A Personality-Aware Hybrid Deep Learning Framework for Anxiety and Depression Detection Using a Neuro-Temporal Sparse Modulation Network

Raksha Rajanna
Department of Computer Science and Engineering, SJCE, JSS Science and Technology University, Mysuru, India
raksha@jssstuniv.in (corresponding author)

Mullur Puttubuddhi Pushpalatha
Department of Computer Science and Engineering, SJCE, JSS Science and Technology University, Mysuru, India
mppvin@jssstuniv.in

Impana Kamalamma Puttaraju
Department of Computer Science and Engineering, JSS Academy of Technical Education, Bangalore, India
impanaraj@jssateb.ac.in

Received: 12 January 2026 | Revised: 7 February 2026 and 13 February 2026 | Accepted: 16 February 2026

 DOI: https://doi.org/10.48084/etasr.17502

ABSTRACT

The early detection of depressive and anxiety disorders is still an active area of research due to the differences between people and the change in time of signs that people show. This study presents a potential advancement towards more accurate identification through a multimodal hybrid model, named Neuro-Temporal Sparse Modulation Network (NTSM-Net). NTSM-Net consists of neuro-inspired sparse encoding, a Bayesian Temporal Transformer for uncertainty-aware temporal modeling, a Behavioral Micro Event Detector to obtain accurate cues, and the application of personality-aware modulation to obtain individualized inferences. NTSM-Net produces trajectories of emotion and estimates of the potential severity of the risk rather than diagnosing someone clinically. NTSM-Net was evaluated using five publicly available multimodal datasets, yielding 96.8% and 96.9% accuracy and F1-scores, respectively, and outperforming recent models. The results of the ablation and robustness analyses determine the contribution of neuro-inspired sparse encoding, probabilistic temporal modeling, behavioral micro-event detection, and personality modulation to the clarification of individuals with depressive/anxiety-related disorders.

Keywords-neuro-inspired sparse coding; Bayesian temporal modelling; personality-aware modulation; micro-event detection; hybrid deep learning; behavioral signal modeling

I. INTRODUCTION

Individuals with mental conditions such as depression and anxiety can show many different kinds of behavior that can change over time. Artificial intelligence methods can analyze large datasets collected from social media and multimodal sources, helping to determine when someone may be experiencing a mental health crisis [1]. Machine learning and deep learning techniques are being used more frequently to help clinicians make informed decisions about diagnoses, while many clinicians allow automated screening to occur outside of their office setting [2]. In [3], it was reported that due to their resemblance to biological learning mechanisms, spiking neural systems produce sparse coding and decorrelation without any means of bias correction, indicating that representations based on sparsity form the basis for robust perceptual systems. Recent studies have started to investigate deep learning architectures for classifying depression, but many do not consider how the neurobiology of cognitive interpretation and emotional expression is related to the development of a classifier for depressive disorders [4].

In [5], a Bayesian Variational Transformer was employed to improve generalization when an item is subjected to operationalized noisy conditions. The benefits of this approach were demonstrated in the identification of equipment defects in the manufacturing process. A significant degree of interest has been paid to the identification of micro-expressions due to the use of HTNet [6], which identifies short-lived emotional indicators based on detailed descriptions of how facial expressions change. Speech-derived personality embeddings [7] are psychologically meaningful ways to conceptualize a person's psychological make-up, while multimedia-based tutoring systems [8] take personality into account to create adaptive human-AI interactions. In [9], a framework was presented to provide sustained clinical monitoring of patients with personalized and multimodal multitask learning, demonstrating that personalized models are essential for effective clinical monitoring systems.

The data collected within micro-events [10] can be used to create trajectories of emotional dynamics. Recently introduced multimodal frameworks for detecting depression incorporate attention-based graph convolution and transformers [11], which are capable of achieving state-of-the-art classification results. This paper presents the development of computational algorithms to assess anxiety and depression utilizing various data streams, such as affective, physiological, and behavioral. According to affective neuroscience and the literature on mental health informatics, symptoms of anxiety and depression can be defined by considering specific patterns of sustained negative affect, increased stress response, decreased variability of emotional state, temporal dysregulation of physiological signals, and so on. The proposed NTSM-Net model can identify anxiety and depression by utilizing observable multimodal data as indicators of the presence and severity of anxiety/depression-level states.

