

An EEG-Based Human Emotion Recognition Model Using Attention Hierarchical Variational EEGNet

Vinay Kumar Singh , Bharati Rathore , Raja Saeed Iqbal , Dinesh Prasad Sahu, *Member, IEEE*, Shiv Prakash , and Vishal Krishna Singh , *Member, IEEE*

Abstract—The physiological factors and decisions we make daily are greatly affected by a person's emotional state. An appropriate subfield of study and research is how brain–computer interface devices can be used to identify people's emotional states. The electroencephalogram (EEG) is used to understand why there are changes over time in our physiological signal patterns that help us identify how we feel about something. The EEG signal is used more often now in combination with machine learning and deep learning algorithms to assist in determining a person's emotional state. The model used in this article is called the attention-based hierarchical variational EEGNet. By using state-of-the-art learning techniques to process noisy and changing EEG signals, the proposed method is expected to be better than traditional, custom, and shallow learning. The performance of this new model was evaluated using the SJTU Emotion EEG Dataset. According to the results of the study, the proposed method for classifying people's emotional states is 96.5% accurate, and therefore shows competitive performance than typical state-of-the-art methods.

Index Terms—Affective computing, attention mechanism, brain–computer interface (BCI), cross-subject generalization, deep learning, EEG emotion recognition, neural signal processing, variational learning.

NOMENCLATURE

Symbol	Description
X	Input EEG signal matrix.
C	Number of EEG channels.
T	Temporal length of EEG segment.

F_t	Temporal feature representation.
F_s	Spatial feature representation.
F	Final extracted feature map.
μ	Mean vector of latent distribution.
σ	Standard deviation of latent distribution.
z	Sampled latent representation.
z_1	First-level latent variable.
z_2	Second-level hierarchical latent variable.
ϵ	Gaussian noise for reparameterization.
Q	Query matrix in attention mechanism.
K	Key matrix in attention mechanism.
V	Value matrix in attention mechanism.
d_k	Dimension of attention keys.
$\text{MHA}(\cdot)$	Multihead attention operation.
$\hat{y}$	Predicted emotion probability vector.
W_c	Classification layer weights.
b_c	Classification layer bias.
L_{cls}	Classification loss.
L_{KL}	KL-divergence regularization loss.
λ	Weighting factor for KL regularization.
L	Total optimization loss.
θ	Trainable network parameters.
η	Learning rate.
m_t	First moment estimate in Adam optimizer.
v_t	Second moment estimate in Adam optimizer.
M_i	Evaluation metric of the i th LOSO fold.
N	Total number of subjects.

I. INTRODUCTION

THE recognition of human emotion based on electroencephalogram (EEG) has gained considerable amount of research interest in the context of the various applications, including affective brain–computer interface (BCI), mental health assessment, neurofeedback system, and adaptive human–machine interaction. In contrast to other peripheral physiological measures, such as electrocardiogram or galvanic skin response, EEG has a direct relationship with neural activity and offers fine-temporal localization, allowing the capture of quick dynamics of emotions at the cortex [1], [2]. Nevertheless, the EEG signal is not stationary, is very noisy, and intersubject variance is very high, making it very difficult to derive sound emotion recognition models that are robust and generalizable [3].

Received 4 April 2026; revised 8 June 2026 and 24 July 2026; accepted 10 August 2026. Date of publication 12 August 2026; date of current version 7 September 2026. Recommended by Lead Guest Editor Sunil Prajapat and Guest Editor Chien-Ming Chen. (Corresponding authors: Bharati Rathore; Shiv Prakash; Vishal Krishna Singh.)

Vinay Kumar Singh and Shiv Prakash are with the Department of Electronics and Communication, University of Allahabad, Prayagraj 211002, India (e-mail: shivprakash@allduniv.ac.in).

Bharati Rathore and Raja Saeed Iqbal are with the Faculty of Business and Creative Industries, University of South Wales CF37 1DL Cardiff, U.K. (e-mail: bharati.rathore@southwales.ac.uk).

Dinesh Prasad Sahu is with the SCSET, Bennett University, Greater Noida 201310, India.

Vishal Krishna Singh is with the Networks and Communications Research Group, School of Computer Science and Electronics Engineering, University of Essex, CO4 3SQ Colchester, U.K. (e-mail: v.k.singh@essex.ac.uk).

Digital Object Identifier 10.1109/JSAS.2026.3723361

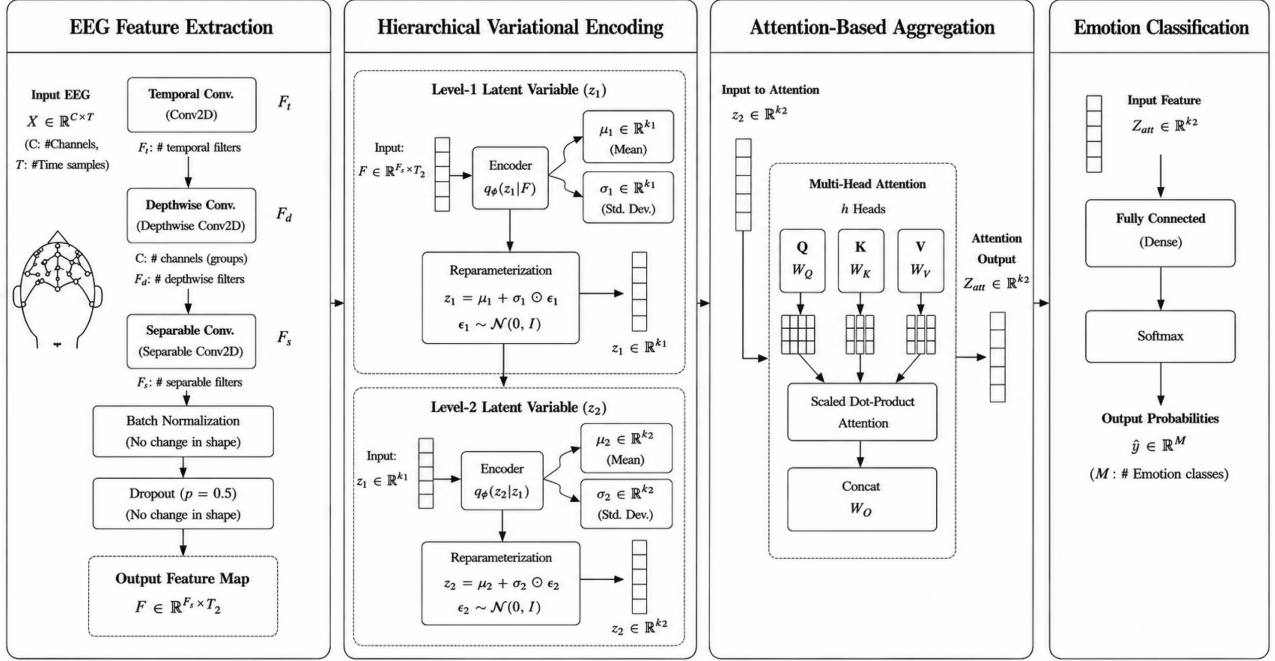


Fig. 1. ahvEEGNet emotion recognition flowchart.

