An AI-Driven Hybrid Approach for Detecting Mental Health Indicators in Multilingual Indian Social Media: Data Acquisition and Analytical Frameworks

K. Alakananda
Department of Computer Science and Engineering, Sahyadri College of Engineering, Mangaluru, Affiliated to Visvesvaraya Technological University, Belagavi, India
alakanandakvidyan@gmail.com

Ananth G. Prabhu
Department of Computer Science and Engineering, Sahyadri College of Engineering, Mangaluru, Affiliated to Visvesvaraya Technological University, Belagavi, India
educatorananth@gmail.com (corresponding author)

K. M. Chaitra
Visvesvaraya Technological University, Belagavi, India
chaitrakm017@gmail.com

Mustafa Basthikodi
Department of Computer Science and Engineering, Sahyadri College of Engineering, Mangaluru, India
mustafa.cs@sahyadri.edu.in

Melwin D. Souza
Department of Computer Science and Engineering, Yenepoya Institute of Technology, Moodabidri, Karnataka, India
mellumerdy@gmail.com (corresponding author)

Received: 29 September 2025 | Revised: 19 October 2025, 31 October 2025, 8 December 2025, and 13 December 2025 | Accepted: 15 December 2025

 DOI: https://doi.org/10.48084/etasr.15214

ABSTRACT

Social networks have become the leading platform for human expression and have therefore been helpful in the early detection of psychological distress. This paper presents a multilingual framework to harvest and examine social media posts in a variety of regional Indian languages, such as Tamil, Telugu, Kannada, Malayalam, and English, to identify signs of different mental health disorders, such as depression, anxiety, and stress. The proposed method combines cutting-edge natural language processing and deep learning approaches, with special attention to exploiting Transformer models in conjunction with psychological lexicon-features, to disentangle complex linguistic and emotional patterns in multilingual text. A high-quality dataset, annotated by mental health experts to capture a variety of mental health signs, served as the basis for model training. The preprocessing pipeline addresses the challenges of multilingual, code-mixed, and transliterated text to provide uniform data quality across languages. The evaluation results show that the proposed hybrid model performs better than classical sentiment analysis approaches, achieving higher accuracy in identifying a variety of mental health signs. By identifying subtle emotional and linguistic signals in social media posts, this work illustrates the viability of scalable, real-world mental health monitoring across India's linguistic diversity. Future work includes the extension of language coverage, model interpretability through explainable AI, and real-time application to assist mental health clinicians in early detection and intervention.

Keywords-multilingual NLP; code-mixed text; lexicon features; mental health signal detection; expert-annotated

I. INTRODUCTION

Mental health disorders, such as depression, anxiety, and stress, remain critical global public health challenges that profoundly shape emotional stability, cognitive functioning, and social well-being. Although the importance of conventional clinical diagnostics—interviews, structured assessments, and self-report questionnaires—remains undisputed, these approaches face persistent challenges, particularly in regions with limited healthcare resources, high treatment costs, and the burden of stigma. Against this backdrop, social media has emerged as a dynamic data source where individuals openly articulate thoughts, emotions, and lived experiences. Unlike traditional methods, these digital traces provide continuous, large-scale, and real-time signals of psychological states, providing unique opportunities for early detection and intervention in mental healthcare.

The global academic landscape increasingly recognizes the potential of social media-based digital surveillance. However, current efforts remain disproportionately focused on English-language data, resulting in frameworks that are poorly adapted to multilingual contexts. With its rich linguistic diversity—including regional languages like Hindi, Tamil, Telugu, Kannada, Malayalam, Marathi, and Bengali—India's online environment exemplifies the need for inclusive digital health solutions. Social media posts in these languages frequently appear in native scripts, transliterated formats, or hybridized code-mixing patterns, further complicated by culturally specific ways of expressing psychological distress. Directly importing Western-centric models into this context risks both cultural misalignment and linguistic exclusion, reinforcing the digital divide in mental health monitoring.

A. Research Gaps

Despite advances in NLP and mental health informatics, key gaps persist. For example, limited high-quality annotated datasets that capture India's linguistic/cultural diversity restrict model robustness across languages. Preprocessing struggles with transliteration, code-mixing, and morphological complexity. Current systems overlook localized idioms and cultural distress expressions, reducing sensitivity. Ethical issues (privacy, consent, explainability) remain underexplored, while few frameworks enable real-time, scalable multilingual surveillance for public health impact.

B. Objectives

- Develop a multilingual NLP pipeline for native scripts, transliterations, and code-mixed content using a hybrid LSTM-Transformer (mBERT/XLM-R) architecture.
- Expand detection from depression to anxiety/stress with expert-guided annotation protocols for clinical/cultural validity.
- Deploy a scalable real-time system to support clinicians, researchers, and policymakers with privacy/ethical safeguards.