II. RELATED WORKS

Recently published literature indicates that there is rapid progress in micro-expression analysis, multimodal depression detection, modeling involving personality factors, sparse neural representations, and uncertainty-aware deep learning. In [12], one of the largest bibliometric surveys ever conducted on the subject of micro-expression recognition was presented, while another large survey on micro-expression spotting and recognition methodology was presented in [13]. In [7], a method to predict personality traits through deep learning merged acoustic and linguistic embeddings. Similarly, in [14], FMFN (Fuzzy Multimodal Fusion Network) used a fuzzy fusion method to add ambiguity to ensemble-based emotion recognition. Unfortunately, this method relied on a generic fusion mechanism with no modeling of either temporal uncertainty or neuro-inspired sparsity. In [15], a dual-modality fusion framework for depression recognition combined speech and text modalities, but was not capable of modeling fine-grained micro-events or probabilistic reasoning related to timing. In [16], a voice pretraining model was presented for the identification of depression, providing further evidence of the strength of using large-scale self-supervised learning to identify mental health signals. In [17], a multimodal fusion method used audio, text, and visual modalities to detect depression. In [18], the first performance evaluation framework to assess emotional states used a model that considered the temporal nature of an evolving emotional state.

In [19], a comprehensive review of machine learning and deep learning approaches for detecting mental diseases was presented. In [20], biologically inspired dynamic sparsity was proposed for energy-efficient perception. In [21], additional research was conducted on adaptive locally competitive sparse coding applied to speech classification. In [22], an audio-based stacking ensemble model was developed to detect depression using diverse models to achieve competitive prediction accuracy. In [23], deep learning models used uncertainty estimation for EEG prediction, demonstrating that calibrating uncertainty improves trustworthiness in neurological applications. In [24], a Bayesian Cooperative Probabilistic Transformer for Remaining Useful Life Predictions was developed. In [25], intelligent digital methods to screen patients for dementia risk used multimodal data and Explainable AI (XAI). The survey in [26] detailed various technological advances in the field of facial expression recognition. In [27], an AU-based micro-expression recognition system could detect driver emotion through micro-physical cues. In [28], a Multi-modal Stress Detection Dataset was presented, composed of both facial and physiological signals. In [29], a compact and efficient framework was developed to predict depression through multiple modalities, demonstrating competitive performance.

Through these studies, several important limitations can be identified, such as no modelling of micro-events/ultra-short temporal cues, very limited use of personality modulated signals, no probabilistic temporal reasoning/Bayesian uncertainty, limited use of sparse neuro-inspired encoders for fine-grained affect signals, and little to no analysis of trajectories to determine severity. The proposed approach aimed to directly address these limitations through the combination of sparse neuro-inspired encoders, Bayesian temporal modelling, personality modulated signals, and micro-event detection within a single diagnostic framework.

III. PROPOSED MODEL

This section describes a new deep learning model, called the Neuro-Temporal Sparse Modulation Network (NTSM-Net), that combines various approaches to analyze different aspects of human behavior in order to identify anxiety and depression. This framework introduces key elements as shown below and in Figure 1.

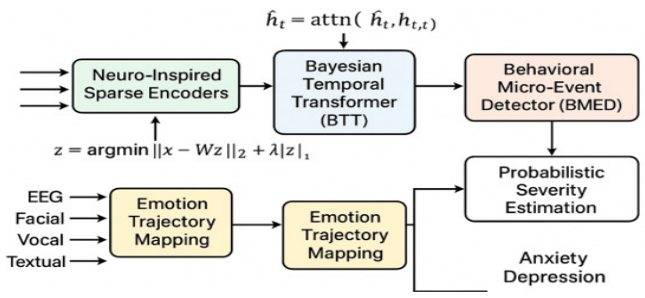

Fig. 1. Design diagram of the proposed NTSM-Net.

A. Neuro-Inspired Sparse Encoders

The first element, neuro-inspired sparse encoders, implements biologically based encoding methods for representing multiple types of signals with minimal energy demand. Neuro-inspired sparse encoders take inspiration from natural systems' cortical sparse coding methods, thereby enforcing local competition among neurons. Using an input audio/visual feature vector x_m, the sparse encoding application is solved as:

$$\alpha_m^* = \arg\min_{\alpha_m} \| x_m - D_m\alpha_m \|_2^2 + \lambda \| \alpha_m \|_1 \quad (1)$$

Here, the learned overcomplete dictionary is represented by the matrix $D_m \in R^{d\times k}$, while the activation vector α_m is sparse, and λ provides a level of control over its sparsity. These elements generate very selective and noise-resistant micro-features (encodings) that are essential for detecting subtle signals that indicate depression or anxiety. The locally-competitive mechanism is described as:

$$\alpha_m^{(t+1)} = \text{ReLU}\left(\alpha_m^{(t)} + \eta\left(D_m^\top x_m - D_m^\top D_m \alpha_m^{(t)} - \lambda\right)\right) \quad (2)$$

This enables the feature selection process to be stabilized through a neuro-inspired pattern of activation.