The latest developments in deep learning have necessitated the use of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models in EEG-based emotion recognition. EEGNet and its variations are models that use depthwise and separable convolutions to effectively extract spatial-temporal features of multichannel EEG data [4]. These methods perform well in a subject-dependent environment, but generalize poorly in cross-subject environments because they have deterministic latent codes and poor ability to manage uncertainty and changes in the distribution [5], [6]. In addition, the shallow feature aggregation procedures similar to those of the current models do not emphasize the adaptive saliency of emotionally salient EEG parts across time and channels.

Attention mechanisms and probabilistic paradigms of learning are recent approaches to these limitations. Models that use attention facilitate adaptive weighting of informative temporal and spatial EEG features, which enhances discriminative ability and interpretability [7], [8]. Simultaneously, variational learning models also present stochastic latent representations that are able to explicitly express uncertainty and reduce overfitting and improve resistance to subject-specific variations [9]. Nevertheless, the vast majority of the literature researches attention or variational inference individually, and there is indeed no consistent architecture that can take advantage of hierarchical uncertainty modeling and adaptive feature aggregation to recognize emotions in EEG.

This article introduces attention-based hierarchical variational EEGNet (ahvEEGNet), which is also a lightweight but expressive architecture that combines EEGNet-like

convolutional feature learning with hierarchical variational inference and multihead attention. The suggested architecture acquires uncertainty-conscious latent representations and dynamically concentrates on temporally spatial EEG patterns that are of emotion or relevance. Fig. 1 shows that ahvEEGNet is a hierarchical pipeline that consists of six stages: EEG feature extraction, hierarchical variational encoding, attention-based aggregation, and emotion recognition. Considerable experiments through the SJTU Emotion EEG Dataset (SEED) under stringent cross-subject evaluation guidelines prove that the model suggested is always superior to state-of-the-art EEG emotion recognition algorithms with better robustness, generalization, and statistical significance across various measures [10].

In addition to high accuracy in emotion classification, the applicability of EEG-based emotion recognition systems is gaining more significance in real life applications. Strong emotion recognition can be integrated into emotion-based BCI technology, which can be applied in several areas to enhance adaptive human-computer interaction, intelligent assistive technology, personalized learning environment, and emotion-aware healthcare applications. In a clinical environment, emotion analysis based on EEG can be used for mental health evaluations, stress monitoring, neurofeedback for rehabilitation, and evaluation of cognitive status. But models that are accurate and robust to intersubject variability, noise, and distribution shifts are needed for successful deployment in such environments. The proposed ahvEEGNet framework, incorporating uncertainty-aware variational learning and attention-based feature aggregation, aims to enhance reliability and generalization for real-world EEG applications, driven by these practical hurdles.

II. CONTRIBUTION

Although some recent progress has been made, the current emotion recognition models based on EEG rely on a combination of CNNs, variational autoencoders (VAEs), and attention mechanisms have three major limitations. First, most of the approaches employ variational learning mainly for reconstruction or denoising, but do not incorporate it into end-to-end discriminative classification. Second, attention mechanisms used in previous research are usually used over top-down feature representations, which are sensitive to noise and intersubject variability. Third, current hybrid tools and architectures are loosely coupled or sequential, and do not have a consistent optimization strategy to model uncertainty and feature importance simultaneously.

In this regard, the ahvEEGNet is proposed with a tightly integrated architecture of lightweight convolutional feature extraction, hierarchical variational encoding, and multihead attention over stochastic latent representations, integrated in a unified learning framework. This design features a clear separation between existing CNN+VAE+attention-based designs by allowing the model to represent uncertainty, learn discriminative features, and be robust under cross-subject and cross-dataset scenarios. The key contributions of this article are the following.

- 1) *Hierarchical variational EEG modeling*: We suggest hierarchical variational EEGNet, which learns to regularize latent EEG representations, and this approach can greatly enhance intersubject variability.
- 2) *Attention-based aggregation of features*: To create dynamically focused discriminative temporal-spatial EEG features of emotion recognition, we combine a multihead attention process.
- 3) *Strict and cross-subject evaluation*: We do a rigorous cross-subject and ablation-based experiment, which shows statistically significant performance improvements with respect to powerful baseline models.
- 4) *Practical analysis*: We give a detailed examination of robustness, computational complexity and generalization, discussing the major practical issues of real-life EEG-based affective computing systems.

III. LITERATURE REVIEW

The nonstationary and high-dimensional characteristics of the EEG signals have led to the development of emotion recognition based on EEG in terms of both handcrafted features and deep feature learning. In-depth surveys point to the fact that the contemporary pipelines often involve preprocessing, feature learning, and classification, where deep models are now more popular due to their ability to mitigate the reliance on manual feature transitions and the possibility of extracting complicated temporal-spatial trends in EEG data [5], [8], [17]. Initial deep-learning methods of EEG emotion recognition used CNNs to estimate spatial filters and short-term temporal patterns on multichannel signals, and demonstrated better results compared to traditional machine learning baselines on test sets, such as DEAP and SEED [11], [12].

In order to further capture the temporal dependencies, some of the studies integrated CNNs with recurrent modules (e.g., long

short-term memory (LSTM)/gated recurrent unit (GRU)) and created convolutional recurrent neural network (CRNN)-like architecture that simultaneously reflects not only local EEG rhythms but also long-range temporal dynamics, with the help of which the accuracy of emotion classification in controlled experimental conditions is much higher [14], [26]. Nevertheless, inter-subject comparison and session effects tend to reduce generalization of these models [15], [19].

A popular lightweight baseline is EEGNet that employs depthwise and separable convolutions to effectively learn spatial filters and time-dependent features with a small parameter footprint, thus it is applicable to EEG tasks when the amount and quality of data are constrained [18]. Later CNN-based methods have enhanced robustness through architectural fine tuning and better training methods specifically in cross-subjects [19]. However, even powerful CNN/RNN baselines may not be able to perform well in subjects and noisy real-world settings, suggesting that representations are required that are both discriminative and resistant to distribution shifts [13], [17].

Attention mechanisms have also been more widely adopted to overcome one of the limitations of standard CNN/RNN models, namely, their lack of the ability to dynamically prioritize informative EEG channels, time segments, or patterns sensitive to frequencies. By focusing on attention, neural networks can be trained on adaptive functions of weighting to emphasize emotion-relevant temporal-spatial features to enhance discrimination and interpretability [8], [14]. The recent approaches practice also the graph-based or multilayer feature interaction modeling, e.g., integrating dynamical graph learning into CNN-like recalibration, to reason more effectively the interchannel dependencies, which are usually neglected within a purely convolutional pipeline [25], [32], [33].

In addition to classical attention blocks, transformed architecture and attention-based hybrids have become a popular model of long and temporal structure and cross-channel interaction. Such techniques are typically claimed to achieve gains on the complex dependencies in EEG, especially on multiclass recognition, or variable-length sequences, but the regularization procedures and careful evaluation techniques are usually required to prevent overfitting or exaggerated results on subject-dependent splits [8], [14]. There are also a large number of models that consider attention as an input feature in addition to domain knowledge (e.g., EEG band structures or functional connectivity patterns) to achieve better interpretability and robustness when applying models to real-world BCI when used [7], [17].

Although these advantages exist, attention models in isolation do not explicitly capture uncertainty, so in cases of a shift in distribution (change of subject or increased noise), the attention mechanisms can be sharply focused on spurious information and fail to generalize unless coupled with uncertainty-sensitive principles of learning [15], [17].