C. Literature Review

The integration of social media and mental health research has emerged very rapidly, creating new opportunities for early detection and intervention in linguistically diverse countries such as India.

Some of the issues in Indian languages have been identified through research, including the lack of access to annotated datasets, complex morphology, and the general usage of code-mixing and transliteration for online texts. Such issues are addressed through the development of custom-designed machine learning tools and language resources that account for these specific linguistic features [1]. Machine learning and deep learning models have been applied in more recent research to detect mental health signals in multilingual and mixed-linguistic social media posts. Diverse feature types, structural and non-structural, enhance the detection of depression in bilingual contexts, successfully capturing linguistic nuances. In [3], emphasis was placed on the identification of emotional cues embedded in regional language posts, as such signals can be significant predictors of mental health disorders. In [4], a multi-head attention model facilitated cross-lingual depression detection, interpreting the decision-making process across languages. In [5], these methods were combined to detect stress and depression with better accuracy and robustness on real-world data. In [6], a hybrid system bridged natural language processing and domain features, resulting in better detection performance in multilingual social media settings. In addition, language-agnostic speech analysis studies show the potential of models that can transfer knowledge from one linguistic and modality boundary to another to implement scalable mental health assistive tools [7].

Systematic reviews of existing tools and datasets indicate the need for culturally contextualized and language-agnostic techniques. In [8], an extended survey of multilingual mental health detection techniques emphasized the need to integrate heterogeneous linguistic features and cultural context. In [9], the pressing need for more abundant, annotated, and culturally contextualized resources for Indian languages was highlighted. In [10], early detection techniques were proposed, showing that early examination of social media updates can facilitate proactive mental health interventions. Similarly, a cross-cultural study [11] highlighted that the linguistic features of depression are heterogeneous across cultures, emphasizing the requirement for culturally contextualized models. In [12], a potential analysis of emotional features extracted from social media updates was presented to identify depression in regional languages, highlighting the need for fine-grained emotion analysis. In [13], contemporary approaches were described, highlighting the use of ethical guidelines to process sensitive material from social media to prevent misuse and respect people's rights. The study in [14] examined the potential and difficulties of conducting mental health studies in different languages on social media platforms, highlighting the importance of developing culturally relevant models and addressing the unavailability of annotated data.

In [15], machine translation was used for cross-lingual transfer learning to improve the accuracy of mental health detection in low-resource languages. In [16], targeted preprocessing and feature extraction techniques were proposed to counter the peculiarities of mixed texts, highlighting their efficacy for proper mental health evaluation in such instances. In [17], a deep neural network model was designed to detect depression in multilingual social media data, combining language-independent features with language-dependent ones and improving generalizability. In [18], existing explainable AI frameworks were described for the detection of mental health conditions in multilingual settings, which foster trust among users and clinicians. In [19], the ethics of social media-based mental health monitoring were outlined, focusing on concerns regarding privacy, consent, and cultural sensitivity for responsible and respectful use of social media data. The study in [20] presented an overview of real-time mental health monitoring systems for multilingual populations through social media stream analysis. This study outlined the potential for integrating state-of-the-art natural language processing techniques with large-scale live data analysis suites, enabling health authorities to act rapidly at the population level.

II. METHODOLOGY

This proposed method is designed to overcome the key challenges of limited annotated data, linguistic complexity, cultural nuance, transparency, and scalability in multilingual mental health detection from social media.

A. Data Acquisition and Annotation

1) Dataset Overview

All models were trained and evaluated on a curated 4,388-post multilingual dataset [21], detailed in Table I, comprising social media posts from Tamil (920), Telugu (900), Kannada (1,000), Malayalam (972), and English (596) sourced from Twitter/X, Facebook, and Instagram. The dataset maintains perfect class balance across depression, anxiety, stress, and neutral categories (1,097 posts each), with code-mixing prevalence ranging from 18.6% (English) to 32.1% (Malayalam), overall 27.4%, reflecting real-world Indian online linguistic diversity that includes native scripts, transliterations, and hybrids. A stratified 70:10:20 split (3,072 training/439 validation/877 test posts) preserved proportional language and class representation with three independent runs using fixed random seeds (42, 2023, 2024). Performance metrics report mean ± standard deviation alongside 95% Confidence Intervals (CI) from the t-distribution.