B. Behavioral Micro-Event Detector (BMED)

The Micro-Event Detector is designed to identify very specific fine-grain, short-duration behavioral indicators that are associated with anxiety and depression, such as micro-vocal tremors, rapid speech pauses, micro-facial muscular activations, or subtle cognitive hesitation during text messaging. Micro-events include micro-expressions, a flick of the eye or gaze, micro-tremors, pauses, and bursts of delay/hesitation during conversations. For sparse temporal sequences $\alpha_m(t)$, the scoring for a micro-event is given by:

$$e_t = \sigma(W_e[\Delta\alpha_m(t); \nabla^2\alpha_m(t)] + b_e) \quad (3)$$

Here, $\Delta\alpha_m(t)$ represents the first derivative of time or onset and cutoff, $\nabla^2\alpha_m(t)$ represents the second derivative or acceleration (micro-tremor), and W_e and b_e are trainable parameters. Micro-event weighting protects against temporal encoding as per:

$$\tilde{\alpha}_m(t) = e_t \odot \alpha_m(t) \quad (4)$$

C. Bayesian Temporal Transformer (BTT)

The Bayesian Temporal Transformer (BTT) offers a model capable of modeling uncertainty throughout time and analyzing changes in the emotional state over time. To model emotional fluctuation and uncertainty, NTSM-Net leverages a Bayesian approach to multi-head self-attention given Q, K, V from the sparse encoders as shown in:

$$\mu_A = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right) \quad (5)$$

$$\sigma_A = \exp(W_s(Q \odot K) + b_s) \quad (6)$$

The stochastic attention is sampled as:

$$A = \mu_A + \epsilon \cdot \sigma_A, \epsilon \sim \mathcal{N}(0, I) \quad (7)$$

The Transformer output of the model is given by:

$$H = AV \quad (8)$$

This model captures the indecisiveness, uncertainty, and unchanging behavior typical of those suffering from anxiety/depression.

D. Personality-Aware Modulation (PAM) Layer

The PAM layer allows each user to be treated as an individual with specific behavioral tendencies due to its five personality traits. The model is environmentally valid and consequently helpful to support NTSM-Net's potential role in providing mental health services in a more real-world context. The PAM layer integrates with the personality embeddings as:

$$p = W_p z + b_p \quad (9)$$

The internal emotional representation H is calculated to modulate using:

$$\widehat{H} = \gamma(p) \odot H + \beta(p) \quad (10)$$

$$\gamma(p) = W_\gamma p \quad (11)$$

$$\beta(p) = W_\beta p \quad (12)$$

These equations allow for individual variability in the amount of feature condensation and shifting of features, reducing the possibility that an introvert may be incorrectly classified as depressed due to personal traits.

E. Emotion Trajectory Mapping

Emotion trajectory mapping accurately maps a continuous affective timeline of how emotional states progress over utterances, segments of text, and micro-expressions. The shape, slope, or stability of the curve illustrates the way in which the aforementioned components convert multimodal signals into meaningful behavior, thus bridging the gap between deep neural representations and clinical temporal insight. The trajectory is calculated using:

$$\tau(t) = \text{GRU}\left(\widehat{H}(t), \tau(t-1)\right) \quad (13)$$

The trajectory $\tau(t)$ indicates emotional stability or volatility. NTSM-Net's Probabilistic Severity Estimation is the last step in decision-making for a clinician. The formulation of severity as a distribution rather than as a single point allows us to provide a greater range of potential severity scores using:

$$y_{\text{pred}} \sim \mathcal{N}(\mu_y, \sigma_y^2) \quad (14)$$

$$\mu_y = W_\mu\tau(T) + b_\mu, \sigma_y = \text{softplus}(W_\sigma\tau(T) + b_\sigma) \quad (15)$$

Final anxiety/depression classification is computed using the activation function and other parameters of including trajectory using:

$$\hat{c} = \arg\max \text{softmax}(W_c\tau(T) + b_c) \quad (16)$$

The reconstruction loss measured with (17) is defined for use within Neuro-Inspired Sparse Encoder architectures.

$$\mathcal{L}_{rec} = \sum_m \| x_m - D_m\alpha_m \|_2^2 \quad (17)$$

To maximize the probability that an output prediction is correct, this method tries to penalize the model for assigning high confidence values to incorrect output states. The Kullback-Leibler (KL) divergence (18) indicates how far apart two distributions A and μ_A are from each other in terms of their relative likelihoods of producing observations.

$$\mathcal{L}_{KL} = D_{KL}(A \parallel \mu A) \quad (18)$$

The categorical cross-entropy loss function is used to train the trajectory classifier to distinguish the mental-state classes: normal, mild anxiety, moderate anxiety, and depression.