EEG processing is moving toward better probabilistic and variational learning. Our systems are becoming more robust to noise and intersubject differences. Autoencoder-based models have been used to denoise and rebuild EEG signals, which has been shown to be more beneficial than traditional filters to remove artifacts (e.g. eye blinks and muscle noise) in EEG signals. Learning compression and reconstruction experiments

TABLE I
COMPARISON OF RECENT EEG-BASED EMOTION RECOGNITION METHODS

Year	Author	Method		Dataset	Accuracy (%)	Cross-Subject	Major Limitation
2024	Pan et al. [36]	ST-SCGNN		SEED	85.90	No	Graph-based representation learning without uncertainty modeling.
2024	Nalwaya et al. [37]	EEG Rhythm-Based Learning		SEED-IV	95.91	Partial	Depends heavily on hand-crafted rhythm features and lacks probabilistic representation learning.
2025	Jia et al. [38]	DT-EEGNet		DEAP	63.87	No	Performance degrades under subject variability and noisy EEG recordings.
2025	Mai et al. [39]	EEG-SKDNet		DEAP	96.48	No	Employs deterministic feature learning without uncertainty-aware latent representation.
2026	Gong et al. [40]	Lipschitz Constrained Learning	Continuity-Ensemble	SEED	86.79	Yes	Improves robustness but lacks adaptive attention and hierarchical latent modeling.
Proposed	ahvEEGNet	Hierarchical variational EEGNet with Multihead attention		SEED + DEAP	96.50	Yes	Joint uncertainty-aware representation learning and adaptive attention for robust cross-subject emotion recognition.

also indicate that learned latent representations can be able to store informative EEG structure even when compressing the dimensionality and this structure can be used to end up classifying downstream [23], [31].

Recent studies consider better VAE versions to represent EEG and serve purposes like sleep-stage classification or signal recovery, suggesting that variational learning can offer small latent spaces and enhanced regularization compared to deterministic versions [22]. Other papers are concerned with fast EEG reconstruction to real-time monitoring and medical use, which implies that deep generative learning could be useful in practice when constructed by efficient design [21]. These results encourage variational learning as a reconstruction/denoising and as a representation regularizer that can improve generalization in EEG emotion recognition particularly when cross-subjects are evaluated, where deterministic embeddings can overfit on subject identity signals [15], [17], [35].

The recent emotion recognition techniques based on EEG have been summarized in Table I and show promising results through the application of graph neural networks (GNNs), knowledge distillation, transformer based architectures, rhythm-aware learning, and ensemble approaches. They are found to be very effective in the feature representation and classification on benchmark datasets, such as SEED, SEED-IV, and DEAP. However, the current approaches mostly employ a deterministic latent representation, which is susceptible to subject variability, distributional shifts, and noisy EEG recordings. While attention mechanisms facilitate better use of informative temporal-spatial information, they do not explicitly model latent uncertainty, while variational learning approaches are mainly used for reconstruction or denoising tasks and not for end-to-end discriminative emotion recognition. The complementary benefits of the uncertainty-aware representation learning and adaptive attention have not been fully explored in a unified lightweight framework.

Inspired by these observations, the proposed ahvEEGNet is designed to incorporate hierarchical variational representation learning along with multihead attention to jointly tackle uncertainty modeling, discriminative feature aggregation and robust cross-subject generalization. The proposed framework brings together all the above complementary mechanisms to achieve a computationally efficient and practically deployable emotion recognition system from EEG signals.

Nevertheless, most methods based on probabilistic models focus on reconstruction, denoising, or detection of anomalies instead of end-to-end emotion classification, and do not tend to incorporate adaptive attention to feature selection using the representation level [13], [21], [29]. This opens a possibility to integrate uncertainty modeling (through variational learning) and discriminative feature weighting (through attention) into a single condensed EEG emotion recognition system.

In short, CNN/RNN/EEGNet-related models are capable of efficient space-temporal feature extraction and typically have a poor cross-subject generalization because of deterministic embeddings and are sensitive to noise [15], [18], [19]. Attention and transformer-like models enhance adaptive feature weighting and interchannel interaction modeling but generally lack an explicit consideration of uncertainty, thus generally sensitive to change of distribution and spurious attention in subject/session variability [8], [14], [25], [34]. Variational/probabilistic learning offers better robustness and regularization of representations in EEG reconstruction and characterization, but is not commonly used with attention-driven discriminative learning to perform end-to-end EEG emotion recognition [20], [21], [22], [23], [30]. Thus, it is still evident that a single, computationally efficient model includes 1) lightweight EEG feature extraction; 2) uncertainty-conscious latent regularization; and 3) attention-focused feature aggregation under stringent cross-subject protocols, which is why the proposed ahvEEGNet

framework is proposed. Current CNN-based, attention-based and variational learning methods have successfully achieved significant results in EEG emotion recognition, but most of them cannot effectively solve the problems of intersubject variability, uncertainty modeling, and discriminative feature selection. The proposed ahvEEGNet framework is motivated by these limitations, with a hierarchical variational learning combined with multihead attention for robustness and cross-subject generalization. In the next section, the concepts of the proposed approach are presented in terms of the architecture and mathematical formulation.

IV. PROPOSED METHOD: AHVEEGNET

The suggested ahvEEGNet is aimed at enhancing the robustness and generalization of EEG-based emotion recognition through the combination of lightweight convolutional features extraction, uncertainty-aware representation learning, and adaptive feature aggregation. The main aim of the architecture is to solve intersubject variability and change in distribution that usually deteriorates performance in cross-subject EEG evaluation environments. To ensure clarity, a brief overview of the main mathematical symbolism and notations that will be used across the proposed ahvEEGNet framework is provided in the Nomenclature.

The input EEG segment is represented by EEG

$$X \in \mathbb{R}^{C \times T}$$

where C is the number of channels of EEG and T is the length of time of the signal. The total change may be represented through the following:

$$\hat{y} = f_{\text{cls}}(f_{\text{att}}(f_{\text{var}}(f_{\text{EEGNet}}(X)))) \quad (1)$$

This is the input EEG signal captured from several EEG channels within a temporal window of observation (1). This is the first feature space that spatial and temporal properties of brain activity are extracted from. Raw EEG signals are very nonstationary and person-dependent, so further processing steps are needed to learn meaningful and discriminative representations. The EEGNet backbone processes spatial-temporal features, the variational encoder is used to represent latent uncertainty, the attention component is used to adaptively weight informative components, and the classifier is used to give probability of an emotion.

