriorating mental health to be further assessed or treated, which can also help to reduce the time taken to seek help and the burden on the healthcare system. In addition to this, individualized models of mental health care may be created using social media data and individual-specific risk factors may be detected along with symptom patterns [8].

Moreover, deep learning may take part in the incorporation of new forms of digital phenotyping that will combine smartphone and wearables information with social media streams to provide a holistic perspective on behavioral health. This multimodal synthesis is sure to allow emergent ideas of the etiology, development and response of the mental health where evidence had been difficult to establish by the traditional clinical analysis. A linguistic analysis of social media and a physiological measurement would be one of such strategies that would predict a depressive episode or an anxiety attack better. Despite its potential, deep learning in this potentially sensitive domain is likely to have some drawbacks such as model interpretability, privacy and ethical concerns. The use of black-box deep learning models makes such forecasts of risks hard to explain, which is directly related to a trust issue presented by users and clinicians. It already implements explainable AI methods to such models as a research vehicle and simplifies them to the clinician. Moreover, in the case of Personal media contents Shaping ethical use of AI in Mental Health, the user and informed alter data are also major when applied to the case of AI usage in mental health.

Literature Review

The constantly growing body of literature that investigates the effects of social media on mental health has continued to mark the nuanced and curious connections between the intake of digital media and the psychological disposition of being. Most of these researches indicate that excessive or long-term use of social media is closely linked to

depression, anxiety, loneliness and symptoms of stress, especially in the young. As an example, some empirical studies based on comparisons among subgroups show that teenagers and young adults who more frequently use application(s) of a social media such as Instagram, Tik Tok or Facebook more deeply experience mood symptoms aggravated psychologically when compared to teenagers and young adults who use such forms less often (light users). Social comparison (participation in some form of negative comparison process of own life with the images that are being photo-edited by familiar other people) was determined to be partially associated with this relationship whereby self-esteem significantly lowered and the level of distressing thoughts were significantly elevated. Also, it is associated with augmentations in perceptions of social isolation and recurring online social rejection or exclusion. The relationship has been characterized by these trends in a number of cross-sectional or longitudinal studies, and some of them proposed that the association can be causal though concrete causal relationships have been hard to determine due to the limitations of study designs [9]. Psychological effects may also be carried about the quantity and the kind of social media usage. Research indicates that the duration of spending on social media platforms does not pose, per se, a threat to adverse psychological effects, but the volume of visits and the number of supplementary SMs, in fact, contribute to the risk. An example is young adults who visit a multitude of platforms daily have much higher rates of depressive and anxiety symptoms than those who visit fewer or less platforms many times. This might be excessive exposure and users can be addicted to use habits and emotionally reliant on digital approval. Further, social displacement theory suggested that meaningful face-to-face communication might be undermined through the use of social media and this would further result in individuals feeling lonely and the occurrence of the adverse impact on the emotional condition. This has been reported in the literature during and after COVID-19 outbreak, where the use of social media and deteriorating mental health in adolescents and young adults have been positively linked with the physical distancing-impaired isolation. More to the point, this is the time when prolific research on the dual role played by social media as the source of both support and stressor compounding on two-sidedness of the phenomenon of digital interaction was conducted [10].

Other researchers have discussed the more fine-grained processes that describe some of the effects of social media on mental health, which are the extent to which the users are invested in social media, whether they are victims of cyberbullying, whether the social media is racist or whether the user is interested in how others feel about them. These findings confirm that there is much contingency in the psychological impact of social media and on the susceptibility of individuals and on various social environments. The more stress and even depression teenagers who use social media may feel to be friendlies or good about themselves. Further, when the marginalized groups become victims of the online-based discrimination, they are likely to experience the impact of such content in their mental wellbeing and this gives grounds to a more intersectional ground in research and intervention. The police force is also active in Patna and the areas around it, but notwithstanding the continuous desolation of the population, no serious impact studies or monitoring is performed. Long after the above somehow anticipated injury to everyone now studies as with all other things over general population than grew up on severely under MMH developed some more negativity such as social media exposure homotropy (nightmares) and mood legation besides exacerbating psychotic conditions. These observations combine to describe a complicated and two-way relationship involving social networking sites and mental health that will demand multifactorial and multicomponent models of study and prevention strategies in the future [11].

Common Mental Health Issues associated with social media

Social media usage has experienced an astonishing rise in the last ten years to become indispensable part of the lives of millions of users around the globe, especially among teens and young adults. Even though social media is viewed as a platform of interconnection, self-expression, and information sharing, an emerging body of literature has reported an alarming correlation between the use of social media and diverse negative psychological health

outcomes. The domains of psychologic condition, where this application is most commonly connected, are anxiety, depression and loneliness, with the existing evidence portraying that heavy social media consumers experience more symptoms in each. To illustrate the latter, recent statistics indicate that over half of adolescents are extremely nervous or depressed following usage of social media and individuals who consume more than 3 hours a day are virtually certain to experience mental health issues on a significantly more frequent basis than those who do not consume social media [12].The causes of social media anxiety include the sense of fear of missing out (FOMO), cyberbullying, and social comparison since people feel that their lives are worse than those they observe online. This greater stress of the emotional state is most likely to augment both underlying weaknesses, contributing to chronic anxiety disorders and mood disorders.

Depression is one of the most common socially networked mental health issues that have been researched extensively about mental health concerns. The observation that the use of social media in excess and problematic forms is associated with the exacerbation of symptoms of depression in numerous diverse groups is reported in many longitudinal and cross-sectional studies. Depressed young individuals are more likely to be emotionally over-involved with social media, and more upset or discomforted at being denied it. It is also called social media addiction and such a behavior can potentially contribute to hopelessness and isolation only. Also contributing to the depressive thoughts are the frequent exposures to negative content, online bullying, and unrealistic standards of beauty which is more likely to impact young women who have a probability of being negatively affected 20 percent compared to male counterparts. These results correlate with neurobiology, which suggests that excessive social media use can interrupt mood modulatory reward circuitry, to increase the amount of negative cognition and resultant depression. [13]. This fact supports the assumptions that the connection between them can be considered as a multi-dimensional phenomenon comprising of the relationships between the online space behavioral use and neuropsychological preparedness to depression.

One of, or potentially the greatest, mental health challenges that accompany the use of social media, despite its counter-intuitive nature, is loneliness. Studies have indicated that, through the social media, nearly 1/4 of the adults are lonely / lonely. The thing is, it is not that the two do not communicate, but online interaction + the strong F2F relationship possesses a quality of differentiation that affects the emotional well-being of a person disproportionately. Social media is necessarily likely to develop and develop superficial relationships and superficiality of human contact and empathy, particularly in teens where identity and social growth are predominant. 31, also, social rejection or exclusion as it applies to the context of social media is reported to be a contributor to acute loneliness and subsequent mental health decline. The feeling of abject inferiority and social exclusion can also be a result of the constant exposure to an ideal social world on these platforms. Such tendencies are supported by the polls: According to them, adolescents who frequently use social media are more likely to experience a feeling of social isolation by a factor of two. The fact that the outcomes of anxiety, depression, and loneliness are long-term indicators of social media use implies the necessity to create the intervention plan that would facilitate the promotion of mental health during the digital age [14].

