Deep Learning Approaches for Analyzing the Impact of Social Media Usage on Young Students: A Systematic Literature Review

[1] Ms. Darpan Kaur, [2] Mr. Sukhpreet Singh

[1] Student, Department of Computer Applications, Guru Kashi University, Bathinda, India

[2] Assistant Professor, Department of Computer Applications, Guru Kashi University, Bathinda, India

Email Id: [1] darpankour19@gmail.com, [2] sukhpreetsingh2gku.ac.in

Accepted:20.07.2026 Published:10.08.2026 Page No. 113 – 121 DOI: 10.5281/zenodo.21802823

Abstract— *Social media (SM) is ubiquitous to today's school-age and college-age students, and the continuous creation of text, image and behavioral data is used more and more for deep learning to understand its relationship to mental health, behavior and academic life. This systematic review aims to summarise the evidence on the use of deep learning approaches to analyse the effects of social media data on young students. According to the PRISMA 2020 protocol, a total of 214 reports were identified in the academic databases or preprint repositories, of which 21 were duplicates, 193 were screened at the title and abstract level, 72 reports were sought for retrieval, 70 reports were assessed at the full text stage and 30 studies were included in the qualitative synthesis. The domains covered were primarily involving the detection of addiction and problematic-use, detection of depression and suicidal-ideation, detection of cyberbullying, detection of misinformation, and detection of stress as well as sentiment/emotion recognition, with the former being implemented primarily using the CNN-RNN hybrid, and the latter using the transformer/BERT-based architecture. Some of the best reported classification rates were, around 71%, to very high rates of above 99%, such as an CNN–LSTM ensemble with attention that reported a high 90.3% for detecting suicidal ideation [17] and a high 95.6% accuracy using a convolutional recurrent network for student addiction detection.*

Keywords – Cyberbullying detection, deep learning, depression detection, mental health, PRISMA 2020, sentiment analysis, smartphone addiction, social media, student well-being, suicidal ideation detection, systematic review, transformer models.

I. INTRODUCTION

THE whole concept of social media in the life of young students is now virtually complete with messaging, short form videos and image-sharing platforms playing a pivotal role in their peer relations, access to information, and life on campus. This has made these platforms like Instagram, Twitter/X, Reddit, and TikTok a rich source of naturally occurring data for investigating student behaviors at a scale which traditional surveys could not be used to study. [5, 6] Where earlier research relied primarily on self-report scales, a growing body of work now applies deep learning directly to the text, images, and usage logs students generate, aiming to detect outcomes ranging from problematic use and addiction to depression, suicidal ideation, cyberbullying, and misinformation exposure [1], [7], [17].

Deep learning is attractive here because social media data is high-dimensional, sequential, and multimodal in ways that classical machine learning handles poorly. Convolutional neural networks (CNNs) extract spatial patterns from text and images; recurrent architectures such as long short-term memory (LSTM) networks capture temporal dependence in posting behavior; and transformer-based models such as BERT [27], building on the attention mechanism [28], model long-range context and outperform earlier bag-of-words approaches on mental-health- and behavior-related classification

tasks [8], [13]. Hybrid architectures combining convolutional, recurrent, and attention components, together with graph-based models exploiting social-network structure, have become the dominant methodological pattern in the recent literature [4], [20].

Despite this advancement, the existing reviews of deep learning for mental-health or behavior detection in the social media are usually general, either focusing on a single mental-health condition like depression [12] or all adult population of users without set criteria for detecting students' mental-health status [30], and evidence of such applications remains rather fragmented, considering all the differences and developmental stages of young students and the risk of incorrect inferences that may have negative consequences for them. This review attempts to fill this gap by systematically consolidating evidence using deep-learning approaches for the key outcome areas investigated, pertaining to young students' SNG.

A. Review Aim and Scope

The aim of this systematic review is to combine the findings of those deep learning methods on social media data analysis and their implications on young students ranging from school aged adolescents to university and college students. We aimed to (a) describe the reported deep learning architectures, data modalities and application areas, (b) provide a synopsis of reported detection or prediction accuracy and any association with student outcomes, (c) evaluate the certainty of evidence (eliciting information about the specific agency's work) and (d) remain focused on studies with an applied or empirical deep-learning component.