TABLE I. SUMMARIZES DATASET STATISTICS, LANGUAGE DISTRIBUTION, AND ANNOTATION QUALITY

Language	No. of posts	Depression	Anxiety	Stress	Neutral	Code-mixed (%)
Tamil	920	230	230	230	230	28.5
Telugu	900	225	225	225	225	31.2
Kannada	1000	250	250	250	250	26.8
Malayalam	972	243	243	243	243	32.1
English	596	149	149	149	149	18.6
Total	4388	1097	1097	1097	1097	27.4

2) Hybrid Annotation Strategy

Pre-labeling using lexicons and hashtags created initial labels, refined through a double-blind expert review by bilingual linguists and mental health specialists for cultural accuracy across Indian languages. Each post uses a multi-label vector representation:

$$y_i = [y_i, 1, \ldots, y_i, k], \quad y_{ij} \in \{0,1\} \tag{1}$$

where $y_{ij} = 1$ indicates the presence of mental health indicator j in post i. Annotation reliability was quantified using Cohen's k:

$$k = \frac{p_o - p_e}{1 - p_e} \tag{2}$$

where p_o is observed agreement and p_e is expected agreement by chance. Three independent annotators, per item, labeled all posts using a structured coding guide that defines linguistic, affective, and contextual cues for each category. Disagreements were resolved by consensus with a supervising clinician. This protocol yielded $k = 0.86$ overall, indicating almost perfect agreement under Landis-Koch, while preserving multi-label cases when criteria co-occurred and excluding mere topical mentions or sarcasm without corroborating self-state.

B. Preprocessing and Linguistic Normalisation

1) Text Cleaning and Standardisation

Posts undergo language identification to detect code-mixing, with native script detection via Unicode ranges to minimize short-text errors. Normalization applies canonical NFC, Indic-specific cleanup (nukta/virama harmonization, diacritic collapse), and standardizes platform elements: URLs → <URL>, mentions → <USER>, hashtags split with <HASHTAG> markers. Emojis convert to sentiment symbols (e.g., <EMO_SAD>). Romanized Indic text transliterates to native script via mapping tables or neural converters, preserving named entities. Code-mixed content receives token-level language tags for language-specific tokenization/ lemmatization. Indian-English colloquialisms and negation cues receive scope tagging (e.g., happy_NEG), while stop words are removed only for sparse features.

The preprocessing transformation for a post x is:

$$P_{norm}(x) = T(S(N(x))) \tag{3}$$

where $N(x)$ is normalization, $S(x)$ is stemming lemmatization, and $T(x)$ is tokenization.

C. Feature Engineering

1) Linguistic and Statistical Features

Term Frequency-Inverse Document Frequency (TF-IDF) measures the importance of a word in a given document concerning its frequency in an entire dataset. It is mathematically defined as:

$$TF - IDF(t,d) = TF(t,d) \times \log\left(\frac{N}{DF(t)}\right) \tag{4}$$

where $TF(t,d)$ is term frequency, $DF(t)$ is document frequency, and N is the total number of documents. This approach combines the word's frequency in the document with

its inverse frequency in all the documents, so that the model can give greater weights to words that are more informative or different for a specific document. Part-Of-Speech (POS) tagging labels words into grammatical categories such that patterns of syntax that can be linked to mental health markers can be recognized. Sentiment analysis predicts text emotional tone based on rule-based dictionaries and machine learning and provides a fine-grained analysis of affective expression.

2) Psychological and Emotional Features

Affective features combine domain lexicons with multilingual projections to cover Indic languages; when native resources are unavailable, high-quality translations are employed and verified on sampled entries. Negation is retained via scoped tags to prevent polarity flips from being washed out by tokenization and lemmatization. Topic models on normalized text reveal latent, culturally grounded themes and frequent code mix patterns that sharpen sensitivity to distress expressions beyond literal sentiment. These lexicon- and topic-derived signals are fused with contextual embeddings from multilingual Transformers prepared on minimally altered text with explicit code mix markers, yielding a representation resilient to transliteration, orthographic variation, and colloquial usage.

D. Model Architecture

1) Hybrid Deep Learning Framework

The proposed model architecture employs a hybrid approach where both sequential and context analysis are employed in multilingual depression detection. Input text is tokenized, embedded, and then processed in parallel with an LSTM network. The following algorithm outlines the step-by-step process of the proposed Hybrid LSTM-Transformer model that leverages both sequential and contextual information for effective multilingual depression detection on social media texts.

```
Algorithm 1: Hybrid LSTM-Transformer for
Multilingual Depression Detection
Input:
  Multilingual social media posts
  X = {x_1, x_2, ..., x_n} tokenized and preprocessed
Output:
  Predicted mental health labels Y = {y_1, y_2, ..., y_n}
  for each post
Preprocessing: For each post x_i, apply
  normalization, transliteration, tokenization,
  and language tagging to handle code-mixed and
  noisy text.
Embedding: Convert tokens of x_i into dense
  vector representations using multilingual
  embeddings (e.g., mBERT/XLM-R embeddings).
Sequential Modeling (LSTM): Pass embeddings
  through an LSTM network to capture sequential
  dependencies: h_t = LSTM(x_t, h_{t-1})
Contextual Modeling (Transformer):
  Simultaneously pass embeddings through a
  Transformer encoder to capture contextual
  relationships via self-attention:
  Context = Transformer(X)
Feature Fusion: Concatenate the final hidden
  state of LSTM h_T and Transformer output
  Context: F = [h_T; Context]
Classification: Apply fully connected layers
  and a softmax function on fused features for
  final prediction: y_i = softmax(WF + b)
Output: Return predicted labels y_i indicating
  mental health indicators (depression,
  anxiety, stress, neutral) for each post.
```