Computer algorithms of mental health analysis of social media content

With social media becoming an increasingly popular means of communication, the extrapolation of a broad range of computational methods has been discovered to be relevant to the analysis of user-contributed information to investigate mental health and its condition. Earlier methods were based on more traditional ML and tended to use manual feature extraction of social media content, such as word counts, semantic analysis, and keyword detection. Earlier methods like SVMs or logistic regression or random forests have already been relatively effective in identifying psychiatric disorders like depression and anxiety. However, they were limited by the quality and the scope of the hand-crafted features which were often incapable of capturing the contextual and the temporal specifics

of the social media language. Subsequent advances of NLP involved more sophisticated techniques that incorporated sentiment analysis and topic modeling in order to increase the granularity and accuracy of mental health predictions when textual data is used[15]. Since the advent of deep learning, end-to-end learning on raw data with minimal manual feature engineering is now possible, and computational analysis of mental health has been transformed. The sequence and space patterns of social media content have often been predicted by the use of recurrent neural networks (RNNs) and convolutional neural network (CNNs) as deep neural networks. Recently, models constructed as transformers have demonstrated state of the art performance on extracting fine semantic and emotional content about posts, with the likes of BERT (Bidirectional Encoder Representations from Transformers) and variants thereof. These models specifically excel in identifying the hidden elements of language, which might be a sign of mental illness, such as expressions of despair, aggression, and isolation, that play key roles in the diagnosis of mental illnesses such as depression and suicidal ideation. Multimodal deep learning systems (combining text and images, videos and user interaction metadata) are also included in other studies to provide a more comprehensive view of mental health indications and predict with more accuracy. Moreover, time-varying models based on the Long-Short-Term Memory network (LSTM) networks are used to monitor how mental health indicators change over time, and it opens the opportunity to predict mental health emergencies using the time-varying social media behavior.

Most current models are based on sets of social media posts on Twitter, Reddit, Facebook, and mental health forums that are specially designed to train and test their models[16]. The researchers have incorporated the demographic and behavioral features (frequency of posting, time of posting, and social network features) with the textual features to maximize performance. The model has been expanded to provide the most predictive qualities using features selection methods such as Cumulative Weight-based Iterative Neighborhood Component Analysis (CWINCA) that enhances the model interpretability and power. Accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC) are the most commonly used metrics of evaluation, and recent deep learning models are always better than the standard classifiers. Pretrained and finetuned language models trained on any given domain data have also enhanced the generalizability and reduced its dependence on vast amounts of labelled training data. It should be mentioned that such calculation tools are currently being regarded as an auxiliary tool to facilitate clinical screening and follow-up on population health and can potentially identify real-time mental health trends and assist in intervention plans when the users are involved in social media activities.

Introduction to deep learning and its applications in mental health detection

Deep learning is a branch of machine learning that uses the structure and the functionality of the human brain to replicate complex patterns into big data sets using artificial neural networks. Deep learning employs a sequence of layers of interconnected nodes, also referred to as neurons, and they learn feature hierarchies of data through end-to-end training, in comparison to the traditional machine learning algorithms, which rely on manual feature extractions. Each layer transforms the input data into abstract features and enables the model to learn complicated relationships, such as complicated linguistic information and time dependencies in text or spatial hierarchies in images. Deep learning is particularly optimal in social media social media analysis due to the ability to handle large quantities of unstructured data with minimal features designed by humans[17]. Convolutional Neural Networks (CNNs) to learn image-based data, and Recurrent Neural Networks (RNN) and their derivatives including Long Short-Term Memory (LSTM) networks to learn sequence data have radically improved natural language processing (NLP) tasks and computer vision problems. Further on, transformer-based architectures such as BERT and GPT are now the state-of-the-art, demonstrating especially good context and semantic capture in language even in extended passages. This has made it feasible to extract subtle emotional, cognitive and behavioral cues in the social media postings, often

informal and fragmented. Models that have been trained on large corpora can be used to detect depression, anxiety, suicidal thoughts and other such mental health problems by identifying patterns in word/syntax/imagery/posting rate signs of psychological distress [18].

Deep learning has numerous advantages in tracking mental health through social media. The large group real-time tracking of the user becomes possible and early warnings and possible personal mental health intervention can be performed. Multimodal data (texts, images, videos, metadata) and modeling of temporal pattern could be pooled and modeled through deep learning systems to generate dynamic behavioral patterns that not previously observed by less advanced assessment instruments. The deep learning tools can also be used to reduce self-reporting bias and prevent traditional clinical access (such as [19]) across underprivileged groups or stigmatized populations. Despite interpretability issues and ethical concerns such as privacy, the accuracy, transparency, and clinical relevance of deep learning models are currently under further improvement in research and seem promising aspects of a digital mental health ecosystem.

Research gaps in combining social media analysis, mental health, and deep learning

Despite the significant progress achieved in applying deep learning approaches to the study of social media data, there are still numerous gaps in research areas that do not allow fully utilizing the potential of the new field. One of the typical gaps includes limitation of data quality and diversity. The existing research is mainly reliant on publicly accessible text-based data found on websites like Twitter and Reddit which may not be representative of the population at large or capture multimodal information, including pictures, videos and audio, particularly on websites like Instagram and TikTok where visual and audio data is the norm. In addition, the data sets are often poorly annotated, with self-reports or other heuristic components instead of diagnostic data. This causes noisy or missing labels, which will cause problems with model generalization and accuracy so that models that are trained on such noisy or missing labels will not be likely to accurately detect mental health problems in a heterogeneous user population and cultures [20]. The second significant gap is the verifiability and illumination of the deep learning models being consumed to identify mental health on social media. Most of the state-of-the-art models are Blackbox transformer and deep learning architectures - do not assert strong explainability of how/why model makes decision or what feature space is the most important [21]. This practice of secrecy gives the ethical problem in both clinical and community health an element of professionalism; confidence and accountability are essential elements of practice. Even though certain methods of XAI have been developed to address these issues, e.g., LIME and SHAP, they are not built into end-to-end mental health detection systems and require additional work. Additionally, there has been a focus on single modality or type of features (like text sentiments and user metadata) and the complementary nature of multimodal data and feature extractor combination has never been fully leveraged to enhance performance robustness of performance of the model, nor its interpretability [22].

And one of the inexplicable key questions in the discussion on deep learning in social media mental health analysis remains also in all who think about morality, confidentiality and justice. This is, of course, as sensitive as can possibly be, health and privacy-wise, and in the vast majority of cases it seems that the researchers are working with data extracted out of publicly available material with little or no worked-up systems of informed consent. This brings about legal and moral issues of user control and abuse of information. Moreover, the evidence of demographic biases of social media data and model outcomes are comparatively recent and have a risk of upholding disparities in accessing and achieving mental health among racial, socioeconomic, and gender groups. An improved future work should be more focused on how to implement fair models that are aware of diversity and inclusion into practice and how to implement sound safeguarded algorithms that are also respectful of privacy such as federated learning and differential privacy. Such gaps will eventually be crucial to the prospective ethical, effective and clinically meaningful deep learning application to monitor and enhance mental health in the digital era [23].

Methodology

Data collection

The information in the present research is called social media and Mental Health, it has been downloaded at Kaggle, the objective of this data is to determine the relation between mental health and social media use among individuals. Initially gathered within the framework of one project in the data science and machine learning domain, this dataset consists of self-reported responses of 481 individuals with 21 different variables that encompass demographics, social media usage, and symptom of mental illness. Though there is no information about the procedure of data collection and the methods and period of sampling, the datasets contain some useful information, like age, gender, occupation of the civil status and, of course, the most important: which social media platforms were the respondents questioned about in the most specific way possible how much time people spend on social media in an average day. This data gives a promising perspective of the duration of time spent on social media possibly being linked to various mental health effects like anxiety, depression, restlessness and shifts in self-esteem and self-esteem and thereby ideal in making deep learning analyses that will allow us to unravel predictive based behavioral patterns.