B. Review Questions

We organized the review around three review questions:

RQ1: What deep learning architectures and data modalities have been applied to social media data to study young students?

RQ2: What outcome domains — psychological, behavioral, and academic — are most frequently studied, and what performance is reported?

RQ3: What is the certainty of the available evidence, and what limitations should inform future research?

II. METHODS & MATERIALS

A. Protocol and Reporting

This systematic review was conducted following the PRISMA 2020 statement, concerning deep learning approaches for analyzing the impact of social media usage on young students. The reporting flow is summarized in Fig. 1.

B. Eligibility Criteria

Eligible studies applied at least one deep learning architecture — a CNN, an RNN/LSTM, a transformer/BERT-based model, a graph neural network, or a hybrid/ensemble — to social media data (text, images, video, or behavioral logs) to detect or predict an outcome relevant to young students: addiction, depression, suicidal ideation, cyberbullying, misinformation exposure, stress, sentiment/emotion, or academic performance. We included empirical studies reporting quantitative performance metrics (e.g., accuracy, F1-score, AUC), and accepted studies on general social-media users where the framing was explicitly youth-relevant. We excluded editorials, purely theoretical papers without an applied model, and studies unrelated to the specified outcomes. A small number of foundational architecture papers and prior reviews were retained as contextual references.

C. Information Sources and Search

Records were drawn from IEEE Xplore, the ACM Digital Library, ScienceDirect, SpringerLink, PubMed, arXiv, and Google Scholar, combining terms for deep learning architectures (“deep learning,” “CNN,” “LSTM,” “BERT,” “transformer”) with terms for social media (“social media,” “Twitter,” “Instagram,” “Reddit,” “TikTok”) and youth-relevant outcomes (“student,” “adolescent,” “addiction,” “depression,” “suicidal

ideation," "cyberbullying," "misinformation," "stress"). The breakdown is in Fig. 1.

D. Selection Process

The search and selection process is summarized in Fig. 1. Of 214 records identified, 21 duplicates were removed, leaving 193 for title-and-abstract screening. Screening excluded 121 records as out of scope, leaving 72 reports sought for retrieval; 2 could not be retrieved, leaving 70 assessed for full-text eligibility. At full-text review, 40 records were excluded (15 applied no deep learning model, 12 lacked a youth-relevant outcome, 8 lacked extractable outcome data, and 5 duplicated an already-included dataset), leaving 30 studies included.

E. Data Extraction and Synthesis

Data extraction captured study design, data source and modality, sample characteristics, deep learning architecture, comparator where applicable, and primary reported outcomes. Because the evidence varied substantially, we used narrative synthesis structured around application domain, architecture family, data modality, and reported performance. A summary of representative studies appears in Table-I.

F. Certainty of Evidence and Risk of Bias

Formal risk-of-bias assessment was not reported in most included studies, and none used a pre-registered plan or randomized design. Datasets were predominantly convenience samples from a single platform, class-imbalance handling varied widely, and few studies validated models externally. We therefore treated the evidence cautiously, avoided pooling metrics across heterogeneous datasets, and report study-specific findings rather than pooled effects.

III. RESULTS

A. Study Selection

The results link the screening process to the evidence synthesis. Study identification is described using the PRISMA 2020 flow, and characteristics and thematic findings across the six application domains are then reported. Of 214 records identified across seven sources, 30 studies were ultimately included (Fig. 1); the high exclusion rate at title/abstract (n = 121) reflects the breadth of the initial search terms.