$$\mathcal{L}_{cls} = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \quad (19)$$

The probabilistic regression loss estimates continuous severity and creates a probabilistic representation of each clinical symptom rather than creating a deterministic outcome.

$$\mathcal{L}_{reg} = \parallel y - \mu_y \parallel_2^2 + \sigma_y \quad (20)$$

This outcome reflects how psychologists assess clinical symptoms rather than simply determining severity levels from observation, and allows the NTSM-Net to generate risk assessment outputs using probability. All four losses are combined into the loss function as:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{rec} + \lambda_2 \mathcal{L}_{KL} + \lambda_3 \mathcal{L}_{cls} + \lambda_4 \mathcal{L}_{reg} \quad (21)$$

where the λ_1 , λ_2 , λ_3 , and λ_4 coefficients control the contribution of each of the sparse encoder loss, Bayesian regularisation loss, trajectory classification loss, and severity regression loss. Next, Bayesian attention optimization is performed by assuming the Gaussian prior $P(A_{ij}) = N(0,1)$. The posterior is approximated via variational inference, and the objective optimizes the Evidence Lower Bound (ELBO) as:

$$\mathcal{L}_{ELBO} = \mathbb{E}_{q(A|X)}[\log p(y|X,A)] - KL(\, q(A|X) |\, p(A)) \quad (22)$$

Monte Carlo sampling with $S = 3$ samples per forward pass is used. The following hyperparameters are used: $\lambda_1 = 0.5$, $\lambda_2 = 0.1$, $\lambda_3 = 0.3$, $\lambda_4 = 0.1$, learning rate of 1e-4 , and Adam optimizer. The multi-objective loss allows the NTSM-Net to learn simultaneously low-level sparse affective features, uncertainty-aware temporal dynamics, high human emotional trajectory classification, and continuous severity estimation.

TABLE I. NOTATIONS LIST

Notation	Meaning
x_m	Input feature vector of modality m
D_m	Learned dictionary matrix for sparse encoding
α_m	Sparse latent representation of modality m
λ	Sparsity regularization coefficient
e_t	Behavioral micro-event salience score at time t
Q, K, V	Query, Key, and Value matrices in the transformer
A_{ij}	Bayesian attention weight between time steps $i$and j
μ_A, σ_A	Mean and standard deviation of attention posterior
p	Personality embedding vector
$\gamma(p), \beta(p)$	Personality-aware modulation parameters
τ^*	Aggregated emotion trajectory representation
$y, \hat{y}$	Ground-truth and predicted class labels
μ_y, σ_y	Predicted severity mean and uncertainty
$\mathcal{L}$	Total training loss function

IV. RESULTS AND DISCUSSION

Experiments were conducted on an NVIDIA RTX 4090 GPU workstation running Ubuntu 22.04 using PyTorch and CUDA for acceleration. The model was trained using the AdamW optimizer with a learning rate of 1×10^{-4}, cosine annealing, a batch size of 16, through 120 epochs with dropout (0.3), weight decay (1×10^{-5}), and topographical pruning from epoch 30 to prevent overfitting. Personality embeddings were z-score normalized, and time-series modalities were temporally aligned before being fused; mixed precision training was employed to increase computational efficiency.

The following datasets were used to examine the performance of the proposed model. DEAP [30] is an extremely popular and widely used multimodal benchmark dataset for emotion and affective state assessment/analysis, comprising 32 subjects with EEG, peripheral physiological signals, and frontal facial video recordings. The 32 subjects viewed 40 affective stimuli and provided their self-rate of the level of arousal, valence, dominance, and liking they experienced on a seven-point scale. The SEED dataset [31], also known as the Emotion EEG Dataset, consists of EEG recordings acquired from 15 participants with a 62-channel EEG setup for viewing emotionally evocative video segments recognized as positive, neutral, and negative. The WESAD dataset [32] contains recordings from 15 people who wore two different types of wearable sensors, the RespiBAN Chest Device and the Empatica E4 wristband, while experiencing three different conditions. RAVDESS [33] is regarded as a multimodal affective database, consisting of 7356 recordings created by 24 professional actors depicting eight classes of emotions: neutral, calm, happy, sad, angry, afraid, disgust, and surprise. The IEMOCAP dataset [34] comprises interactions between two actors who were either improvising or performing from a script. Each interaction is annotated with emotion labels such as angry, sad, happy, frustrated, excited, and neutral.