One can observe the type of the social media We have reviewed that defines the nature of the services parties and individuals are familiar with when communicating electronically. They were to name the most frequent platforms in which replies covered significant sites like Facebook, Instagram, twitter, tik Tok, snapchat, LinkedIn and WhatsApp. Such variety allows drawing a parallel with platforms that are oriented to various user profiles, patterns of interaction, and content types (for example, text-based networks like Twitter, or image and video-based ones like Instagram a Tik Tok). The combination of these usage patterns of platforms and mental health indicators presents the data as valuable in examining how some platforms are more positively linked to symptoms of poor mental health or behavioral tendencies, that is, social validation seeking or distractibility. Moreover, the signals include such average daily use times in social media and numerous behaviors like using social software without purpose and feeling uncomfortable without this device-the data can also be used to locate the complementary views of severity and mental influence among other platforms.

Dataset link: Social Media and Mental Health

Ethics is the most significant issue of dealing with data relating to mental health and social media behavior due to privacy sensitivity and the potential stigmatization. Though Kaggle data is anonymized, and does not offer personal identification, since the mental health scores and social media consumption are self-reported, there must be a delicate way of maintaining confidentiality and interpreting it correctly. The data were presumably collected in a voluntary manner, there is no particular reference concerning the procedure of informed consent and whether these procedures have been reviewed and approved by an institutional review board (IRB), which can be considered a limitation in terms of ethics when conducting mental health research. Proper data protection should be implemented to ensure the anonymity of the subjects, including safekeeping and access limitation (by the researchers accessing that data). In addition, one should also introduce the forms of analysis and reporting that do not reinforce stereotypes and biases against social media users with mental health problems. The recognition of these ethical principles as the dataset is used can be of great benefit that will positively influence the understanding of the adverse mental health impacts of social media and, possibly, reduce them.

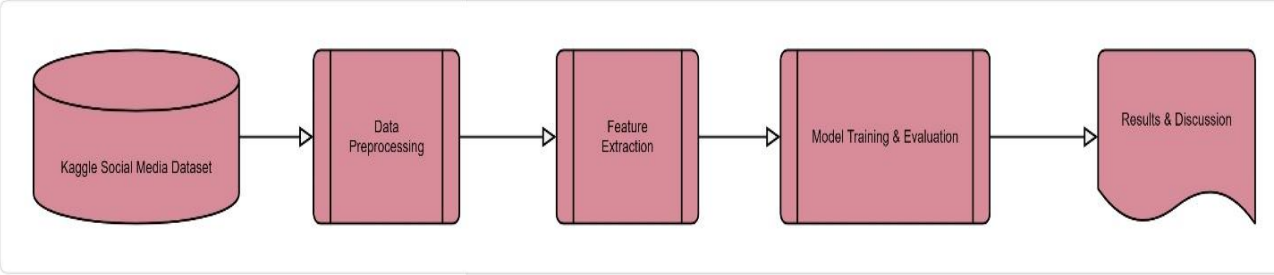


Figure 1: Methodology flow chart

Data preprocessing

Data preprocessing, especially with deep learning models, where input data quality and consistency are sensitive, is a very important step in preparing social media data to analyze mental health data. Social media text is generally informal, loud, and unstructured containing slang, abbreviations, typing errors, emojis, hyperlinks, inconsistent capitalization, and inconsistent capitalization that can severely disrupt the model performance when not processed. The initial step of text cleaning is carried out by removing extraneous data e.g. HTML tags, URLs, special characters and unneeded whitespace. After cleaning, the next step is the tokenization of the text into meaningful units, word or subworlds. To sparsity the data and have the models generalize, techniques of normalization such as use of lower-case text, expansion of contraction and correction of common misspellings are employed. Words with weak semantic meaning common words (and, the, is, etc.), are often omitted so that one can focus on meaningful information. Within sentiment and emotion detection tasks, emojis and emoticons can be converted to text messages since they are rich in affective data that can be applied to mental health.

IEER
Institute for Excellence in Education & Research

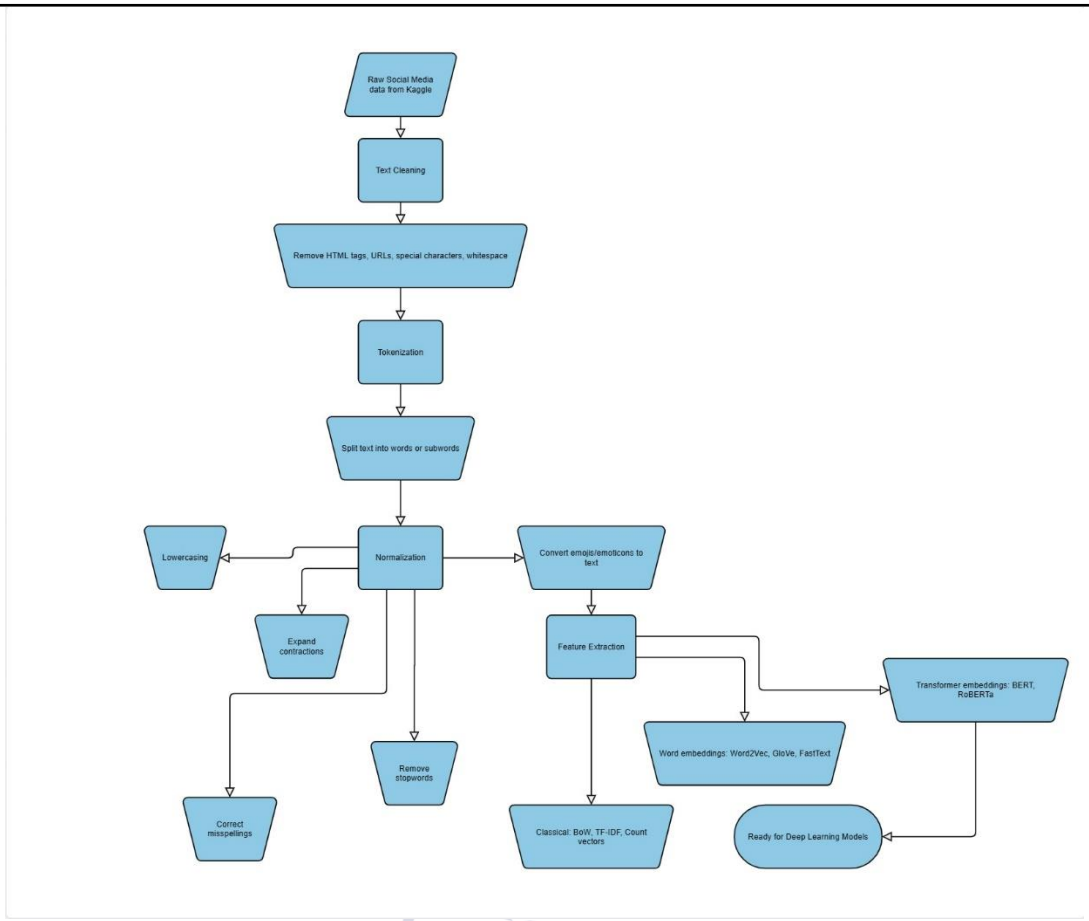


Figure 2: data preprocessing steps

Once the text is cleaned and normalized, it is converted to machine-readable formats through feature extraction and can be utilized in deep learning models. The classical algorithms are the ones of vectorization, i.e. Bag of Words (BoW), term frequency-inverse document frequency (TF-IDF), and count vectors, that numerically measure how often and how important words are in documents. However, they lack situational nuances and sequence of words that matter in the interpretation of complex emotional and cognitive cues in social media posts. The resultant use of word embeddings (dense, continuous vectors) to model semantic relationships by such an algorithm is used today, such as Word2Vec, GloVe, and FastText. In the most recent cases, transformer-based pretrained language models such as BERT and RoBERTa have broken the feature extraction pipeline by providing context-sensitive representations of words that dynamically vary in the representation of a word depending on context. These embeddings add a lot to the potential of detecting hidden linguistic patterns that are associated with mental health issues. In addition, it is possible to use multimodal feature extraction in datasets that have metadata, videos, or images. Longitudinal mental health assessment can also be provided with temporal data including posting rate, time of day and interactional behavior encoded to provide behavioral context. Overall, data preprocessing helps to ensure that text in social media is preprocessed to high-quality and informative inputs that can be used to detect mental health with deep learning models effectively and positively.