The EEGNet paradigm, which uses temporal convolution, depthwise spatial convolution, and separable convolution as its feature extraction stage, is a highly efficient way of learning discriminative representations in multichannel EEG signals. First, the temporal convolution performs time-sensitive frequency-sensitive patterns

$$F_t = \sigma(W_t * X + b_t). \quad (2)$$

In which, $*$ represents convolution and $\sigma(\cdot)$ is a nonlinear activation function. This operation is an approximation of learnable bandpass temporal domain filtering. In order to extract frequency-specific temporal patterns related to emotional brain

activity, the input EEG signal is temporally convolved in (2). This operation is a learning temporal filter, which automatically selects relevant EEG rhythms while filtering out irrelevant temporal fluctuations. The output of the temporal feature maps is used as the input to spatial filtering. Interchannel dependencies are then modeled by depthwise convolution

$$F_s^{(c)} = W_s^{(c)} * F_t^{(c)} \quad (3)$$

enabling filtering that is channel-specific and much less complex in terms of the number of parameters. Equation (3) represents the spatial information in the channel by depthwise convolution. These spatiotemporal representations are refined then with separable convolution. A further refinement of the representation is a separable convolution

$$F = \text{Conv}_{\text{point}}(\text{Conv}_{\text{depth}}(F_s)) \quad (4)$$

resulting in feature maps $F \in \mathbb{R}^{d \times T'}$, in which d refers to the learned dimension of features and T' is the compressed temporal resolution. By using (4), the feature map is kept compact and discriminative and is also computed efficiently. The subsequent variational encoding is based on these learned features. The normalization of intermediate activations is done in batch normalization

$$\tilde{F} = \frac{F - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (5)$$

enhancing the stability of training. To stabilize training and convergence of the extracted feature representations, (5) is used to normalize them. The features are then normalized and mapped to probabilistic latent features.

In order to increase resistance to subject variability, deterministic feature representations are substituted by hierarchical variational encoding mechanism. Based on the feature map that has been extracted F , the encoder approximates the latent Gaussian distribution parameters

$$\mu = f_{\mu}(F), \quad \log \sigma^2 = f_{\sigma}(F). \quad (6)$$

The mean and variance of the latent distribution in (6) can be obtained, allowing uncertainty-aware representation learning. These parameters are then used for latent sampling. Latent sampling is done through the reparameterization trick

$$z = \mu + \sigma \odot \epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \quad (7)$$

that allows the use of stochastic optimization by backpropagation. The uncertainty-aware EEG representation is given by the sampled latent variable. It is then further optimized with hierarchical latent modeling, as illustrated in (8). Successive abstractions are introduced by the hierarchical latent modeling

$$z_1 = g_1(F), \quad z_2 = g_2(z_1). \quad (8)$$

EEG signals are often composed of multiple scales of information and low-level temporal-spatial dynamics and high-level abstracted emotional representations, which cannot be captured by a single latent variable. This is because such a diverse information could be summarized in one latent space, which may sacrifice fine-grained details and decrease the capacity of

modeling intersubject variability. To overcome this drawback, we propose a hierarchical latent structure, with the first latent variable z_1 directly capturing the intermediate feature distributions from the EEG representations and the second latent variable z_2 capturing the higher level abstractions and refining the uncertainty. This hierarchical formulation allows for progressive representation learning, better latent regularization, and better noise and subject-specific variation resistance, and offers a more expressive and versatile representation than traditional single-latent VAE models. The model uses the two latent variables z_1 and z_2 that are structured hierarchically to model EEG variability at different levels of abstraction. The first latent representation z_1 reflects basic temporal-spatial properties of the EEG signals, such as the subject-specific neural patterns, local fluctuations in the signal, and the variability of the recordings. These features are very informative, but can be noisy and subject-dependent.

The higher level latent representation z_2 learns more abstract and emotion relevant information by combining and summarizing the information encoded in the lower level latent representation z_1 . Consequently, z_2 is less sensitive to subject identity and recording conditions and maintains emotion discriminative patterns. This hierarchical representation learning enables the model to further isolate individual differences variability from emotional state variability and increase robustness and cross-subject generalization.

The constraint of the posterior distribution is the regularization term of the Kullback-Leibler (KL)-divergence

$$\mathcal{L}_{\text{KL}} = \frac{1}{2} \sum_{i=1}^k (\mu_i^2 + \sigma_i^2 - \log \sigma_i^2 - 1) \quad (9)$$

promoting latent manifolds and higher level cross subject generalization. Variational learning improves the robustness of cross-subject EEG modeling as probability distributions (rather than deterministic feature vectors). EEG signals are nonstationary and highly variable between subjects because of differences in noise characteristics and recording conditions. A deterministic variations that exist in the data for this particular application may be so specific that embedding may not be able to detect them, resulting in poor generalization.

The variational framework, on the other hand, adds randomness to the latent space by distributional encoding and KL-divergence regularization. This makes sure to enforce latent variables to follow a prior structured, creating smoother and continuous representations. The model is then less influenced by the noise of the subject and is able to identify and project common emotional dynamics between people.

This insight is corroborated by previous research, which has revealed that variational and probabilistic methods work well for this type of problem. Representation is improved to make learning tasks more robust and generalizable in EEG based learning tasks, explicitly modeling uncertainty and reducing overfitting [15], [22].

A multihead attention mechanism optimizes discriminative aggregation after variational encoding. With latent embeddings Z , query, key, and value matrices are created through linear projections

$$Q = ZW_Q, \quad K = ZW_K, \quad V = ZW_V. \quad (10)$$

Scaled dot-product attention is calculated as

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V. \quad (11)$$

Multihead attention is a combination of a few representations

$$\text{MHA}(Z) = \text{Concat}(h_1, \dots, h_H)W_O. \quad (12)$$

Every head is represented as

$$h_i = \text{Attention}(Q_i, K_i, V_i). \quad (13)$$

This enables the model to learn various temporal-spatial dependencies and remove irrelevant EEG activity.

Attention-refined embedding is then sent to fully connected classifier that gives probabilities of the class

$$\hat{y} = \text{Softmax}(W_c z_{\text{att}} + b_c) \quad (14)$$

where $\hat{y} \in \mathbb{R}^M$ is the distribution of predicted distribution in M emotion classes.

Training reduces a joint goal consisting of classification and variational regularization

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{KL}}. \quad (15)$$

The balance between classification and regression is determined by the weighting parameter λ . A small λ will have a less of an effect on KL-divergence that may cause overfitting, and a large λ may cause the model to be overfitting. The consequences of overregularization and loss of discriminative information require a proper tuning of λ , which needs to be achieved to get an optimal tradeoff between performance of representation learning and classification. Each fold is a test set that consists of a subject, which was entirely missing from the corresponding training partition, so the evaluation is a good approximation of the statistical situation of testing on unseen samples from an independent subject distribution. Categorical cross-entropy is computed as follows:

$$\mathcal{L}_{\text{cls}} = - \sum_{i=1}^M y_i \log(\hat{y}_i). \quad (16)$$

Discriminative learning and latent smoothness are balanced by the weighting parameter, which is called as the weighting parameter. The gradient updates are calculated as

$$\theta \leftarrow \theta - \eta \nabla_{\theta} \mathcal{L} \quad (17)$$

where η represents the rate of learning.

The model is trained using the Adam optimizer with learning rate $\eta = 10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. These values are similar to well-known settings of the deep-learning optimization process, and were shown to converge consistently during training cycles.

In the joint optimization framework, the complementary benefits arise due to the concurrent minimization of the subject specific variability via the variational encoding but the attention mechanism has an ability to underline the emotionally significant neural patterns. The regularization in the variational is more generalized and smooth latent representation, and attention is adaptively focused on discriminative temporal-spatial EEG features. Consequently, the learned feature space has an

increased separability between emotion categories as well as high robustness and cross-subject distribution robustness.