- Sequential Modeling (LSTM): The LSTM architecture processes input sequences by updating hidden states through gated mechanisms that selectively retain or forget information from previous time steps, enabling the capture of long-term dependencies in social media text. LSTM Hidden State Update, h_t, at time t is defined as:

$$h_t = \sigma(W_h \times [h_{t-1}, x_t] + b_h) \quad (5)$$

where x_t is the input token, W_h and b_h denote the weights and the bias parameters, and σ is the activation function. This complete definition ensures clear mapping between sequence input and temporal dependency learning.

- Contextual Modeling (Transformer): The self-attention mechanism computes weighted relationships between all words in a sentence by calculating attention scores through query-key interactions, allowing the model to understand contextual meaning regardless of word position. To reflect attention weighting and value aggregation, Transformer Self-Attention is defined as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (6)$$

where Q, K, and V are the query, key, and value matrices, respectively, and d_k is the key dimension used for normalization.

- Feature Fusion and Classification: The final classification layer combines the sequential representations from LSTM with the contextual embeddings from the Transformer through concatenation, and then applies a softmax function to predict mental health indicators.

$$P\langle y|x\rangle = softmax(W_c \times [h_T; TR(x)] + b_c) \quad (7)$$

where h_T is the final LSTM state, $TR(x)$ is the Transformer output, and W_c and b_c are classifier weights and bias.

E. Model Training and Evaluation

1) Experimental Setup

All experiments were implemented in TensorFlow 2.15 and PyTorch 2.3, with fixed random seeds for reproducibility. Each model's performance was averaged over three independent runs to account for training variance.

2) Training

The model is trained by minimizing the cross-entropy loss, which measures how far the predicted probabilities are from the true class labels for every category. The cross-entropy loss function may be written as:

$$L = \sum_{i=1}^{C} y_i \log(\hat{y}_i) \quad (8)$$

where y_i represents the true label, $\hat{y}_i$ is the predicted probability for class i, and C is the number of classes. To facilitate efficient and adaptive learning, the Adam optimizer is employed, enabling the model to adjust its learning rates dynamically and converge more effectively during training.

Performance was evaluated based on accuracy, precision, recall, F1-score, and AUC-ROC. The results of the hybrid model are compared with baseline classifiers and deep learning models in isolation.

F. Real-Time and Scalable Deployment

The system runs continuously to analyze Indian-language social media in real time, ingesting Twitter and Facebook streams via scalable queues for high-throughput multilingual processing. Incoming posts undergo instant language ID, transliteration, and normalization before analysis to standardize noisy, code-mixed inputs. A hybrid deep learning stack couples LSTM sequence modeling with Transformer-based context encoding to detect mental health signals from text. The models are refreshed through periodic batch retraining and online updates, keeping pace with evolving vocabulary and expressions. Outputs feed a live dashboard that surfaces trends and flags early warning indicators for rapid clinical response. Figure 1 illustrates the end-to-end streaming pipeline.

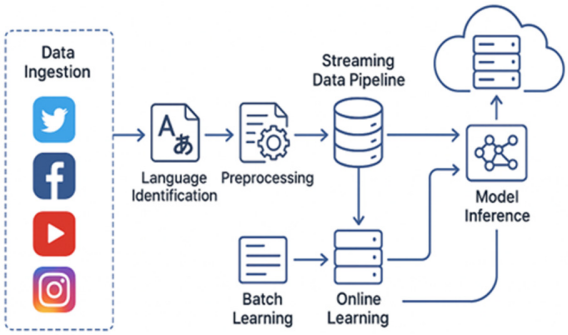

Fig. 1. End-to-end social media data pipeline: From ingestion to real-time model inference

G. Ethical and Cultural Considerations

All data are anonymized and collected in accordance with platform policies and established research ethics, with annotation and modelling informed by local language and cultural norms to ensure contextual validity. This approach pairs expert manual labeling with machine translation to expand scarce labeled datasets and improve robustness, while multilingual preprocessing tailored to code-mixing and transliteration reduces noise and increases accuracy. Culturally aware, affective, and lexical features, combined with advanced feature engineering, help surface subtle, context-specific mental health markers that generic models miss. Privacy is protected through dedicated safeguards alongside interpretable modeling, and the system scales for continuous real-time monitoring across languages and communities to support ongoing mental health surveillance and timely intervention.