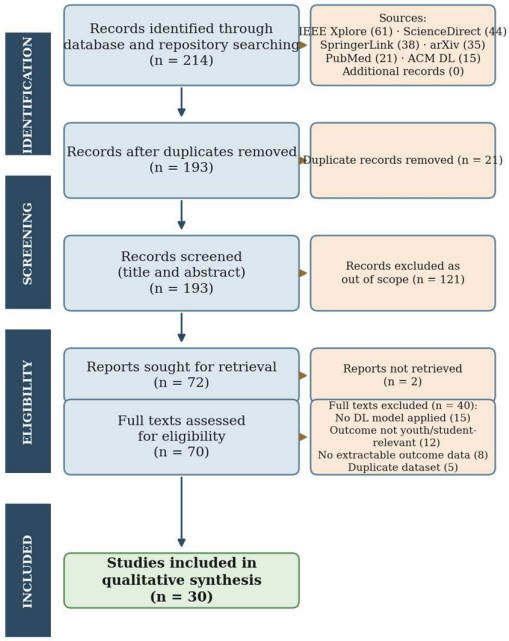

Fig. 1. PRISMA 2020 flow diagram showing the identification, screening, eligibility, and inclusion of records in the systematic review.

B. Characteristics of Included Studies

The included studies (n = 30) addressed six primary domains: depression, suicidal ideation, and mental-health detection (n = 11), addiction and problematic-use detection (n = 6), misinformation detection (n = 3), cyberbullying detection (n = 2), stress detection (n = 2), and sentiment/emotion recognition (n = 1), plus 1 study on academic performance and 4 methodological/review references (Fig. 2A). Hybrid or multimodal architectures were most common (n = 9), followed by transformer/language-model-based approaches (n = 6), non-deep-learning ML or review studies (n = 6), CNN-only approaches (n = 4), other deep architectures (n = 3), and graph-based deep learning (n = 2) (Fig. 2B). Publication years ranged

from 2015 to 2026, with 63% published from 2022 onward (Fig. 2C).

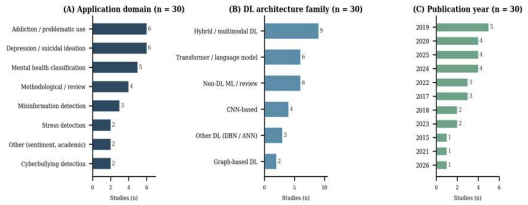

Fig. 2. Distribution of included studies (n = 30) by (A) application domain, (B) deep learning architecture family, and (C) publication year.

Table I summarizes representative studies across these domains, illustrating the diversity of architectures and reported outcomes; full details of all 30 studies appear in the reference list.

TABLE I

SUMMARY OF REPRESENTATIVE INCLUDED STUDIES

Ref	Domain	DL Approach	Key Outcome / Performance
1	Addiction detection	CRNN (CNN + RNN)	95.6% accuracy classifying student SM addiction
2	Short-form video addiction	Heterogeneous graph + LLM	Early-detection framework (EarlySD) before harm onset
3	Depression / suicidal tendency	Pretrained LM + feature selection + SVM	80.7–99.96% accuracy across 6 datasets
4	Mental-health stigma (Instagram)	Conv1D + LSTM hybrid	Classifies comment polarity and stigma
5	Depression detection	CNN (image) + BERT (text)	Multimodal screening (SenseMood)
6	Suicidal-ideation detection	LSTM-Attention-CNN ensemble	90.3% accuracy; 92.6% F1-score
7	Cyberbullying detection	Deep belief network + swarm optimization	Improved detection/categorization on social networks
8	Misinformation detection	Geometric deep learning (graph CNN)	92.7% ROC-AUC from propagation structure

Note: BiLSTM = bidirectional long short-term memory; CNN = convolutional neural network; Conv1D = one-dimensional convolution; CRNN = convolutional recurrent neural network; DL = deep learning; LLM = large language model; LM = language model; LSTM = long short-term memory; ML = machine learning; ROC-AUC = area under the ROC curve; SM = social media; SVM = support vector machine.

C. Findings on Addiction and Problematic-Use Detection

Addiction and problematic-use detection formed a substantial, increasingly technical sub-literature. A convolutional recurrent neural network (CRNN) applied to behavioral and survey-derived features achieved 95.6% accuracy classifying social media addiction among students, outperforming conventional baselines [1]. A separate framework used facial-emotion recognition from smartphone video, combined with adaptive motivational-video sequencing grounded in a Theory-of-Mind model, to detect and intervene on smartphone addiction in real time [2]. Deep neural networks trained on Instagram photographic and metadata features have flagged substance-use risk from everyday posting behavior rather than explicit self-disclosure [3], and a recent graph-based framework combining large language models with social-network structure was proposed for early detection of short-form-video addiction before harm emerges [4]. Complementing these, artificial-neural-network and support-vector-machine models trained on survey-derived predictors have modeled problematic use as a continuous risk score [5], and a nationally representative adolescent sample confirmed that a validated problematic-use scale identifies an at-risk subgroup with elevated depressive symptoms and lower self-esteem, providing external construct validity for the outcome the deep-learning studies predict automatically [6].