The labels derived from those datasets were considered proxy indicators of anxiety and depression based on the body of evidence supporting the relationship between emotions. Consequently, the learning task is framed as a supervised detection of affective states associated with anxiety/depression that allows for a valid comparative evaluation between models. The evaluation protocol used was k-fold subject-independent cross-validation without overlap of subject identity across splits. The statistical significance of the results was evaluated through paired t-tests conducted over five independent runs with different random seeds (significance level: $p < 0.05$). The cross-dataset harmonization step ensured the consistency of the datasets that had varying sampling rates, modal types, and annotation methods. All modalities were temporally aligned using a shared 100 ms reference grid. All numeric features were normalized using a z-score or min-max approach based on the modal types of those features. A hybrid imputation approach was used to handle missing data with the interpolation of small audio and video gaps. A GRU-based forward-fill model was used to fill in missing physiological windows, and the collapse of missing whole modal types was accomplished through cross-modal knowledge distillation in the NTSM-Net.

Many different types of performance evaluations were used to analyze the effectiveness of the proposed model in classifying emotions. The accuracy, precision, recall, and F1-score metrics were used to evaluate the correct class assignments of the model. The area under the ROC curve was used to measure how well NTSM-Net can discriminate between different classes for varying thresholds of decisions. The generalized strong performance of the model, based on data from five different datasets representing human emotion, is evident from the performance comparisons as shown in Figure 2. Specifically, in the DEAP dataset, NTSM-Net achieved a high level of accuracy (92.4%) with equally balanced Precision, Recall, and F1-Score of approximately 92%. The AUC was also notably high at 95.3%. This indicates that the model reliably differentiated between emotional states. In the WESAD dataset, it was able to deliver even better results; while still achieving a high degree of accuracy at 94.8%, it also produced an AUC of 96.8%. The increased accuracy on the WESAD dataset was due to the model's performance being based on more than one type of basic sensor data when classifying physiological emotions. The model's performance topped out at 96.8% for the SEED dataset, indicating that it has the capacity to perform high-quality feature extraction with EEG-based emotion classification. In addition, NTSM-Net achieved accuracies of 91.2% on the RAVDESS dataset and 89.7% on the IEMOCAP dataset.

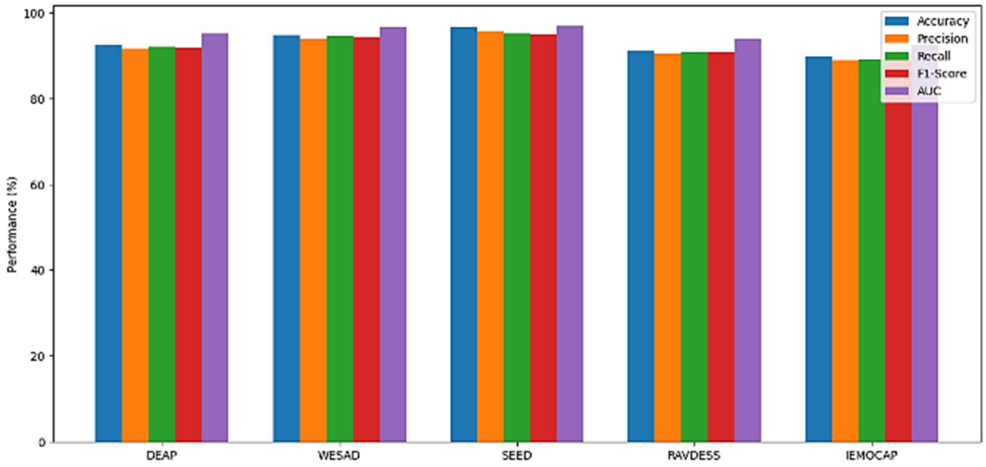

Fig. 2. Overall performance of NTSM-Net on all datasets.

The results derived from ablation studies provide a clear overview of the contribution of each aspect that makes up NTSM-Net. The combined model achieved the greatest performance with an accuracy of 96.8% and an F1 score of 96.9%. The removal of the Neuro-Inspired Sparse Encoder showed a substantial decline from the full baseline, specifically with a drop of 5.1%, attesting to the significance of sparse feature encoding for capturing subtle affective cues. Additionally, removing the Temporal Modulation had a negative impact on the model's performance, resulting in a 3.9% decrease in accuracy from the full baseline. The greatest reduction (6.7%) resulted from the removal of the attention-based Graph Attention Network module, which emphasizes the need to utilize attention to highlight discriminating temporal-neural patterns.