Significance was tested via paired, two-tailed t-tests comparing Hybrid vs. each baseline on run level macro F1 and ROC AUC (n = 3; *seeds*: 42, 2023, 2024; df = 2). In addition, per-language (Tamil, Telugu, Kannada, Malayalam, English) macro-F1 tests were run with Benjamini-Hochberg FDR control (q = 0.05). Effect sizes were Cohen's d for paired samples. The t distribution used 95% CIs for mean differences.

III. RESULTS

This section presents a comprehensive empirical validation of the proposed hybrid LSTM-Transformer model for multilingual mental health detection in Indian social media posts. The results demonstrate consistent superiority over established baselines. The evaluation encompasses traditional machine learning approaches alongside advanced deep learning architectures, highlighting the robustness of the proposed model in handling code-mixed and transliterated content.

A. Overall Model Performance

Table II summarizes the performance of five baselines—Naïve Bayes, SVM, CNN, and Transformer models (mBERT/XLM R)—against the proposed hybrid LSTM Transformer. Naïve Bayes uses a multinomial variant with TF-IDF features (maximum 5,000 terms, smoothing α=1.0), while the SVM baseline applies a linear kernel with TF-IDF features (maximum 5,000 terms) and *class_weight* set to balanced. The CNN model operates at the sentence level with 300-dimensional embeddings, two convolutional layers (128 and 256 filters with kernel sizes 3 and 4), max pooling, a 512-unit dense layer, and dropout of 0.5. The transformer baseline relies on base mBERT/XLM R configurations (12 layers, 768 hidden units, maximum sequence length of 256), fine-tuned with AdamW at a learning rate of 2×10^{-5}. The hybrid LSTM Transformer fuses a 256-unit LSTM over token embeddings with mBERT encoder outputs, trained with AdamW (encoder learning rate 2×10^{-5}, LSTM/classifier learning rate 1×10^{-3}), batch size 64, and early stopping on validation macro F1. The proposed hybrid architecture achieved the best overall results, with 91.5% accuracy, 89.8% macro F1 score, and 0.94 AUC-ROC, clearly outperforming the baselines. By combining sequential modeling from LSTM with the contextual representations of Transformers, the proposed hybrid model handles multilingual and code-mixed text more robustly than any single architecture.

TABLE II. COMPARATIVE PERFORMANCE WITH VARIABILITY (MEAN±SD OVER 3 SEEDS)

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
NaïveBayes	78.2±0.6 [77.5, 78.9]	76.5±0.7 [75.6, 77.4]	74.8±0.8 [73.8, 75.8]	75.6±0.7 [74.7, 76.5]	0.81± 0.01 [0.80, 0.82]
SVM	81.5±0.5 [81.0, 82.0]	79.9±0.6 [79.1, 80.7]	78.3±0.6 [77.5, 79.1]	79.1±0.5 [78.5, 79.7]	0.84±0.01 [0.83, 0.85]
CNN	83.2±0.6 [82.4, 84.0]	81.6±0.6 [80.8, 82.4]	80.5 ± 0.6 [79.7, 81.3]	81.0 ± 0.5 [80.4, 81.6]	0.86±0.01 [0.85, 0.87]
Transformer (mBERT/XLM-R)	88.3±0.4 [87.9, 88.7]	87.0±0.5 [86.4, 87.6]	86.2 ± 0.5 [85.6, 86.8]	86.6 ± 0.4 [86.2, 87.0]	0.91±0.01 [0.90, 0.92]
Hybrid (proposed)	91.5±0.4 [91.1, 91.9]	90.2±0.4 [89.8, 90.6]	89.5 ± 0.4 [89.1, 89.9]	89.8 ± 0.4 [89.4, 90.2]	0.94±0.01 [0.93, 0.95]

1) Model Performance Across Languages

Figure 2 reports macro F1 scores per language for the main deep learning models. Across Tamil, Telugu, Kannada, Malayalam, and English, the hybrid LSTM Transformer consistently achieved the highest F1 score, with particularly strong gains in Tamil and Malayalam, where code mixing and transliteration are most frequent. This pattern indicates that joint sequential and contextual modeling is especially beneficial in highly variable linguistic settings.