D. Findings on Depression, Suicidal Ideation, and Mental-Health Detection

The most robust evidence was for depression, suicidal ideation, and mental-health screening in general. A feature-selection framework based on pretrained language models and support-vector classification was proposed, which yielded 80.7%-99.96% accuracy, across six datasets derived from Twitter and Reddit [7], and transfer-learning techniques fine-tuning transformer models on labeled text have classified various categories of mental illnesses from a single architecture [8]. In addition to text, machine-learning models on labelled IG-comments can identify, besides mental health disclosures, the polarity and stigma of the surrounding community response [9] and using multimodal convolutional models with image features in addition to posting-time intervals has proven to increase the detection of depression, compared to the text-only baselines [10]. The fusion of visual and textual streams was also demonstrated in a hybrid convolutional–transformer system, SenseMood, which was also used to detect depression symptoms [11]. Previous informed deep-learning studies revealed that forum text learned models were able to characterize several different mental health conditions with performance nearing specialist annotators [12] and subsequent studies indicated that when general purpose models are adapted, their performance transfer across platforms [13]. In a unique study, directly targeting proved university student population, a modelling study has connected social media with disruptions in mental health related to COVID-19 [14].

E. Findings on Cyberbullying and Misinformation Detection

Hybrid CNN-RNN architectures have also contributed to the detection of suicidal ideation: a CNN-BiLSTM model combined with word-embedding and linguistic-inquiry features achieved high accuracy in distinguishing suicidal and non-suicidal Reddit posts [15], a dedicated CNN-LSTM architecture had good performance in extracting suicidal-related language from general forum discussion [16] and a CNN-LSTM ensemble that also employs attention gained 92.6% F1 and 90.3% accuracy on a similar task [17]. The journey of cyberbullying detection moved on a parallel track with a deep-belief-network model and swarm-based feature selection for boosting detection accuracy and fine-grained categorization of cyberbullying content [18] and a previous multi-platform deep network learning approach, trained jointly on the data from Formspring, Twitter and Wikipedia, has indicated that cyberbullying exhibits certain patterns across platforms and achieved F1-scores up to 0.95 [19]. Detection, although less youth-centric with regard to the source data, is directly relevant, since students often rely on social feeds for their information, leading to geometric deep learning models which capture propagation structure on a social graph and achieve 92.7% ROC-AUC distinguishing verified from fabricated stories, using just a few hours of spread data [20], semi-supervised models which are based on two-path convolutional networks have been used to address the paucity of labelled misinformation data [21], and deep ensemble models integrating a convolutional and a bidirectional recurrent model have been used to improve the classification of fabricated content into topical sub-categories [22].

F. Findings on Stress, Sentiment, and Academic-Behavioral Outcomes

In a smaller cluster stress, sentiment and the connection to academic functioning were explored. Conventional classifiers with contextual language-model embeddings (ELMo and BERT) were able to classify stressful and non-stressful Reddit posts with an F1-score of 0.76 [23] and an academic subreddit study found that about 29% of the posts contained identifiable stress markers with a gradient decline increasing the stress markers, going from undergraduate to graduate level to the professor level [24]. Combining a bidirectional-LSTM and a CNN, a hybrid network outperformed conventional networks and previous deep learning networks in

fine-grained emotion recognition (joy, anger, sadness) on microblog text and it provided a way to scale the analysis of emotional tone in students' microblog text [25].