TABLE II. ABLATION STUDY OF NTSM-NET

Model variant	Accuracy (%)	F1-score (%)
Full NTSM-Net	96.8	96.9
Without Sparse Encoder (NSEM)	90.3	89.7
Without Temporal Modulation (T-Fusion)	91.4	90.6
Without Attention (GAT-Mod)	88.2	87.8
Without Residual Modulation	92.1	91.5
Only CNN Baseline	85.7	84.9

As shown in the results of cross-data evaluations in Table III, NTSM-Net's ability to generalize across many types of emotion and physiology is exceptional. For example, using DEAP to train a model and testing it with SEED data resulted in 72.1% accuracy when using a baseline deep neural network, and 76.8% accuracy when employing a transformer-based model, while NTSM-Net achieved an outstanding 82.6% accuracy. NTSM-Net was the only model to achieve an impressive 81.2% accuracy score, more than 10% higher than the baseline model, and over 7% higher than transformer-based approaches. Moreover, the model continued to show superior results across all physiological protocols, achieving 80.9% and 76.4% accuracy scores, respectively, connecting to physiologically different datasets (WESAD to DEAP) and transferring emotions through audio and video across two different datasets.

TABLE III. CROSS-DATASET GENERALIZATION PERFORMANCE

Train → Test	Baseline (%)	Transformer (%)	NTSM-Net (%)
DEAP → SEED	72.1	76.8	82.6
SEED → DEAP	70.4	74.1	81.2
WESAD → DEAP	69.7	73.5	80.9
RAVDESS → IEMOCAP	65.2	69.8	76.4

Figure 3 shows a comprehensive comparison of state-of-the-art methods for emotion recognition from EEG data. The proposed NTSM-Net outperformed all other current methods on this dataset by achieving an accuracy of 96.8%. This is significantly greater than EmoSTT (a Spatial-Temporal Transformer model) and the Robust 2D-CNN applied to Differential Entropy (DE) features, which had accuracies of 92.67% and 93.05%, respectively. PESD, which uses a pre-trained encoder trained with sensitive data, achieved an accuracy of 93.14% while Region-based BiLSTM EEG Modeling and STGATE achieved 93.05% and 90.3%, respectively. The spike-based EESCN method had lower accuracy at 79.65%. The high level of accuracy can be attributed to using multimodal fusion of data as well as subject-of-care evaluations; however, performance is subject to change as the patient base becomes increasingly diverse and clinically varied.

The proposed NTSM-Net is a research framework for affective computing and behavioral modeling. It was not developed as a clinical diagnostic or medical decision support system. Any deployment in a healthcare setting must be done through IRB-approved validation studies, have clinical supervision, and comply with the regulations established by GDPR and HIPAA.

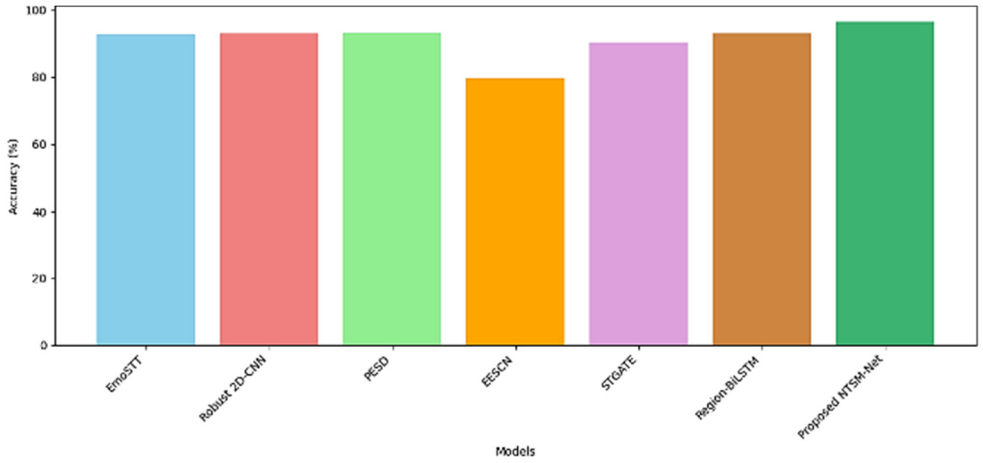

Fig. 3. Accuracy comparison on EEG-based emotion recognition on the SEED dataset.

V. CONCLUSION

This study introduced the NTSM-Net framework, comprised of Neuro-Inspired Sparse Encoding, Bayesian temporal reasoning, micro-event analysis, and individual personality modulation, for accurate classification of anxiety and depression. Through extensive experimentation, NTSM-Net achieved an accuracy of 96.8% and an F1-score of 96.9%. The model also demonstrated better generalization, greater robustness to noisy signals, and enhanced ability to detect weak behavioral cues across all evaluated datasets. It should be noted that removing architectural modules changes parameter capacities. In order to accommodate this, the learning rates were retuned for each ablation model, monitoring the convergence performance. However, this study did not strictly enforce perfect parameter normalization, which will be addressed in future research. Future work will also focus on real-time deployment on lightweight edge devices, examining the impact of cross-cultural behavioral differences, and investigating continual learning strategies for long-term mental health monitoring.