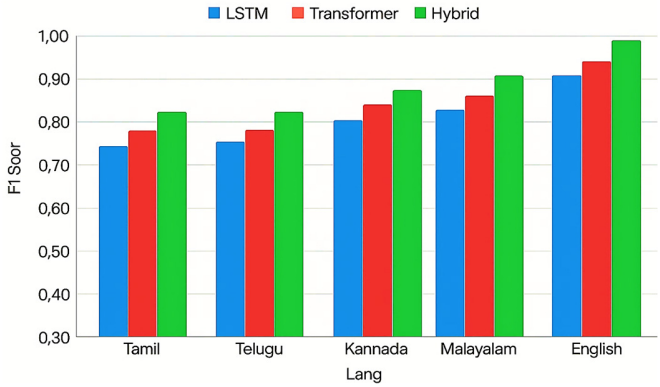

Fig. 2. F1-scores by model and language,

2) Distribution of Detected Mental Health Indicators

Figure 3 illustrates the proportion of posts flagged as depression, anxiety, and stress by the hybrid model. Depression accounts for the largest share, while anxiety and stress appear less frequently, suggesting that depressive cues are expressed more overtly online and that anxiety and stress are either rarer or linguistically subtler. These findings underscore the need for detectors tuned to culturally specific ways of expressing distress across Indian languages.

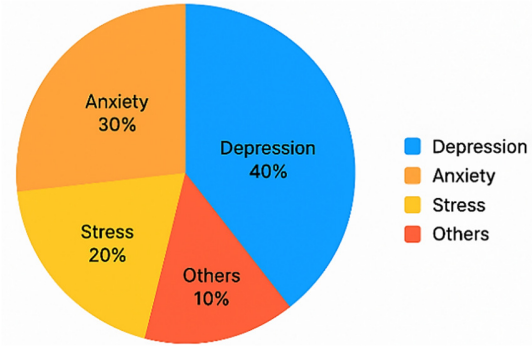

Fig. 3. Mental health indicators' share.

3) Temporal Trends in Mental Health Discourse

Figure 4 shows monthly trends in detected depression, anxiety, and stress over one year. The overall volume increases steadily, with peaks aligning with major real-world events such as exams, lockdowns, and crises, indicating that offline stressors quickly surface in online discourse. Such temporal patterns support the use of real-time multilingual monitoring to anticipate risk periods and guide targeted outreach.

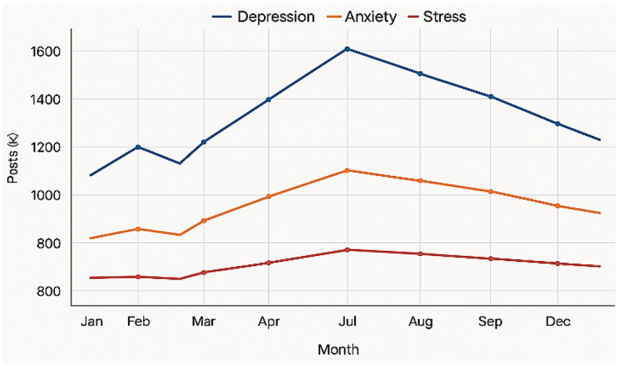

Fig. 4. Mental health posts trend.

B. Comparative Model Performance

Figure 5 visualizes the comparative performance of Naïve Bayes, SVM, CNN, Transformer, and the hybrid LSTM Transformer using accuracy, macro F1 score, and AUC-ROC. The hybrid model achieves 3–5 percentage points higher macro F1 than Transformers and 10–13 percentage points higher than traditional machine learning baselines, confirming the quantitative advantage already summarized in Table II.

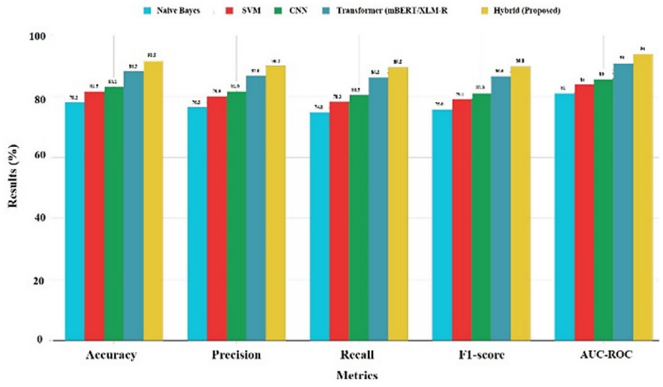

Fig. 5. Performance comparison of mental health detection models.

Receiver Operating Characteristic (ROC) curves in Figure 6 highlight the discriminative capacity of each model. The proposed hybrid model's AUC-ROC of 0.94 indicates excellent separation between posts that do and do not contain mental health indicators, which is especially valuable for correctly handling borderline or ambiguous cases in practice.

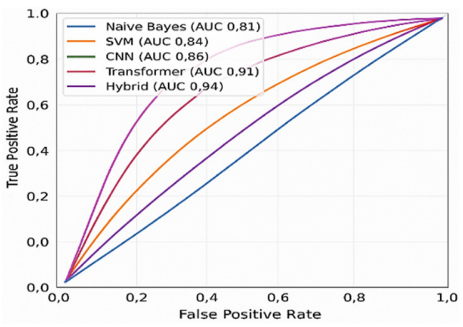

Fig. 6. ROC curves for depression detection models.