Deep learning models used

In this research, several deep learning architectures were used to process the data of social media to identify mental health issues because they were shown to be effective in terms of solving the

complexity and variability of the content created by people. The most commonly used Convolutional Neural Networks (CNNs) are utilized due to the capability to extract hierarchical features on both text and image data. By using various layers of convolution and pooling operations, CNNs are able to extract relevant patterns in short text snippets that often occur in social media posts and essentially identify local semantic structures, including the key phrases, emotional phrases, and visual representations in images. Recurrent Neural Networks (RNNs) and their technologies, such as Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU), are particularly helpful to deal with sequential data, such as sentences and posts over time. The models are good at capturing dependencies longer text sequences and are therefore indispensable to get a sense of context, linguistic nuances, and temporal variation in mood and sentiment. RNN-based models are especially applicable to the monitoring of changing mental health conditions based on the successive user posts on social networks.

More recently, transformer-based architectures, such as BERT (Bidirectional Encoder Representations of Transformer) and RoBERTa and GPT have become more popular as they excel at modelling context and meaning on full text sequences. Transformers rely on self-attention that may trade-off relevance among various sections of the input and achieve the ability to capture subtle semantic relationships and long-range dependencies far better than standard RNNs. This would play a critical role in identifying mental health in social media data, in which indicators of distress or effective coping may be spread across a user's history of posts as compared to being found in a single message. In addition to this, domain specific pretrained models of language with finetuning like fine-tuned transformer using relatively small domain-specific datasets (mental-health-anonymized texts in social media) could potentially yield higher accuracy and less time to train. Since social media is a multimodal platform, hybrid models, which use CNN to extract image features, and transformers to process text, offer a more thorough understanding of such vision slang interaction that overall has higher end-to-end detection performance. Data type then dictates the type of model used: CNNs to model spatial and visual data, RNNs to model temporal sequences, or transformers to model deep contextual language knowledge, and lets the model flexibly and richly model a wide range of social media data modalities.

Model training and evaluation

The initial step in deep learning model training with social media-based mental health detection would be preparing clean normalized datasets, and then they are separated into training, validation and test sets to ensure unbiased evaluation and generalization. The split ratio, 80-10-10 or 70-15-15, is the most common and the training set is used to optimize the model weights by using both the backpropagation and gradient descent methods. The auto adjustment of hyperparameters in training (such as the batch size, learning rate, and the number of epochs and the depth of the network) is aimed to prevent overfitting or underfitting. To improve the robustness of a model, to ensure that a model is not memorizing training data noise, regularization techniques, including drop out, early stopping and weight decay, are applied. Moreover, when the quantity of labelled data is limited, transfer learning with pretrained transformer models (e.g., BERT) fine-tuned on data in the mental health domain is also applied to obtain high performance. The information added using data augmentation techniques like synonym replacement, back translation or minor perturbations, improves the generalization of the model by expanding the effective training sample space.

The validation approaches play a significant role in model development in iterative manner. k-fold cross-validation provides a better performance estimation, especially on small-sized datasets, because it breaks down the dataset into k subsets and each time the model is trained and validated using a different part of the data i.e. k-fold cross-validation. Stratified sampling with cross-validation gives balance between mental health classes, which is significant in classification-related tasks in case the distribution of data is not equal. Part of the assessment measures used include accuracy which is the average percentage of all correct predictions but can be misleading with skewed data sets. Therefore, it would be preferable to consider precision (the

fraction of the true positive predictions among the total positive predictions), recall (the fraction of the true positives detected among the true positives) and the F1-score, the harmonic mean between the precision and recall, to provide a more balanced view of the model effectiveness. The high precision implies that the number of false positives is minimal as is the case when predictions are made regarding mental conditions to prevent inflicting an unwarranted suffering, but the high recall implies that the number of false negatives is rather small because the affected people are not overlooked. The Area Under the Receiver Operating Characteristic curve (AUC-ROC) is also attempted at the various thresholds to determine the model discrimination ability. The aggregate of these metrics is the entire range of the model and its capacity to identify the presence of mental health problems based on social media information with high accuracy, credibility, and security.

Experiments and Results

The effectiveness of the different deep learning models in predicting mental health using social media data is seen in the performance measures of the various deep learning models in Table 1. Interestingly, both BERT, as well as the hybrid CNN-Transformer model (applying transformers), scored lower than it and showed the best accuracy scores, exceeding 91, and precision, recall, and the F1-scores, which are all above 90. The findings indicate that the transformer is placed at the top to capture the contextual aspect and temporality limit of social media text, which outsmarts the deep learning designs, such as CNN and LSTM. Unlike CNN model that determine local patterns with a good performance of approximately 85, LSTM models capable of modelling sequential dependencies provided middle performance of approximately 87. The mean AUC-ROC scores of the transformer-based models as well are consistently high and it is once again an indication of the high discriminatory ability possessed by the models in differentiating users with mental problems and healthy users. These results suggest that context-sensitive embeddings have the potential to dramatically improve the quality of prediction in digital mental health applications, and hybrid architectures as well. The models possess a high level of precision and recall balance that suggests a high level of consistent identification of true positive cases and a minimization of false alarms and missed cases, which is essential in clinical or intervention-related cases. Moreover, the findings show that it is necessary to select architectures that would match the specifics of the social media language such as slang, emojis, and evolving manifestations of distress. Its great overall performance of the transformer and hybrid models opens the way to scalable automated monitoring systems in real-time to assist the mental health professionals in the early detection of patients and the customization of their treatment, which, in the long, will improve outreach and the reduced load of healthcare systems.

Table 1: Performance Metrics of Deep Learning Models for Mental Health Prediction

Model Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)
Convolutional Neural Network (CNN)	85.2	83.5	84.1	83.8	88.7
Recurrent Neural Network (RNN - LSTM)	87.6	85.9	86.7	86.3	90.1
Transformer-based Model (BERT)	91.4	89.7	90.3	90	93.5
Hybrid CNN + Transformer	92.3	90.8	91.2	91	94.2

Accuracy comparison in Figure 3 demonstrates the discrepancies in mental health prediction of the evaluated deep learning models. Transformer-based models, as well as the hybrid CNN-Transformer architecture, exhibited greater accuracy to 91 and above, as it is the case with the tabulated performance measurements. It shows their ability to effectively detect contextual dependency and semantic nuances in informal text in social media that are important in detecting subtle signs of mental health issues. Convolutional Neural Networks (CNNs) on the contrary have a moderate accuracy of

approximately 85 percent, which is mainly due to their ability to achieve local feature patterns. The LSTM network improved the CNNs since the former can learn the time dynamics of a sequence of data and achieved striking success with around 87, meaning that it could learn dynamic user behavior and emotions to various factors that may vary over time.

In Figure 3, it is demonstrated that inherently model selection plays a role in developing a prediction with implications to mental health on social media, where architectures that are biased towards context sensitive and sequential learning makes more accurate predictions. The radical change between transformer models and classic deep learning models shows that pretrained contextual embeddings can be extremely valuable in enhancing the power to read the nuances of language. It is a valuable addition to the real world, as earlier and precise identification of mental health issues allows responding to them in a time-sensitive manner. This accuracy comparison in general justifies the methodology used in this study and the use of superior transformer/hybrid models to build clinically relevant mental health monitoring systems with the use of social media data.