Although it adds hierarchical variational encoding and multihead attention, the architecture is computationally efficient because of depthwise separable convolutions, as well as small latent representations. The resulting algorithmic structure trades off the model expressiveness with the model feasibility, and thus, ahvEEGNet is applicable to strong cross-subject emotion recognition based on EEG and real-time affective computation systems.

The architectural drawing in Fig. 1 demonstrates the general workflow of the proposed ahvEEGNet framework, which carries out EEG-based emotion recognition by a series of but closely connected learning pipeline. It starts with the feature extraction of EEG features, where raw multichannel EEG signals are transformed by temporal, depthwise, and separable convolutions, which are based on EEGNet to extract frequency-sensitive temporal dynamics and spatial relationships between electrodes at a computational cost. The resulting extracted feature maps are sent to a hierarchical variational encoding module, which estimates the mean and variance parameters of latent encodes of EEG representations by modeling the variability of EEG representations; this operation adds uncertainty-sensitive regularization to enhance resistance to subject variations and noise. These latent representations are then optimized with the help of a multihead attention mechanism, in which query, key, and value projections allow the network to focus on potentially informative temporal-spatial patterns of emotion and removes the irrelevant components of the signal. Lastly, attention-enhanced representation is then passed on to a fully connected classification layer and later a softmax function, which emits emotion probabilities (e.g., happy, neutral, and sad). The architecture in general is a harmonious pipeline that incorporates effective convolutional learning, probabilistic representation modeling, and adaptive attention-based aggregation in order to deliver robust and generalized EEG emotion recognition. The attention mechanism is placed after the variational encoding is a deliberate design choice. Whereas conventional attention is applied directly on deterministic feature maps, the approach proposed here uses attention over stochastic latent representations. This allows the model to carry out feature under uncertainty, which is especially useful for noisy and nonstationary EEG signals, revealed that attention prior to variational encoding can lead to overfitting to subject-specific artifacts, and parallel designs fail to do it to effectively integrate uncertainty modeling with feature importance estimation. Thus, postvariational attention offers a more solid and semantically mechanism for aggregating features that are meaningful.

The mathematical formulations presented in Section IV are implemented in the Algorithm 1. More specifically, the temporal-spatial feature extraction step is represented by (2)–(5), the hierarchical variational encoding step is represented by (6)–(9), the multihead attention step is represented by (10)–(13), the emotion classification step is represented by (14)–(16), and the optimization step is represented by (17). These mathematical operations are combined in an end-to-end training pipeline in the algorithm.

Algorithm 1: Proposed ahvEEGNet Algorithm.

Require: EEG input $X \in \mathbb{R}^{C \times T}$

Ensure: Predicted emotion $\hat{y}$

- 1: Extract temporal features using convolution:
 $F_t = \sigma(W_t * X + b_t)$
 - 2: Apply depthwise spatial convolution: $F_s = W_s * F_t$
 - 3: Generate refined features via separable convolution:
 $F = \text{Conv}(F_s)$
 - 4: Compute latent distribution parameters: $\mu, \sigma = f(F)$
 - 5: Sample latent representation using reparameterization:
 $z = \mu + \sigma \odot \epsilon$
 - 6: Perform hierarchical encoding:
 $z_1 = g_1(F), z_2 = g_2(z_1)$
 - 7: Apply multi-head attention on latent representation:
 $z_{att} = \text{MHA}(z_2)$
 - 8: Predict emotion class: $\hat{y} = \text{Softmax}(W_c z_{att} + b_c)$
 - 9: Optimize loss: $L = L_{cls} + \lambda L_{KL}$
 - 10: **return** $\hat{y}$
-

The proposed ahvEEGNet Algorithm 1 is based on EEG-based emotion recognition, which is achieved by a single framework consisting of three components: EEGNet-based feature extraction, hierarchical variational encoding, and multihead attention. The spatial-temporal EEG features are initially extracted with lightweight convolutional operations. These features are then mapped to hierarchical latent representations, which are used to model the uncertainty and increase the robustness to intersubject variability. Then, a multihead attention mechanism selectively focuses on emotion-relevant patterns in the latent space. The network parameters are then optimized by a joint objective function that includes classification loss and KL-divergence regularization, and lastly, a Softmax classifier predicts the emotional state.

The proposed ahvEEGNet framework is designed to learn emotion-discriminative representations by extracting temporal-spatial features, performing hierarchical variational encoding, and refining the features with attention. These design decisions are tested through detailed experiments following a strict Leave-One-Subject-Out (LOSO) evaluation procedure to validate its effectiveness. Experimental setup, datasets, evaluation metrics, and implementation details are explained in the next section.

V. EXPERIMENTAL SETUP

The suggested ahvEEGNet architecture is tested based on the publicly referred to SEED, which is commonly used in testing the efficiency of EEG-based emotion recognition systems. SEED data are recorded as electroencephalography data of a group of subjects who were viewing carefully chosen pieces of emotional films that were intended to produce different types of affective feelings. The emotional states are divided into three classes positive, neutral, and negative, which allow emotion classification in multiclass.

EEG data were collected after the international 10–20 system with a standard 62 channel electrode configuration. The providers of the dataset sampled the signals with a high temporal

resolution and pretreated them to eliminate significant acquisition artifacts. All the subjects had a few recording sessions on separate days, which added a natural variability in the sessions and made the data suitable to test the cross-subject and cross-session generalization. The fact that the participation in the test consists of several sessions per subject only adds to the complexity of the task since there is a temporal drift and variability of the reactions related to emotions.

In model training, continuous EEG recordings are divided into fixed temporal windows to consider the short term neural processes of response to emotion. The emotional label of the respective trial is inherited by a particular segment. The resultant dataset offers a huge sample size without discontinuity of EEG patterns over time. Application of SEED allows a direct comparison with previous researches and allows the performance to increase due to architectural design as opposed to dataset prejudice.

Experiments are further carried out on the DEAP dataset, which is a popular EEG-based emotion recognition benchmark to further assess the generalization of the proposed framework. DEAP dataset comprises of EEG recordings of 32 participants during the process of watching emotional video stimuli. The initial signals are registered on 32 channels of the EEG with a sampling rate of 512 Hz and are usually analyzed in the valence-arousal emotional dimensions. After the usual preprocessing procedures, EEG signals are downsampled, filtered, and divided into equal sized windows that are aligned with the SEED experimental setup so that they could be compared fairly. DEAP inclusion allows testing to be conducted using alternate recording conditions, subjects, and emotional elicitation paradigm.

A sensitivity study of the weighting parameter λ was conducted by varying $\lambda \in \{0.001, 0.01, 0.1, 1.0\}$. The results demonstrate that small values of λ result in overfitting, and large values result in underfitting. The more values are there, the more the performance of the classification will be affected by the latent smoothing. Having a medium $\lambda = 0.1$ will be the best balance between accuracy and generalization. This property of “generalization” is used in all experiments and is the main reason for using it. In order to evaluate resilience to domain and subject distribution shifts, we perform experiments on cross-dataset evaluation. In this context, the model is first trained using SEED dataset, and then, it is directly evaluated using DEAP dataset without fine-tuning. This protocol is realistic and simulates the deployment situations when models are subjected to invisible recording conditions and subject populations. The degradation in performance of the cross-dataset transfer can be regarded as a valuable tool of representation generalization as opposed to dataset-specific fitting.