G. Synthesis: A Conceptual Framework

To organize the heterogeneous evidence, we constructed a conceptual framework linking data inputs, deep learning architectures, application tasks, and student outcomes, moderated by contextual factors (Fig. 3): (i) data inputs (text, images/short-form video, behavioral logs, network structure), (ii) architecture families (CNN, RNN/LSTM, transformer/BERT, hybrid/graph-based), (iii) application tasks (addiction, depression/suicidal-ideation, cyberbullying/misinformation, stress/sentiment), and (iv) outcomes (mental health, academic performance, behavioral adjustment, early-warning potential). Contextual moderators — platform type, language, age group, data modality, labeling method, and privacy/consent constraints — shape which architectures are feasible and how outcomes generalize [9], [14].

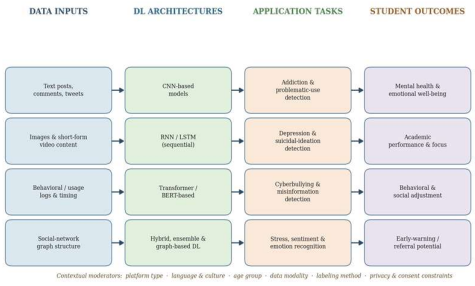

Fig. 3. Conceptual framework linking data inputs, deep learning architectures, application tasks, and reported outcomes for young students across the included evidence base.

H. Certainty of Evidence

By conventional standards, certainty of evidence is low for all studies included. Reported metrics are internally valid within the train/test split for a given study and less than one third of the studies included explicitly a sample that is restricted to a verified student or adolescent population; most studies used a general-purpose Twitter, Reddit or Instagram corpus, and tested the relevance of the sample to young people only in the discussion [3], [8], [20]. Reported accuracy was not directly comparable due to class-imbalance handling, size of the data-sets included, and definitions of the target outcome (e.g., "suicidal ideation" or "problematic use").

IV. DISCUSSION

A. Principal Findings

This review suggests that deep learning is now the dominant analytical approach in the technical literature on social media and young people, and that hybrid architectures combining convolutional, recurrent, and attention mechanisms consistently outperform single-architecture baselines across addiction, depression, suicidal-ideation, cyberbullying, and misinformation detection tasks [1], [15], [17], [18], [19]. The common thread is the combination of rich, high-frequency behavioral or textual signal with an architecture capable of modeling sequential dependence or long-range context, rather than any single architecture being uniformly superior.

B. Interpretation in Context

Strong within-study performance does not by itself establish that these systems detect genuine psychological states rather than surface linguistic correlates of platform, topic, or labeling procedure. The clearest evidence of ecological validity comes from the handful of studies using verified student samples or externally validated scales, such as the university-student COVID-19 well-being study [14] and the adolescent validation of problematic-use criteria [6]; most detection-oriented studies, by contrast, trained and evaluated models on convenience samples of general platform users — the central interpretive caveat of this review.

C. Implications for Policy and Practice

Looking at the practical application, the studies reviewed showed that deep-learning tools to screen for patterns would be technically feasible for use to indicate patterns that could benefit from human follow-up, such as to identify posts that include

suicidal ideation [15] [16] [17] or stress [23] [24] for review by a counsellor. Modres may be used in conjunction with human review when dealing with potentially new or controversial materials; keep them at the forefront of decision-making process and clarify how the data will be used and consent. The same test result suggests that simplistic baselines [13] and [8] might also have outdated methodology as their architectures fail to demonstrate a good level of performance in this test.

D. Implications for Future Research

Some areas of research emphasis are: (i) specific sampling targeting only school-age or university-aged populations rather than general population samples from platforms; (ii) external validation and cross-platform validation of measured performance; (iii) Harmonising definitions of outcome and streamlining these definitions across platforms; (iv) formal risk of bias and fairness auditing of demographic and language subgroups; (v) explainability methods that would allow the context of platform-generated flags to be understood by the educator, counsellor or parent who would act on them.