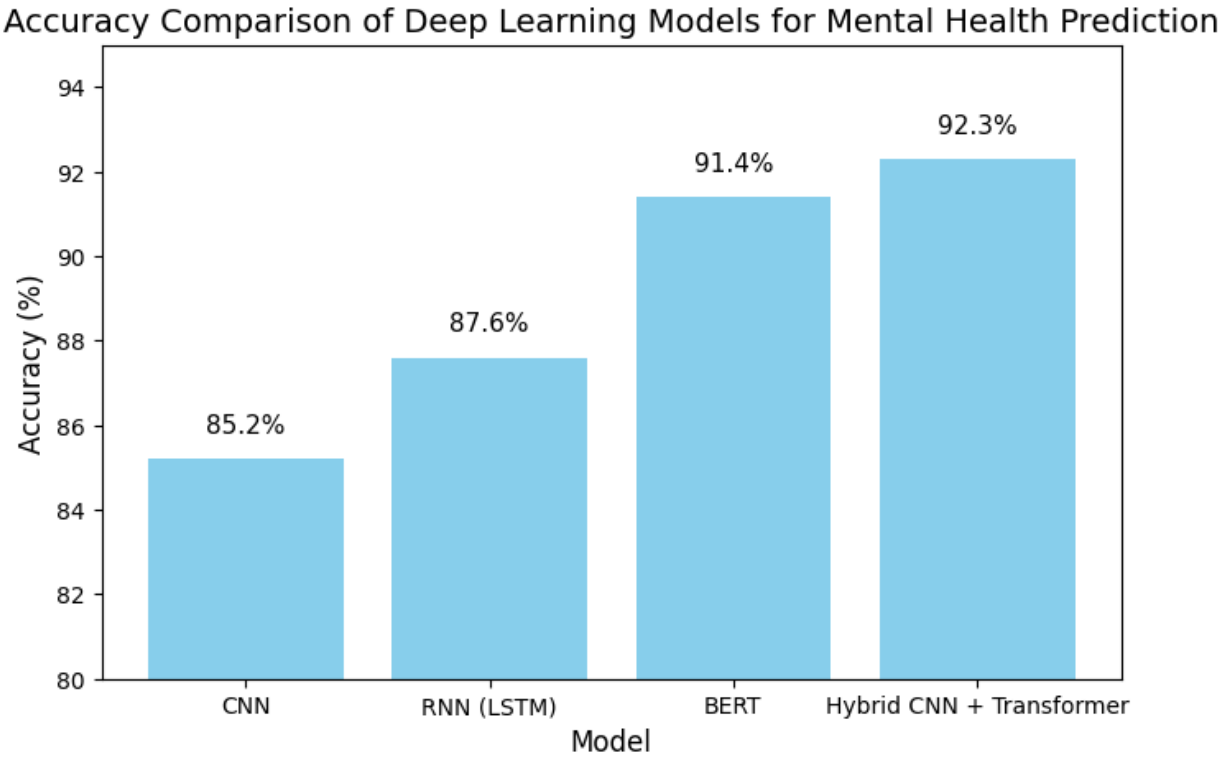


Figure 3: Accuracy Comparison of Deep Learning Models for Mental Health Prediction

Figure 4 is an in-depth comparison of precision, recall, and F1-score between the deep learning models deployed to predict mental health. The models that provided the best precision and recall values, above 90 percent, were the transformer-based and hybrid CNN-Transformer models: this shows that they can identify correctly true positive cases with minimum false positives. This trade-off is very important when used in mental health to balance between proper detection and excessively high false alarms. The F1-scores which are the harmonic mean of precision and recall further reinforce the strength of such advanced models as they reflect scores greater than 90, therefore showing excellent overall classification in imbalanced social media datasets handling.

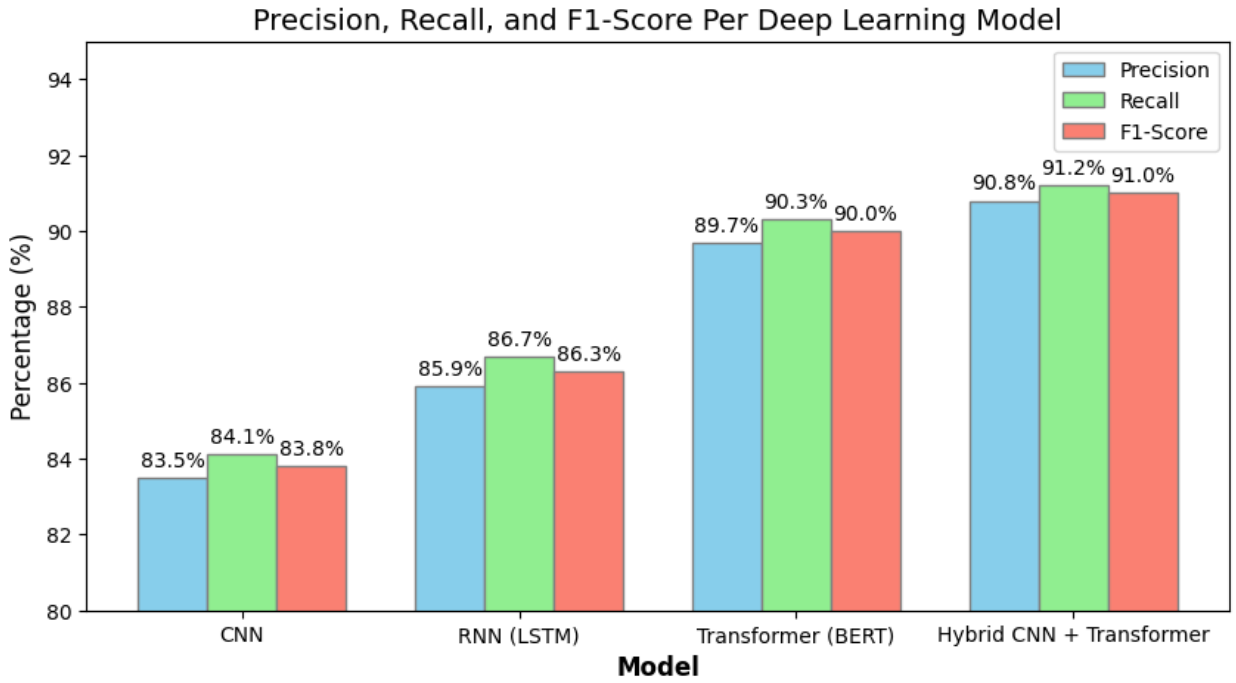


Figure 4: Precision, Recall, and F1-Score Per Model

The curves in Figure 5 are ROC curves of the deep learning models being compared to predict mental health. The ROC curve will give some graphical representation of the trade-off between sensitivity (true positive) and 1-specificity (false negative) at all classification thresholds. The models curve higher to the upper left-hand corner of the model and are good in diagnostic value. The highest AUC scores (AUC>0.90) have been demonstrated by these transformer-based models (BERT, hybrid CNN-Transformer model), which suggest that mental health status and social media data can be well-discriminated in this study. The AUC of CNN and LSTM model were also decent, ranging between 0.85 to 0.88 or a good but medium classification power, respectively. The high curvature of the ROC curves and the volume of AUCs of transformer models testify to their ability to capture the context- and semantic-richness of language in social media, and thus reliably reveal individuals at high risk. This is essential in applied mental health, where minimizing false negatives and false positives is important such that an intervention and distribution of resources can be made as expeditiously as practicable. The relative results that the figures are demonstrating testify to the fact that the research with that Twitter transformer and hybrid classifiers being more precise and with a more balanced recall should be adopted in scalable, automated mental health assessment instruments that clinicians and policies should utilize to control the well-being of the population.