DECLARATION OF COMPETING INTERESTS

The authors declare no competing interests.

ACKNOWLEDGEMENT

Not applicable in this paper.

DATA AVAILABILITY

The datasets used can be found in [30-34].

REFERENCES

[1] M. Mansoor and K. Ansari, "Early Detection of Mental Health Crises through Artificial-Intelligence-Powered Social Media Analysis: A Prospective Observational Study," *Journal of Personalized Medicine*, vol. 14, no. 9, Sept. 2024, Art. no. 958, https://doi.org/10.3390/jpm14090958.

[2] B. H. Bhavani and N. C. Naveen, "An Approach to Determine and Categorize Mental Health Condition using Machine Learning and Deep Learning Models," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13780–13786, Apr. 2024, https://doi.org/10.48084/etasr.7162.

[3] M. A. Ruslim, M. J. Spencer, H. Hogendoorn, H. Meffin, Y. Lian, and A. N. Burkitt, "Emergence of Sparse Coding, Balance and Decorrelation from a Biologically-Grounded Spiking Neural Network Model of Learning in the Primary Visual Cortex." Neuroscience, Dec. 10, 2024, https://doi.org/10.1101/2024.12.05.627100.

[4] B. H. Bhavani, M. Sreenatha, and N. C. Kundur, "Diagnosis and Classification of Depressive Disorders using ML and DL Models," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 21383–21389, Apr. 2025, https://doi.org/10.48084/etasr.10017.

[5] Y. Xiao, H. Shao, J. Wang, S. Yan, and B. Liu, "Bayesian variational transformer: A generalizable model for rotating machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 207, Jan. 2024, Art. no. 110936, https://doi.org/10.1016/j.ymssp.2023.110936.

[6] Z. Wang, K. Zhang, W. Luo, and R. Sankaranarayana, "HTNet for micro-expression recognition," *Neurocomputing*, vol. 602, Oct. 2024, Art. no. 128196, https://doi.org/10.1016/j.neucom.2024.128196.

[7] M. Lukac, "Speech-based personality prediction using deep learning with acoustic and linguistic embeddings," *Scientific Reports*, vol. 14, no. 1, Dec. 2024, Art. no. 30149, https://doi.org/10.1038/s41598-024-81047-0.

[8] Z. Liu, S. X. Yin, G. Lin, and N. F. Chen, "Personality-aware Student Simulation for Conversational Intelligent Tutoring Systems," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 2024, pp. 626–642, https://doi.org/10.18653/v1/2024.emnlp-main.37.

[9] M. Song *et al.*, "Empowering Mental Health Monitoring Using a Macro-Micro Personalization Framework for Multimodal-Multitask Learning: Descriptive Study," *JMIR Mental Health*, vol. 11, Oct. 2024, Art. no. e59512, https://doi.org/10.2196/59512.

[10] G. Pushpa *et al.*, "An advanced AI framework for mental health diagnostics using Bidirectional Encoder Representations from Transformers with gated recurrent units and convolutional neural networks," *Ingénierie des systèmes d information*, vol. 30, no. 1, pp. 213–220, Jan. 2025, https://doi.org/10.18280/isi.300118.

[11] X. Jia, J. Chen, K. Liu, Q. Wang, J. He, and College of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, Shaanxi, China, "Multimodal depression detection based on an attention graph convolution and transformer," *Mathematical Biosciences and Engineering*, vol. 22, no. 3, pp. 652–676, 2025, https://doi.org/10.3934/mbe.2025024.

[12] A. Ahmad *et al.*, "A comprehensive bibliometric survey of micro-expression recognition system based on deep learning," *Heliyon*, vol. 10, no. 5, Mar. 2024, Art. no. e27392, https://doi.org/10.1016/j.heliyon.2024.e27392.

[13] A. Jain and D. Bhakta, "Micro-expressions: a survey," *Multimedia Tools and Applications*, vol. 83, no. 18, pp. 53165–53200, Nov. 2023, https://doi.org/10.1007/s11042-023-17313-6.

[14] X. Han, F. Chen, and J. Ban, "FMFN: A Fuzzy Multimodal Fusion Network for Emotion Recognition in Ensemble Conducting," *IEEE Transactions on Fuzzy Systems*, vol. 33, no. 1, pp. 168–179, Jan. 2025, https://doi.org/10.1109/TFUZZ.2024.3373125.

[15] Z. Xu *et al.*, "Depression detection methods based on multimodal fusion of voice and text," *Scientific Reports*, vol. 15, no. 1, July 2025, Art. no. 21907, https://doi.org/10.1038/s41598-025-03524-4.