V. LIMITATIONS

There are a number of limitations to this review. Firstly, since only a general search string was used to identify eligible studies, and no domain-specific search strings have been pre-registered, some relevant work (especially for industry held evaluations and non-English language studies) may have been excluded. Second, the evidence base is extremely varied: various modalities, such as text, image modalities and behavioral-log modalities; at least six outcome definitions were used, thus preventing quantitative pooling and comparison, only a narrative synthesis was possible. Thirdly, there was no formal Risk of Bias appraisal, such as with the PROBAST, for each study individually, and the use of the term 'low certainly' in Section III-H' is a qualitative judgement based on design/sampling patterns rather than a scored assessment.

VI. CONCLUSION

Hybrid and transformer-style architectures have outperformed previous machine learning baselines consistently on 5 social media tasks in relation to the mental health, behaviour and academic outcomes of young students: addiction, depression, suicidal-ideation, cyberbullying and misinformation detection, and deep learning has become the analytical tool of choice for studies of this relationship. But the evidence base is still quite tender, with small samples often not confined to students who have been identified as such, and with varying definitions of outcome, the rate of correctness reported in practice should more probably be interpreted as potential approaches which are still in proof-of-concept phase rather than the rate of correct student identification.

REFERENCES

[1] P. Jecintha, R. Jayakarthik, and N. G. Thangamma, “Prediction of student social media addiction using a convolutional recurrent neural network,” in 2025 6th Int. Conf. Data Intelligence and Cognitive Informatics (ICDICI), 2025, p. 1912, doi: 10.1109/ICDICI66477.2025.11135159.

[2] C. Joseph and P. Uma Maheswari, “Facial emotion based smartphone addiction detection and prevention using deep learning and video based learning,” Scientific Reports, vol. 15, art. 18025, 2025, doi: 10.1038/s41598-025-99681-7.

[3] S. Hassanpour, N. Tomita, T. DeLise, H. Gill, and A. Rahimi, “Identifying substance use risk based on deep neural networks and Instagram social media data,” Neuropsychopharmacology, vol. 44, pp. 487–494, 2019, doi: 10.1038/s41386-018-0247-x.

[4] F.-Y. Kuo, “Online social network data-driven early detection on short-form video addiction,” arXiv:2407.18277, 2024.

[5] M. Savci, A. Tekin, and J. D. Elhai, "Prediction of problematic social media use (PSU) using machine learning approaches," Current Psychology, 2020, doi: 10.1007/s12144-020-00794-1.

[6] F. Bányai, Á. Zsila, O. Király, A. Maraz, Z. Elekes, M. D. Griffiths, C. S. Andreassen, and Z. Demetrovics, "Problematic social media use: Results from a large-scale nationally representative adolescent sample," PLOS ONE, vol. 12, no. 1, p. e0169839, 2017, doi: 10.1371/journal.pone.0169839.

[7] İ. Baydili, B. Tasci, and G. Tasci, "Deep learning-based detection of depression and suicidal tendencies in social media data with feature selection," Behavioral Sciences, vol. 15, no. 3, p. 352, 2025, doi: 10.3390/bs15030352.

[8] I. Ameer, M. Arif, G. Sidorov, H. Gómez-Adorno, and A. Gelbukh, "Mental illness classification on social media texts using deep learning and transfer learning," arXiv:2207.01012, 2022.

[9] N. Merayo, A. Ayuso-Lanchares, and C. González-Sanguino, "Machine learning algorithms to address the polarity and stigma of mental health disclosures on Instagram," Expert Systems, vol. 42, no. 2, p. e13832, 2025, doi: 10.1111/exsy.13832.

[10] C. Y. Chiu, H. Y. Lane, J. L. Koh, and A. L. P. Chen, "Multimodal depression detection on Instagram considering time interval of posts," Journal of Intelligent Information Systems, 2020.

[11] C. Lin, P. Hu, H. Su, S. Li, J. Mei, J. Zhou, and H. Leung, "SenseMood: Depression detection on social media," in Proc. 2020 Int. Conf. Multimedia Retrieval (ICMR), 2020, pp. 407–411.

[12] G. Gkotsis, A. Oellrich, S. Velupillai, M. Liakata, T. J. P. Hubbard, R. J. B. Dobson, and R. Dutta, "Characterisation of mental health conditions in social media using Informed Deep Learning," Scientific Reports, vol. 7, p. 45141, 2017, doi: 10.1038/srep45141.