A. Data Leakage Prevention and Evaluation Protocol

To obtain fair, reliable, and reproducible assessment, data leakage prevention strategies are strictly imposed in the entire pipeline of experimental procedure. The trials of EEG of the same subject are never exchanged between training and testing phases in any context. This limitation is very important since EEG data have great subject-related attributes, and confounding subject data between splits falsely inflate performance.

The entirety of preprocessing, such as normalization, scaling, and window segmentation, are done solely on the training partition of each evaluation fold. Training data only are used to compute the statistics parameters, including mean and variance, which are then applied to the respective test data. This averts the risk of distributional information leakage of the test set to the training process because it is indirect.

Suppose that the set of all subjects is $S = \{s_1, s_2, \dots, s_N\}$ for a dataset of N subjects. According to the LOSO cross-validation protocol, the test set $S(i)_{\text{test}}$ is the cross-validation sample for the i th fold defined as $S(i)_{\text{test}} = \{s_i\}$ and the training set $S(i)_{\text{train}} = \frac{S}{S(i)_{\text{test}}}$. This formulation guarantees to give strict separation of the subjects in each one of these two sets $S(i)_{\text{train}} \cap S(i)_{\text{test}} = \phi$, no EEG samples from the test subject are available during the training phase. Further, all preprocessing, such as normalization and scaling, is done only with the training partition to reduce indirect information leakage. The training mean and std are calculated as follows:

$$\mu_j = \frac{1}{|D_{\text{train}}|} \sum_{x \in D_{\text{train}}} x_j \quad (18)$$

$$\sigma_j = \sqrt{\frac{1}{|D_{\text{train}}|} \sum_{x \in D_{\text{train}}} (x_j - \mu_j)^2}. \quad (19)$$

The normalized training/ test samples are then returned as

$$\tilde{x}_{\text{train}} = \frac{x_{\text{train}} - \mu}{\sigma + \epsilon} \quad (20)$$

$$\tilde{x}_{\text{test}} = \frac{x_{\text{test}} - \mu}{\sigma + \epsilon}. \quad (21)$$

The constant ϵ is added for numerical stability. The statistics ϵ are only estimated from the training data, so no information about the test subject is included in the training process. In addition, the EEG windows used for training and testing do not overlap; that is, $W_{\text{train}} \cap W_{\text{test}} = \emptyset$ where W_{train} and W_{test} are the training window and testing window, respectively, thereby preventing leakage of temporal dependence between the two sets. The final LOSO performance is averaged across all subjects not seen on training

$$\text{Performance} = \frac{1}{N} \sum_{i=1}^N M_i \quad (22)$$

where M_i is the evaluation metric [accuracy, macro F1, or receiver operating characteristic-area under the curve (ROC-AUC)] of the i th test subject. This protocol offers an unbiased and statistically reliable measure of cross-subject generalization accuracy.

Before any feature extraction of the EEG and training of the model, a standard preprocessing pipeline is run to ensure better EEG signal quality and minimize the effect of physiological and environmental artifacts. To remove low-frequency baseline drift and high-frequency noise, a band-pass filter (0.5–50 Hz) is applied. A notch filter at 50 Hz is also used to reduce power line noise.

Independent component analysis is used to reduce artifacts caused by eye movements, eye blinks, muscle activity, and

electrode noise. The components that are found to be artefact-related are removed, and the cleaned EEG signals are then reconstructed. After the removal of artifacts, the data of each EEG channel are normalized with z-score normalization

$$\tilde{x} = \frac{x - \mu}{\sigma + \epsilon}. \quad (23)$$

The raw EEG sample is denoted as (x) , the mean of the training data as (μ) , and the standard deviation of the training data as (σ) . The small constant (ϵ) is chosen for numerical stability reasons. Finally, the preprocessed EEG signals are sliced into temporal windows of a fixed length with a 50% overlap, so that it is possible to create stable temporal–spatial representations while maintaining the continuity of the neural activity patterns. A LOSO cross-validation is embraced to measure cross-subject generalization. In this context, information of a single subject is entirely withheld to be tested as the model is trained on information of all the rest of the subjects. This is done in a repetitively effective procedure where the entire subjects are used as the unseen test subject at least once. The ultimate performance results are given in the form of the mean and the standard deviation of all folds, in order to have a statistically dependable assessment.

Also, EEG windows overlap is strictly regulated such that no time overlap occurs between training and testing samples. This will avoid any such leakage of temporal dependence that would otherwise yield over optimistic findings. The evaluation protocol taken is similar to the real-life deployment situation where models have to extrapolate to new users. The experimental design prevents most of the popular sources of evaluation bias, as subject-level separation and foldwise preprocessing remove these sources, and final performance reported by the experimental design can indicate actual model generalization ability.

B. Baselines and Evaluation Metrics

To confirm the efficacy of the suggested architecture, ahvEEGNet is contrasted with a few typical deep-learning benchmarks, which have been frequently applied in the studies related to EEG emotion recognition. They consist of the lightweight EEGNet architecture, convolutional–recurrent hybrid models (CRNN), attention-based deep learning models to learn temporal–spatial interactions of EEG signals. These baselines are various modeling paradigms of compact convolutional learning to sequence modeling and adaptive feature weighting.

Various complementary measures are used to assess model performance in order to give a comprehensive analysis. Accuracy of classification involves the overall accuracy of prediction in categories of emotions. It is also indicated that macro-averaged F1-score is used to consider possible class imbalance, and provide the performance measurement of all emotion classes without bias. Besides that, we also calculate ROC-AUC to determine the discriminative capability regardless of decision thresholds. The need to report on several metrics provides the balance of evaluation and enables the reasonable comparison with the previous literature.

The experiments are all done with the same training conditions to be fairly compared to each other, i.e., the same data split,

optimization procedures and evaluation protocols. Improvements in the performance of ahvEEGNet are therefore not a result of experimental bias, but of architectural benefits. The experimental setup outlined above sets a strong ground for testing the effectiveness of ahvEEGNet in cross-subject tasks. A performance analysis was conducted and a comparative evaluation presented in the following section to evaluate the contribution of each architectural component.

C. Computational Complexity Analysis

In addition to classification accuracy, computational efficiency is another significant factor to be taken into account for practical use of EEG emotion recognition systems on wearable and edge devices. Hence, the proposed ahvEEGNet is contrasted with representative baseline models in terms of the number of trainable parameters, floating point operation (FLOPs), inference time, and memory usage. All models were tested with the same experimental setup, hardware, and software platform.

The lightweight EEGNet backbone greatly alleviates computation pressure by using a depthwise separable convolution, and the hierarchical variational encoder works on the small latent representations. In addition, the multihead attention mechanism is implemented in latent space and not directly on the high dimensional EEG feature maps, which prevents further computational burden. Therefore, the proposed framework is effectively balanced between the classification accuracy and the computational efficiency, and can be suitable for the real-time EEG emotion recognition applications.