[16] X. Huang *et al.*, "Depression recognition using voice-based pre-training model," *Scientific Reports*, vol. 14, no. 1, June 2024, Art. no. 12734, https://doi.org/10.1038/s41598-024-63556-0.

[17] M. Nykoniuk, O. Basystiuk, N. Shakhovska, and N. Melnykova, "Multimodal Data Fusion for Depression Detection Approach," *Computation*, vol. 13, no. 1, Jan. 2025, Art. no. 9, https://doi.org/10.3390/computation13010009.

[18] Z. Tan *et al.*, "Detecting Emotional Dynamic Trajectories: An Evaluation Framework for Emotional Support in Language Models." arXiv, 2025, https://doi.org/10.48550/ARXIV.2511.09003.

[19] Y. Zhang *et al.*, "Employing Machine Learning and Deep Learning Models for Mental Illness Detection," *Computation*, vol. 13, no. 8, Aug. 2025, Art. no. 186, https://doi.org/10.3390/computation13080186.

[20] S. Zhou, C. Gao, T. Delbruck, M. Verhelst, and S. C. Liu, "Exploiting neuro-inspired dynamic sparsity for energy-efficient intelligent perception," *Nature Communications*, vol. 16, no. 1, Nov. 2025, Art. no. 9928, https://doi.org/10.1038/s41467-025-65387-7.

[21] S. Bahadi, E. Plourde, and J. Rouat, "Efficient Sparse Coding with the Adaptive Locally Competitive Algorithm for Speech Classification." arXiv, Aug. 29, 2025, https://doi.org/10.48550/arXiv.2409.08188.

[22] S. Mamidisetti and A. M. Reddy, "A stacking-based ensemble framework for automatic depression detection using audio signals," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 7, 2023.

[23] M. Tveter *et al.*, "Advancing EEG prediction with deep learning and uncertainty estimation," *Brain Informatics*, vol. 11, no. 1, Dec. 2024, Art. no. 27, https://doi.org/10.1186/s40708-024-00239-6.

[24] S. Xie *et al.*, "Bayesian cooperative probabilistic Transformer for remaining useful life prediction with uncertainty estimation in industrial equipment," *Advanced Engineering Informatics*, vol. 67, Sept. 2025, Art. no. 103515, https://doi.org/10.1016/j.aei.2025.103515.

[25] I. H. Haraldsen *et al.*, "Intelligent digital tools for screening of brain connectivity and dementia risk estimation in people affected by mild cognitive impairment: the AI-Mind clinical study protocol," *Frontiers in Neurorobotics*, vol. 17, Jan. 2024, Art. no. 1289406, https://doi.org/10.3389/fnbot.2023.1289406.

[26] T. Kopalidis, V. Solachidis, N. Vretos, and P. Daras, "Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets," *Information*, vol. 15, no. 3, Feb. 2024, Art. no. 135, https://doi.org/10.3390/info15030135.

[27] P. Malik, J. Singh, F. Ali, S. S. Sehra, and D. Kwak, "Action unit based micro-expression recognition framework for driver emotional state detection," *Scientific Reports*, vol. 15, no. 1, July 2025, Art. no. 27824, https://doi.org/10.1038/s41598-025-12245-7.

[28] M. Hosseini *et al.*, "A multimodal stress detection dataset with facial expressions and physiological signals," *Scientific Data*, vol. 12, no. 1, Nov. 2025, Art. no. 1844, https://doi.org/10.1038/s41597-025-05812-0.

[29] E. Lim, M. Jhon, J. W. Kim, S. H. Kim, S. Kim, and H. J. Yang, "A lightweight approach based on cross-modality for depression detection," *Computers in Biology and Medicine*, vol. 186, Mar. 2025, Art. no. 109618, https://doi.org/10.1016/j.compbiomed.2024.109618.

[30] S. Koelstra *et al.*, "DEAP: A Database for Emotion Analysis ;Using Physiological Signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–31, Jan. 2012, https://doi.org/10.1109/T-AFFC.2011.15.

[31] W. L. Zheng and B. L. Lu, "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, Sept. 2015, https://doi.org/10.1109/TAMD.2015.2431497.

[32] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, Oct. 2018, pp. 400–408, https://doi.org/10.1145/3242969.3242985.

[33] S. R. Livingstone and F. A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English," *PLOS ONE*, vol. 13, no. 5, May 2018, Art. no. e0196391, https://doi.org/10.1371/journal.pone.0196391.

[34] C. Busso *et al.*, "IEMOCAP: interactive emotional dyadic motion capture database," *Language Resources and Evaluation*, vol. 42, no. 4, pp. 335–359, Dec. 2008, https://doi.org/10.1007/s10579-008-9076-6.