C. Class-wise Error Patterns

Table III reports the true versus predicted counts for depression, anxiety, and stress for the Naïve Bayes, SVM, CNN, Transformer, and hybrid LSTM Transformer models. Naïve Bayes shows substantial leakage from anxiety and stress into depression, whereas SVM reduces some of these misclassifications but still exhibits considerable confusion between anxiety and stress. CNN increases correct detections for all three classes, yet maintains noticeable overlap between anxiety and stress, highlighting the limits of purely convolutional filters. In contrast, the Transformer and especially the hybrid LSTM Transformer attain the strongest diagonal concentrations and the fewest off-diagonal entries, indicating a more reliable separation of depression, anxiety, and stress in multilingual social media posts.

TABLE III. CLASS-WISE PREDICTION COUNTS FOR MENTAL HEALTH DETECTION MODELS

Model	True Class	Predicted Depression	Predicted Anxiety	Predicted Stress
Naïve Bayes	Depression	120	30	25
	Anxiety	30	90	35
	Stress	22	27	70
SVM	Depression	135	20	15
	Anxiety	20	110	30
	Stress	15	20	80
CNN	Depression	145	10	15
	Anxiety	15	120	25
	Stress	10	15	90
Transformer (mBERT/XLM-R)	Depression	150	6	5
	Anxiety	10	130	8
	Stress	4	7	125
Hybrid LSTM-Transformer	Depression	170	5	3
	Anxiety	7	145	5
	Stress	2	4	135

D. Statistical Validation

Paired two-tailed t-tests over three runs confirm that the hybrid model's gains in macro F1 and AUC-ROC over each baseline are statistically significant at $p < 0.01$. Per language tests with Benjamini-Hochberg false discovery rate control show significant improvements in Tamil, Kannada, and Malayalam, with English and Telugu differences not reaching significance due to lower code mixing intensity and smaller effect sizes. These results support the robustness and practical relevance of the proposed hybrid architecture for real-world multilingual mental health detection.

IV. CONCLUSIONS

This study presents a robust multilingual model for identifying mental health indicators, i.e., depression, anxiety, and stress, in Indian social media updates based on contemporary natural language processing and deep learning frameworks to tackle the linguistic and cultural diversity in the country. The proposed LSTM-Transformer-based model outperformed traditional machine learning and individual deep learning approaches, with robust accuracy and well-balanced classification in various Indian languages and code-mixing updates. With the integration of sequential and contextual modeling, the system can capture nuanced linguistic and affect cues, which allows for more precise detection of psychological distress across diverse online communities. The results highlight the importance of culturally and linguistically responsive AI systems for real-time mental health monitoring, with enhanced detection accuracy. This study also highlights the necessity for the development of large annotated datasets and the integration of ethical considerations, such as privacy, consent, and transparency, in the use of digital mental health technologies in sensitive contexts. Future research must expand language coverage, improve model explainability, and create scalable, real-time systems capable of supporting early intervention and mental health care among India's diverse population.

REFERENCES

[1] M. Garg, "Mental Health Analysis in Social Media Posts: A Survey," *Archives of Computational Methods in Engineering*, vol. 30, no. 3, pp. 1819–1842, Apr. 2023, https://doi.org/10.1007/s11831-022-09863-z.

[2] F. Rehmani, Q. Shaheen, M. Anwar, M. Faheem, and S. S. Bhatti, "Depression detection with machine learning of structural and non-structural dual languages," *Healthcare Technology Letters*, vol. 11, no. 4, pp. 218–226, Aug. 2024, https://doi.org/10.1049/htl2.12088.

[3] P. Ta, N. Tran, H. Nguyen, and H. D. Nguyen, "Detecting signs of depression on social media: A machine learning analysis and evaluation," *Sustainable Futures*, vol. 10, Dec. 2025, Art. no. 100827, https://doi.org/10.1016/j.sftr.2025.100827.

[4] V. Vajrobol, N. Aggarwal, U. Shukla, G. J. Saxena, S. Singh, and A. Pundir, "Explainable cross-lingual depression identification based on multi-head attention networks in Thai context," *International Journal of Information Technology*, vol. 17, no. 5, pp. 2997–3012, June 2025, https://doi.org/10.1007/s41870-023-01512-3.

[5] V. Tejaswini, K. S. Babu, and B. Sahoo, "Depression Detection from Social Media Text Analysis using Natural Language Processing Techniques and Hybrid Deep Learning Model," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 1, pp. 1–20, Jan. 2024, https://doi.org/10.1145/3569580.