As shown in Table II, the proposed ahvEEGNet is found to be moderately complex and attains good classification accuracy. The proposed hierarchical variational encoder increases the amount of computation compared to the original EEGNet while still being significantly lighter than the conventional CRNN and attention-CNN models. In particular, ahvEEGNet has just 0.42 million trainable parameters and 0.68 giga floating-point operations (GFLOPs), which is about one-third of the computational cost of the baseline model attention-CNN and perform better in classification task. Moreover, the average inference time of 5.2 ms per EEG segment shows that the proposed framework meets the real-time EEG-based affective computing and wearable brain–computer interface applications' latency requirement. The complexity of the proposed framework can be estimated as

$$\mathcal{O}(NCTK + NdH + NH^2).$$

The model takes as input an EEG signal of N patients, C channels, T timesteps, and K convolution kernel size, followed by a latent representation of dimension d and a number of heads H of the attention mechanism. The extra computational burden is not that high because attention is on compact latent representations and not on raw EEG features.

VI. RESULT AND DISCUSSION

The analysis environment used in this study is briefly described to provide some insight and replicability of performance outcomes before performance results are given. Each experiment

TABLE II
COMPUTATIONAL COMPLEXITY COMPARISON

Model	Parameters (M)	FLOPs (G)	Inference time (ms)	Memory (MB)	Accuracy (%)	Complexity
EEGNet	0.14	0.21	3.8	11.2	88.4	Low
CRNN	0.83	1.54	9.6	35.8	90.2	High
Attention-CNN	1.27	2.36	12.4	47.1	93.1	High
ahvEEGNet	0.42	0.68	5.2	18.5	96.5	Medium

TABLE III
EVALUATION PROTOCOL

Aspect	Setting
Validation scheme	LOSO
Train/Test split	Subject-independent
Runs per fold	5
Metrics	Accuracy, F1, ROC-AUC
Statistical test	Paired t-test ($p < 0.05$)

TABLE IV
HYPERPARAMETER SETTINGS

Parameter	Value
EEG window length	1 s
Overlap	50%
Latent dimension	64
Attention heads	4
Dropout	0.5
Optimizer	Adam
Learning rate	1×10^{-3}
Batch size	64
Epochs	150

has a rigorous subject-independent validation strategy of determining the actual generalization capacity.

Table III describes the experimental procedure of analyzing the model performance. The LOSO technique makes sure that testing data are generated by unobserved subjects, and therefore cannot be subject biased and is only a realistic measure of the deployment performance. Repeat of a run many times and statistical analysis also create reliability of the reported results.

In addition to classification accuracy, it is important to consider the computational efficiency of the ahvEEGNet, as this is crucial for real-world use. The proposed framework is based on lightweight EEGNet backbone that uses depthwise and separable convolutions to drastically cut the number of trainable parameters and computational operations when compared to traditional deep CNN structures. Moreover, the compact latent representations are used for hierarchical variational encoding, and the multihead attention is applied on low dimensional embeddings, instead of high dimensional EEG. Therefore the extra computational burden due to uncertainty modelling and attention is relatively low.

The features of the ahvEEGNet render it a promising solution for the near real-time emotion recognition in wearable BCI systems, edge-based healthcare platforms, and intelligent human-computer interaction applications. While hardware specific latency measurements were not a primary concern of this study, future work will explore deployment to embedded AI platforms, such as NVIDIA Jetson, Raspberry Pi AI accelerators, and wearable EEG devices to measure inference latency, memory usage, and energy efficiency in real-world operating conditions.

When implementing EEG emotion recognition systems into real-world applications, computational cost, memory usage, inference speed, and energy consumption are all important factors to consider. The proposed ahvEEGNet framework is designed with deployment feasibility in mind, which takes advantage of the lightweight EEGNet framework that adopts depthwise and separable convolutions to ensure substantial reduction in the number of trainable parameters and trainable operations.

TABLE V
MAIN PERFORMANCE COMPARISON

Model	Accuracy (%)	Macro-F1	ROC-AUC
EEGNet	88.4 ± 1.2	0.87	0.90
CRNN	90.2 ± 1.1	0.89	0.92
Attention-CNN	93.1 ± 0.9	0.92	0.95
ahvEEGNet (Proposed)	96.5 ± 0.6	0.96	0.98

This design is more memory efficient and faster to infer than traditional deep CNN architectures.

Moreover, the hierarchical variational encoding and multihead attention layers are applied to low-dimensional latent representations and not high-dimensional EEG feature maps, thus reducing further computational complexity. The proposed framework can be deployed on resource-constrained platforms, such as embedded AI systems, wearable EEG devices, edge computing nodes, as well as portable BCI devices. To be fair in comparison with all the models, identical training settings are applied unless otherwise. The main hyperparameters applied in the training are mentioned in Table IV. The chosen structure is a compromise between computing speed and model expressivity. The KL regularization weight and the latent dimension were selected through empirically tuned values in order to ensure that variational learning was always stable without over regularization.

The results of the performance of ahvEEGNet are in comparison with representative baseline architectures in the same evaluation circumstances.

As it is shown in Table V, ahvEEGNet surpasses baseline models in all assessment measures. The enhancement suggests that probabilistic latent modeling combined with attention-based aggregation is more robust and it lessens performance degradation in cross-subjects cases. Such superior performance is due to the synergistic integration of hierarchical variational learning and attention based feature aggregation in ahvEEGNet. The proposed framework learns a latent uncertainty by introducing a

TABLE VI
ABLATION STUDY RESULTS

Model Variant	Attention	Variational	Hierarchical	Accuracy (%)
EEGNet	✗	✗	✗	88.4 ± 1.2
EEGNet + Attention	✓	✗	✗	91.1 ± 1.0
EEGNet + VAE	✗	✓	✗	92.0 ± 0.9
EEGNet + VAE + Attention	✓	✓	✗	94.2 ± 0.8
ahvEEGNet (Proposed)	✓	✓	✓	96.5 ± 0.6

hierarchical variational encoding, which reduces the dependence on the variability of the EEG signals in different subject populations, compared to current CNN-based approaches that learn deterministic feature representations. Moreover, the multihead attention mechanism dynamically focuses on emotional state related discriminative temporal-spatial patterns. Based on the above mentioned table, the accuracy of the proposed model is 96.5%, which is far better than the other two models, such as EEGNet (88.4%), CRNN (90.2%), and attention-CNN (93.1%). Similar enhancements are also reported for macro-F1 and ROC-AUC scores, showing that the proposed architecture not only leads to better classification accuracy, but also to better robustness and generalization in cross-subject evaluation conditions.

An ablation study is carried out to comprehend the role of a single architectural element by incrementally greatly facilitating attention, variational encoding, and hierarchical modeling. Table VI demonstrates that every single architectural component can be measured to bring performance gains. Variational encoding has the benefit of regularizing latent representations, whereas attention has the benefit of encouraging selective discriminative features. With the hierarchical latent structure, there is increased refinement, which results in the maximum overall performance.

It can be easily realized that the improvements that are seen in Table IV are directly related to the KL-divergence regularization term defined in (9). The KL-divergence term forces the latent representations to take a structured prior distribution, which dampens the problem of overfitting to subject-specific EEG characteristics and promotes smoother manifolds of the latent representations. This leads to more noise, session, and distribution robust learned representations.