[13] I. Sekulic and M. Strube, "Adapting deep learning methods for mental health prediction on social media," in Proc. 5th Workshop on Noisy User-generated Text (W-NUT 2019), 2019, pp. 322–327.

[14] V. Das Swain, J. Ye, S. K. Ramesh, A. Mondal, G. D. Abowd, and M. De Choudhury, "Leveraging social media to predict COVID-19-induced disruptions to mental well-being among university students: Modeling study," JMIR Formative Research, vol. 8, p. e52316, 2024, doi: 10.2196/52316.

[15] T. H. H. Aldhyani, S. N. Alsubari, A. S. Alshebami, H. Alkahtani, and Z. A. T. Ahmed, "Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models," Int. J. Environmental Research and Public Health, vol. 19, no. 19, p. 12635, 2022, doi: 10.3390/ijerph191912635.

[16] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of suicide ideation in social media forums using deep learning," Algorithms, vol. 13, no. 1, p. 7, 2020, doi: 10.3390/a13010007.

[17] S. Renjith, A. Sreekumar, and M. Jathavedan, "An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms," Journal of King Saud University – Computer and Information Sciences, vol. 33, no. 10, pp. 1379–1392, 2021.

[18] S. Neelakandan, M. Sridevi, S. Chandrasekaran, K. Murugeswari, A. K. S. Pundir, R. Sridevi, and T. B. Lingaiah, "Deep learning approaches for cyberbullying detection and classification on social media," Computational Intelligence and Neuroscience, vol. 2022, p. 2163458, 2022, doi: 10.1155/2022/2163458.

[19] S. Agrawal and A. Awekar, "Deep learning for detecting cyberbullying across multiple social media platforms," in Advances in Information Retrieval (ECIR 2018), Lecture Notes in Computer Science, vol. 10772, 2018, pp. 141–153.

[20] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," arXiv:1902.06673, 2019.

[21] X. Dong, U. Victor, S. Chowdhury, and L. Qian, "Deep two-path semi-supervised learning for fake news detection," arXiv:1906.05659, 2019.

[22] A. Roy, K. Basak, A. Ekbal, and P. Bhattacharyya, "A deep ensemble framework for fake news detection and classification," arXiv:1811.04670, 2018.

[23] S. Inamdar, R. Chapekar, S. Gite, and B. Pradhan, "Machine learning driven mental stress detection on Reddit posts using natural language processing," Human-Centric Intelligent Systems, vol. 3, no. 2, pp. 80–91, 2023, doi: 10.1007/s44230-023-00020-8.

[24] N. Oryngozha, P. Shamoi, and A. Igali, "Detection and analysis of stress-related posts in Reddit's academic communities," IEEE Access, vol. 12, pp. 14932–14948, 2024, doi: 10.1109/ACCESS.2024.3357662.

[25] C. Li and F. Li, "Emotion recognition of social media users based on deep learning," PeerJ Computer Science, vol. 9, p. e1414, 2023, doi: 10.7717/peerj-cs.1414.

[26] M. Owusu-Acheaw and A. G. Larson, "Use of social media and its impact on academic performance of tertiary institution students: A study of students of Koforidua Polytechnic, Ghana," Journal of Education and Practice, vol. 6, no. 6, pp. 94–101, 2015.

[27] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. 2019 Conf. North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 2019, pp. 4171–4186.

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems 30 (NeurIPS 2017), 2017, pp. 5998–6008.

[29] M. J. Hasan, S. H. Shifat, J. Matubber, R. Hossain, M. A. Rahman, B. M. T. Haque, and M. J. Hossen, "An in-depth exploration of machine learning methods for mental health state detection: A systematic review and analysis," Frontiers in Digital Health, vol. 7, p. 1724348, 2026, doi: 10.3389/fdgth.2025.1724348.

[30] Y. Cao, J. Dai, Z. Wang, Y. Zhang, X. Shen, Y. Liu, and Y. Tian, "Machine learning approaches for mental illness detection on social media: A systematic review of biases and methodological challenges," Journal of Behavioral Data Science, vol. 5, no. 1, 2024, arXiv:2410.16204.