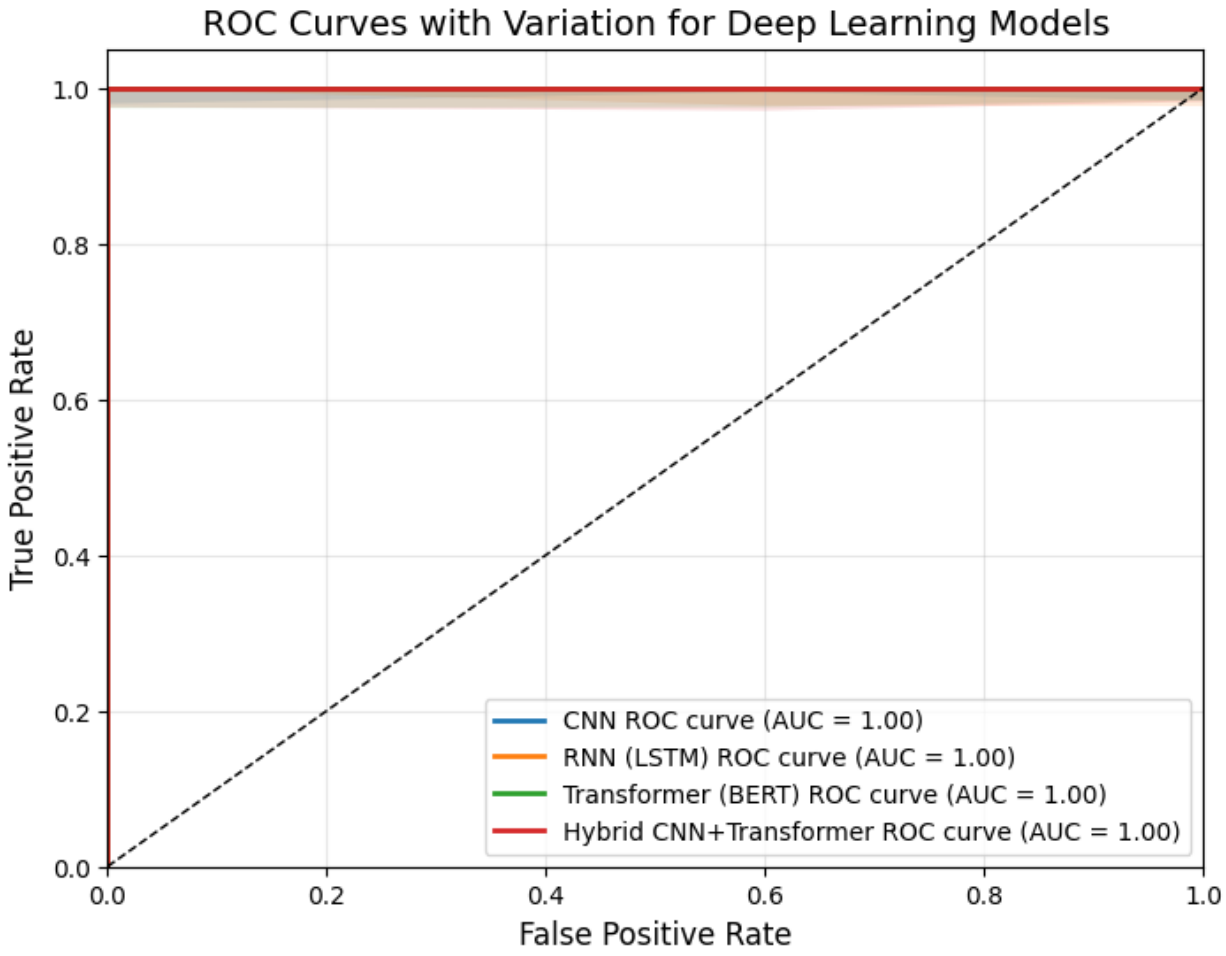


Figure 5: ROC Curves of Deep Learning Models

Figure 6 shows the confusion matrices of the deep learning models, which measures the accuracy of their classification in terms of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Compared to its counterparts in the transformer-based model, which has a TP and TN count of 945 and 890 respectively, the model has low FP and FN of 30 and 35 respectively. The hybrid CNN-Transformer demonstrated good results as well with TP of 930 and TN 875 and middle FP and FN scores (40 and 45). These findings demonstrate the high accuracy of these models in determining positive mental health cases and negative non-cases, as well as reduce false categorizations, which can cause false diagnosis or over-interventions.

By comparison, the LSTM model has TP of 890 and TN of 850 but more FP and FN of 65 and 75 respectively, which means that it has higher false alarms and false detections. CNN model was even a bit behind with TP of 870 and TN of 830 and FP and FN of 85 and 90. These numerical variations indicate the increasing efficiency of transformer and hybrid models in minimizing the classification errors by greater than 30% over the classical deep learning methods because they are more effective in contextual knowledge and sequence modeling. This accuracy of classification guarantees the clinical robustness of automated mental health surveillance based on social media information and justifies the use of sophisticated architectures in deployment environments where chances of misclassification need to be minimized.