[6] Vandana, N. Marriwala, and D. Chaudhary, "A hybrid model for depression detection using deep learning," *Measurement: Sensors*, vol. 25, Feb. 2023, Art. no. 100587, https://doi.org/10.1016/j.measen.2022.100587.

[7] K. Daly and O. Olukoya, "Depression detection in read and spontaneous speech: A Multimodal approach for lesser-resourced languages," *Biomedical Signal Processing and Control*, vol. 108, Oct. 2025, Art. no. 107959, https://doi.org/10.1016/j.bspc.2025.107959.

[8] A. Khan and R. Ali, "Unraveling minds in the digital era: a review on mapping mental health disorders through machine learning techniques using online social media," *Social Network Analysis and Mining*, vol. 14, no. 1, Apr. 2024, Art. no. 78, https://doi.org/10.1007/s13278-024-01205-0.

[9] S. T. Ibrahim, M. Li, J. Patel, and T. R. Katapally, "Utilizing natural language processing for precision prevention of mental health disorders among youth: A systematic review," *Computers in Biology and Medicine*, vol. 188, Apr. 2025, Art. no. 109859, https://doi.org/10.1016/j.compbiomed.2025.109859.

[10] Y. Cao *et al.*, "Machine Learning Approaches for Depression Detection on Social Media: A Systematic Review of Biases and Methodological Challenges.," *Journal of Behavioral Data Science*, vol. 5, no. 1, pp. 67–102, Feb. 2025, https://doi.org/10.35566/jbds/caoyc.

[11] T. Amorese *et al.*, "Detecting depression in speech using verbal behavior analysis: a cross-cultural study," *Frontiers in Psychology*, vol. 16, May 2025, Art. no. 1514918, https://doi.org/10.3389/fpsyg.2025.1514918.

[12] M. K. Myee, R. D. C. Rebekah, T. Deepa, G. D. Zion, and K. Lokesh, "Detection of Depression in Social Media Posts using Emotional Intensity Analysis," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16207–16211, Oct. 2024, https://doi.org/10.48084/etasr.7461.

[13] B. G. Teferra *et al.*, "Screening for Depression Using Natural Language Processing: Literature Review," *Interactive Journal of Medical Research*, vol. 13, Nov. 2024, Art. no. e55067, https://doi.org/10.2196/55067.

[14] M. E. Aragón, A. P. López-Monroy, M. Montes-y-Gómez, and D. E. Losada, "Adapting language models for mental health analysis on social media," *Artificial Intelligence in Medicine*, vol. 168, Oct. 2025, Art. no. 103217, https://doi.org/10.1016/j.artmed.2025.103217.

[15] X. Shi, X. Liu, C. Xu, Y. Huang, F. Chen, and S. Zhu, "Cross-lingual offensive speech identification with transfer learning for low-resource languages," *Computers and Electrical Engineering*, vol. 101, July 2022, Art. no. 108005, https://doi.org/10.1016/j.compeleceng.2022.108005.

[16] M. Kanahuati-Ceballos and L. J. Valdivia, "Detection of depressive comments on social media using RNN, LSTM, and random forest: comparison and optimization," *Social Network Analysis and Mining*, vol. 14, no. 1, Feb. 2024, Art. no. 44, https://doi.org/10.1007/s13278-024-01206-z.

[17] W. B. Tahir, S. Khalid, S. Almutairi, M. Abohashrh, S. A. Memon, and J. Khan, "Depression Detection in Social Media: A Comprehensive Review of Machine Learning and Deep Learning Techniques," *IEEE Access*, vol. 13, pp. 12789–12818, 2025, https://doi.org/10.1109/ACCESS.2025.3530862.

[18] T. S. Kumar, "A Deep Learning Framework with a Hybrid Model for Automatic Depression Detection in Social Media Posts," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 4, pp. 3217–3231, June 2024.

[19] A. Gupta and R. Katarya, "Social media based surveillance systems for healthcare using machine learning: A systematic review," *Journal of Biomedical Informatics*, vol. 108, Aug. 2020, Art. no. 103500, https://doi.org/10.1016/j.jbi.2020.103500.

[20] A. Montejo-Ráez, M. D. Molina-González, S. M. Jiménez-Zafra, M. Á. García-Cumbreras, and L. J. García-López, "A survey on detecting mental disorders with natural language processing: Literature review, trends and challenges," *Computer Science Review*, vol. 53, Aug. 2024, Art. no. 100654, https://doi.org/10.1016/j.cosrev.2024.100654.

[21] "South Indian languages social Media Posts Dataset · Issue #1." GitHub, [Online]. Available: https://github.com/melwin-boop/South-indian-Language-social-Media-posts-/issues/1.