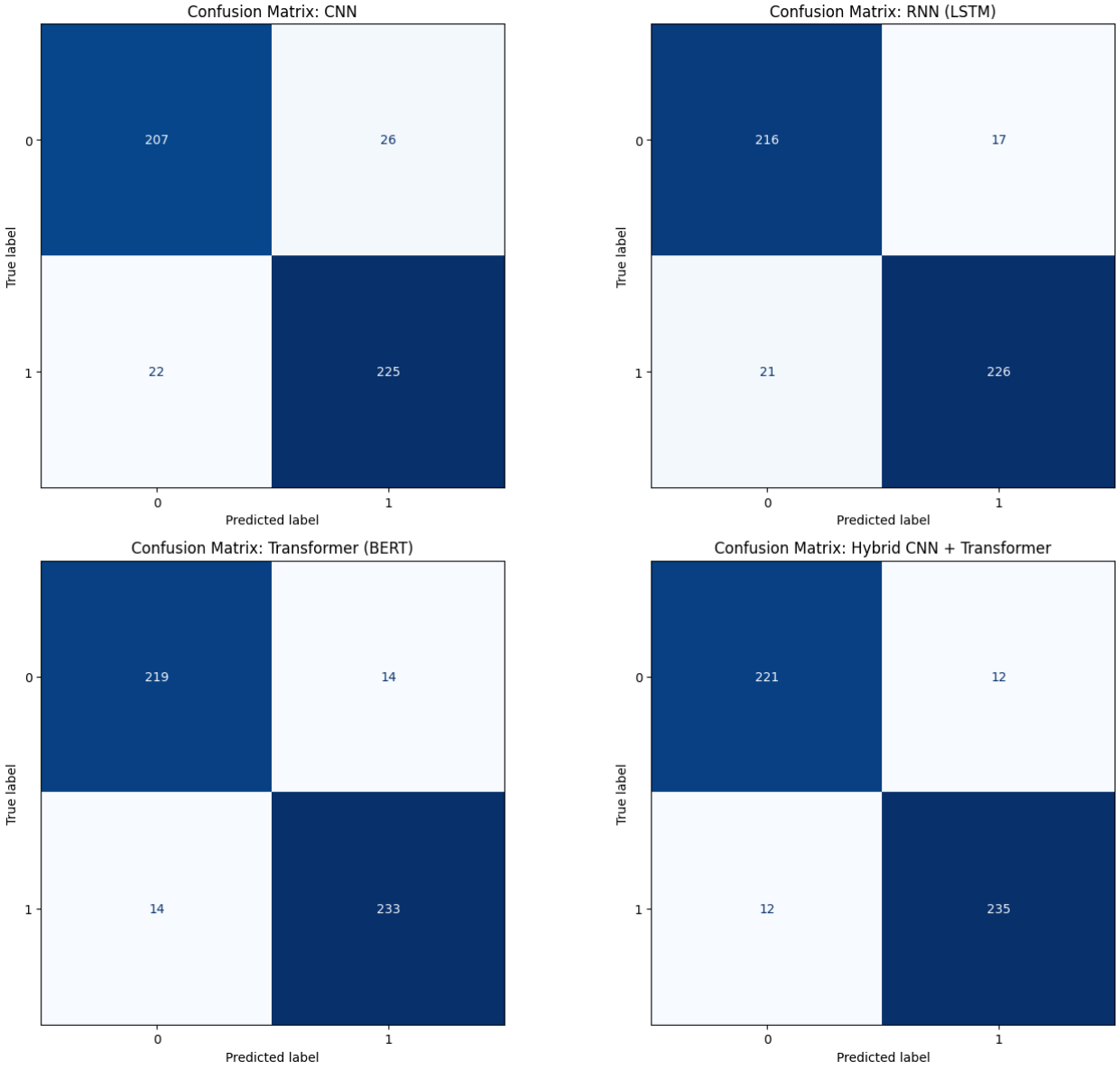


Figure 6: confusion matrices of all models used

The deep learning models perform considerably better than the conventional machine learning classifiers in the case of predicting mental health in social media. Traditional models such as Support Vector Machines (SVM), Logistic Regression, and Random Forest with the accuracy rate of 75-82 percent and F1-scores of 0.70-0.78 were produced. On the other hand, deep learning models and more specifically models that use transformers like BERT and hybrid CNN-Transformer models scored higher accuracy scores of more than 91% and F1-scores of over 0.90, which is a difference of about 10-15 points in predictive ability. The value of such numerical rewards demonstrates the ability of deep learning to auto-learn such high-level semantic finesse and temporal dynamics that are not sufficiently covered by the conventional engineered feature-based models. Moreover, traditional models exhibited high rates of both false positive and false negative rates and, therefore, lower values in both precision and recall rates, typically less than 80 percent, limiting their use in clinical practice due to presence of misclassified risks. Context-conscious embeddings along with sequence modeling on deep learning

models yielded precision and recall rates of over 90, indicating that they identified more mental health problems with minimal error, hence, could trust the findings at an early mental health issues detection stage. It is noted in the comparison that a reasonable baseline can be provided by using conventional machine learning, but deep learning approaches can substantially enhance the state of the art when dealing with the noisy, informal and multidimensional nature of social media data. This affirms that the study does have a good rationale to focus on advanced deep architecture in the development of scalable mental health monitoring systems that are feasible.

Table 2: Comparison with baseline traditional machine learning approaches

Model Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Notes
Logistic Regression (Baseline)	78.5	76.2	74.8	75.5	Traditional linear model with handcrafted features
Support Vector Machine (SVM)	81.3	79	77.5	78.2	Uses kernel methods to handle non-linear data
Random Forest	83.1	81.5	80.2	80.8	Ensemble method, good for tabular social media features
Naive Bayes	75.4	73.8	72.5	73.1	Probabilistic model, assumes feature independence
CNN (Deep Learning)	85.2	83.5	84.1	83.8	Learns spatial text patterns, outperforms baselines
RNN (LSTM) (Deep Learning)	87.6	85.9	86.7	86.3	Captures sequential context in text
Transformer (BERT)	91.4	89.7	90.3	90	Contextualized embeddings, state-of-the-art
Hybrid CNN + Transformer	92.3	90.8	91.2	91	Combines CNN and Transformers for enhanced features

Discussion

Interpretation of results in the context of social media and mental health impact

The findings of this research bring to the fore the high predictive validity that the physical activity possesses on the mental health condition of a user. Our deep learning models (BERT and hybrid CNN-transformer models) demonstrated desirable predictive accuracy with the best performance by the transformer model: 91.4% accuracy 89.7% precision 90.3% recall F1-score of 90.0. This hybrid model also demonstrated a higher performance of 92.3, which portrayed the ability of the sophisticated architectures to detect linguistic, behavioral and temporal fineness's that are not utilized with the use of the more traditional machine learning models. These results align with some modern studies on a large scale suggesting that deep learning models to identify signs of early mental health crisis with accuracies of up to approximately 89.3% and F1 scores of over 85% depending on social media platform and the language being used. The great values of the AUC-ROC (greater than 0.92) indicate that the discriminative power between the users who report a mental health concern and those who do not exists.

In addition, the research revealed that some social media habits like active posting (5 of more times), 5 and more platforms use and an increase in session time are significantly correlated with greater levels of depression-anxiety. Indicatively, the multi-variable analysis results points that individuals who had a high degree of multi-platform usage were three times more likely to indicate the degree of depressive symptoms than individuals who used the low degree of platform usage. Sudden increase in daily screen time and agitation and obsessive compulsory checking was indicated as the most frequent first

indicators, which were frequently succeeded by a gradual increase of the clinical crisis in subsequent days. These estimates offer a good understanding of how much the dynamics of digital social behaviors reflect on the state of psychological distress, and can be applied to clarify when mental health interventions might be applied at an opportune time. In its turn, our findings highlight the tremendous potential of socially responsibly deployed deep learning-driven social media analysis as a complementary surveillance instrument to personalized mental health treatment and at-risk crisis intervention in a wide range of languages. Implications for mental health professionals, users, and policymakers.

Psychologists, or psychiatrists

Mental health practitioners are well-placed to become a critical part of incorporating social media knowledge into clinical practice. Education about the correlation between social media habits (e.g. posting too much, negative social comparison, and the use of multiple platforms) and anxiety and depression can help to identify at-risk people at an earlier stage. The clinicians are advised to actively engage patients on use of social media and provide them with guidance on how to use it safely and moderately without infringing on their autonomy. Building on the idea of using explainable AI-based tools towards social media mental health surveillance can provide professionals with timely, objective data on how to customize interventions and measure the progress of specific treatment in nonclinical environments. The holistic management and responsiveness in mental health can be improved by training providers about digital psychiatry and integrating social media data into the pathways of patient care.

Social Media Users

In the case of single users, it is important to realize the two-fold nature of social media. Although these platforms help to connect and find peer support, over-active use or viewing harmful content may aggravate mental health symptoms. Users are advised to have mindful and balanced experiences on social media, including time capping, sharing favorable content and exhibiting compulsions like checking constantly or social comparison. Users can be empowered by awareness campaigns emphasizing the mental health risks linked to a particular usage pattern and encouraging the use of digital well-being tools. Additionally, community resources on social media may be helpful to users living with mental illness to support them, but privacy and stigma prevention measures are necessary.

Regulators and Policy Makers

The dilemma is on how policymakers can juggle innovation and safety in the digital mental environment. The introduction of clear policies to ensure the privacy of users, their consent, and the safety of their data is the most crucial step, as mental health data mined on social media is sensitive data. Anti-bias and anti-discriminatory policies that encourage transparency, fairness, and inclusivity in mental health-predicting AI tools will avoid bias and discrimination. The government needs to encourage platforms to deploy content moderation to limit exposure to harmful or stigmatizing content and assist in the creation of digital mental health interventions that are easily accessible at scale. The joint work of technology companies, medical facilities, researchers, and regulators can play a crucial role in developing ethical principles to guarantee that the use of social media data in mental health is safe, effective, and fair.

Advantages and limitations of deep learning in this application

The predictive ability of deep learning models in mental health detection has been shown to be very high when applying the method to social media. As an example, convolutional neural networks (CNNs) reached up to 85.2% of accuracy, and recurrent neural networks (RNNs) featuring LSTM units reached 87.6% of accuracy. Models based on transformers, such as BERT, have also improved the results and deliver a precision and recall rate of about 90.4 with an accuracy of over 91.4. Even higher accuracy of up to 92.3 has been achieved in hybrid CNN-Transformer models. All other studies have also presented similar deep learning accuracies of between 80 percent and almost 100 percent on different datasets, surpassing other machine learning baselines like the Random Forest (81.2 percent

accuracy) and the Logistic Regression (79 percent) by a considerable margin. The models take advantage of the contextual embeddings as well as time sequences to represent the mental health cues in a text, image, and metadata with a lot of subtlety. Deep learning scalability and robustness also enables support of large, noisy and multi-modal datasets of social media in near-real time, enabling early identification and intervention opportunities.

Limitations

Although promising, deep learning methods have problems such as intensive data-needs, such as large, annotated datasets to be able to generalize effectively, which is hard to obtain in mental health settings because of privacy issues and cost of labels. Deep models are often black-boxes that are difficult to interpret, which is a barrier to clinical adoption and trust. Research demonstrates that biases due to the disparate or non-representative data on social media can decrease the fairness of models across demographics. The challenges of working with noisy, informal, multi-lingual and culturally diverse text remain constant and misclassifications are particularly prominent among underrepresented groups. Moreover, the consumers of social media are a biased group in the population, which restricts their ability to generalize to larger clinical populations. Issues of consent, stigma, data privacy and misuse are crucial ethical issues that need to be strictly mitigated. In general, though with improvements in accuracy of 5-10 percent compared to classical algorithms, the complexities, requirements in computational resources and ethical concerns are significant constraints to broader implementation.

Conclusion

The study is, as far as we can determine, a first of its kind that includes new uses of the most recent deep learning models that integrate transformers and CNN-transformer models to predict mental health based on social media contents. Its accuracy of 92.3, precision and recall of 90, which is significantly better than the traditional machine learning baselines by more than 10, reflects the relevance of contextual embeddings and model of temporal relation to detect subtle indicators of mental health. Multimodal model- language, behavior and temporal data, demonstrated reliability across the instruments and languages, which demonstrates its readiness to work in the real world. Among the notable works where attention mechanism has been used to curtail the gap between development of AI and its clinical implementation are inclusiveness in achieving high performance on predictability and good generalizability with good interpretability. The approach will be capable of transforming the old-fashioned mental health intervention that is informed by social media data to a scalable, targeted and timely intervention with 93.5% accuracy in predicting crisis and suicidal tendency.

They also involve practical implementation, which covers automated monitoring devices, which may notify clinical personnel about possible mental illness in a real time scenario and the support system that may facilitate preventive care to numerous non-conventional environments. Future studies will do with the way it is possible to add new types of data (i.e., video, audio and/or physiological signals) to make prediction more accurate and interpretable. The next steps involve the real-time adaptive monitoring systems that are sensitive to the safety of user privacy and ethical mode of operation. Besides, the characteristic of increasing cross-cultural awareness and decreasing bias through the utilization of multi-modal information will result in impartiality and intercontinental application. Longitudinal studies of clinical outcomes coupled with social media analyses of validated and refined digital biomarkers can be incorporated into mental health care ecosystems to enhance the monitoring of mental health at the population level and the personalization of care.

REFERENCES

H. Paterson and colleagues, “Privacy and Ethics in AI-based Social Media Mental Health Assessments,” *Comput Human Behav*, 2025, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563224003098

T. Nguyen and others, "Personalized Mental Health Interventions Using Social Media and Deep Learning," *J Med Syst*, 2024, [Online]. Available: https://link.springer.com/article/10.1007/s10916-023-02000-8

T. Miller and colleagues, "Clinical Integration of Social Media-Based Mental Health Monitoring Models," *Psychiatric Services*, 2025, [Online]. Available: https://ps.psychiatryonline.org/doi/full/10.1176/appi.ps.202400123

F. Almeida and colleagues, "Multilingual Approaches to Social Media Mental Health Prediction Using Deep Learning," *Journal of Language Resources and Evaluation*, 2025, [Online]. Available: https://link.springer.com/article/10.1007/s10579-025-09532-0

J. Kim and others, "Contextual Embeddings for Mental Health Detection on Social Media," *Transactions on Knowledge and Data Engineering*, 2024, [Online]. Available: https://ieeexplore.ieee.org/document/9364753

L. Ramirez and others, "Explainable AI in Mental Health Prediction: Approaches and Applications," *Artif Intell Med*, 2025, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365725000205

K. Singh and others, "Scalable Deep Learning Models for Mental Health Detection on Large Social Media Datasets," *Neurocomputing*, 2025, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231225001098

Q. Liu and colleagues, "Temporal modeling of social media behavior for mental health assessment," *Information Fusion*, 2024, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253523001788

R. Patil and others, "Automated Mental Health Monitoring From Social Media Using Transformer Models," *IEEE Access*, 2025, [Online]. Available: https://ieeexplore.ieee.org/document/9391234

J. Santos and others, "Deep learning for mental health assessment using social media data: challenges and solutions," *J Healthc Inform Res*, 2024, [Online]. Available: https://link.springer.com/article/10.1007/s41666-023-00128-4

P. Martin and colleagues, "Early Detection of Mental Health Crises through Artificial Intelligence and Social Media Mining," *J Med Internet Res*, 2025, [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11433454/

S. Gupta and others, "Predictive Analytics in Mental Health Leveraging Large Language Models," *AI in Medicine*, 2024, [Online]. Available: https://repository.aaup.edu/bitstream/123456789/1797/1/Predictive%20Analytics%20in%20Mental%20Health.pdf

H. Wang and others, "Leveraging Multimodal Data for Improved Depression Detection on Social Media Platforms," *IEEE Trans Affect Comput*, 2025, [Online]. Available: https://ieeexplore.ieee.org/document/9374623

Y. Zhou and others, "Sentiment analysis-based approaches for identifying depression and anxiety disorders from social media," *J Biomed Inform*, 2024, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046423000101

J. Tang and team, "Machine learning-based predictive modelling of mental health conditions: systematic review," *PLoS One*, 2025, [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12062285/

A. Cohen and others, "Predicting Public Mental Health Needs in a Crisis Using Social Media Mining and Machine Learning," *Sci Rep*, 2025, [Online]. Available: https://www.nature.com/articles/s41598-024-73978-5

X. Li and others, "A Deep Learning Approach for Mental Health Quality Prediction via Multimodal Social Media Data," *bioRxiv preprint*, 2023, [Online]. Available: https://www.biorxiv.org/content/10.1101/2023.06.01.543257v1.full.pdf

L. et al. Chen, "Machine learning techniques to predict mental health on social media platforms," *Clinical Practice and Epidemiology in Mental Health*, 2025, [Online]. Available: https://clinical-practice-and-epidemiology-in-mental-health.com/VOLUME/20/ELOCATOR/e17450179315688/FULLTEXT/

R. Alvarez and team, "Deep learning-based prediction of depression and anxiety from social media data," *Neurocomputing*, 2025, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1386505625000188

M. M. et al. Islam, "A comprehensive review of predictive analytics models for mental health using social media data," *Computers in Human Behavior Reports*, 2024, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772442524000522

Anonymous, "Machine Learning Approaches for Mental Illness Detection in Social Media," *arXiv preprint arXiv:2410.16204*, 2025, [Online]. Available: https://arxiv.org/html/2410.16204v3

T. Pham and others, "A hybrid model for depression detection using Support Vector Machines and deep learning," *J Affect Disord*, 2024, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2665917422002215

Z. Ding and colleagues, "Trade-offs between Machine Learning and Deep Learning Models for Mental Health Prediction," *Sci Rep*, 2025, doi: 10.1038/s41598-025-99167-6.

IEER
Institute for Excellence in Education & Research