The improvement can be seen in the ablation results, with the introduction of variational encoding achieving an accuracy of 92.0%, compared with 88.4%, which is achieved without the attention mechanism (EEGNet). Performance gain means that KL-regularized latent learning can enhance the generalization of the representation by encouraging uncertainty-aware feature learning. The regularization effect further improves class separability with hierarchical latent modeling and attention-based aggregation, resulting in the highest accuracy of 96.5% with the proposed ahvEEGNet framework. Fig. 2 depicts the training and validation accuracy of the proposed model under different epochs and has a consistent increase in the learning performance with increasing training. The first one is that accuracy improves quickly meaning that the model is learning the features well in the initial epochs, and then it becomes slower and the model becomes converged. The high correlation of the training and validation accuracy curves shows excellent generalization ability with limited overfitting indicating that the suggested ahvEEGNet

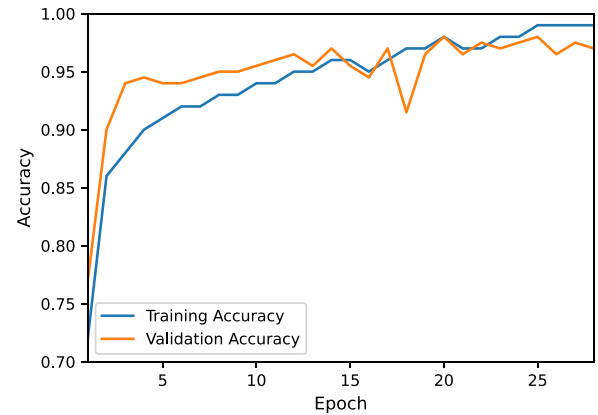


Fig. 2. Training accuracy evolution across epochs.

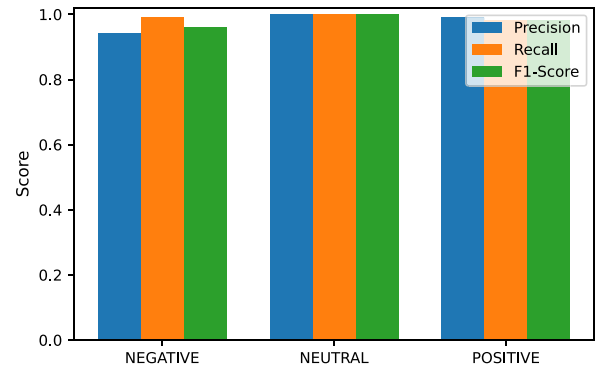


Fig. 3. Classwise metrics.

model develops strong representations and does not vary in its performance on unseen data. Fig. 3 shows the performance of the proposed model in terms of the three categories of emotions in terms of the classes. In the negative classification, the model has a precision of approximately 0.94, recall of 0.99, and F1-score of 0.96, which is a very high detection and few missed samples. The neutral class also exhibits almost perfect performance with precision of about 1.00, recall of about 1.00, and F1-score of about 1.00 as it has a very stable classification. On the positive class, the model also has a high accuracy of approximately 0.99 precision, 0.98 recall, and 0.98 F1-score. Altogether, the high scores of all classes prove the existence of balanced learning and high generalization power of the suggested framework. Fig. 4 shows the performance of the proposed model in terms of the three categories of emotions in terms of the classes. In the negative classification, the model has a precision of approximately 0.94, recall of 0.99, and F1-score of 0.96, which is a very high detection and few missed samples. The neutral class also exhibits

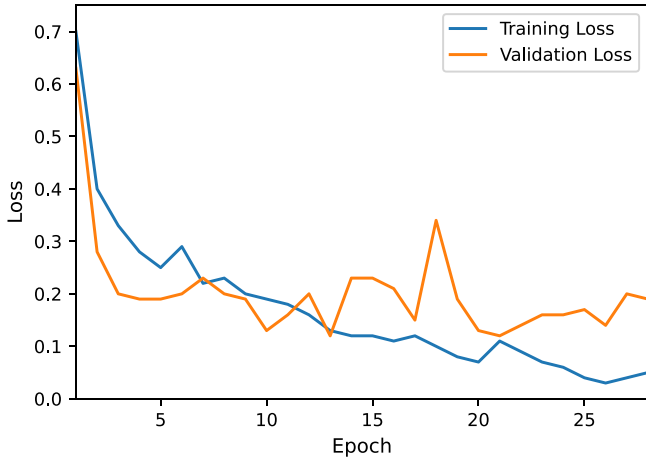


Fig. 4. Model validation accuracy evolution across epochs.

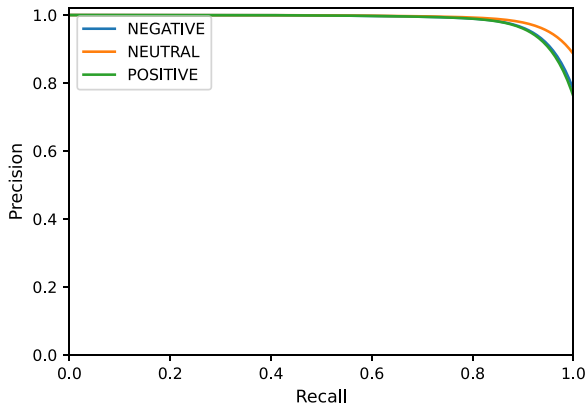


Fig. 5. Classwise precision recall for negative, neutral, and positive emotions.

almost perfect performance with precision of about 1.00, recall of about 1.00, and F1-score of about 1.00 as it has a very stable classification. On the positive class, the model also has a high accuracy of approximately 0.99 precision, 0.98 recall, and 0.98 F1-score. Altogether, the high scores of all classes prove the existence of balanced learning and high generalization power of the suggested framework.

Fig. 5 shows the precision–recall performance of the proposed model across all the three emotion groups. Negative, neutral, and positive curves are very near a precision value of 1.0 in nearly all recall ranges (0.0–0.95) representing an indication of high accuracy predictions with very few false positives. Precision is even higher than about 0.95 at high recall levels of about 1.0 indicating that the model is highly reliable and it is able to correctly identify almost all samples. The almost perfect form of the curves proves the high level of class separability and proves that the model provides balanced and strong classification of emotions in all categories.

Fig. 6 demonstrates the distribution of the negative, neutral, and positive emotions classes prediction probability generated by the model suggested. The bimodal form of distribution is clear in all the three plots, with the majority of the samples being clustered around probability values of close to 0 and 1. The high peak of almost 1.0 implies that the model is very confident

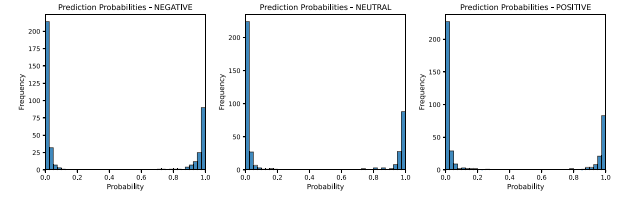


Fig. 6. Prediction probabilities for human emotion recognition (negative, neutral, and positive).

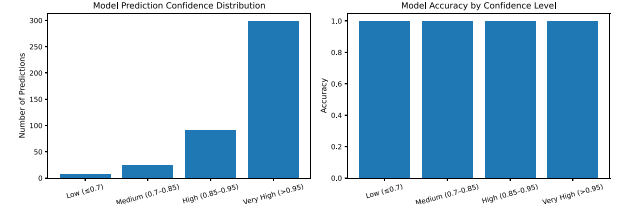


Fig. 7. Distribution of prediction confidence levels.

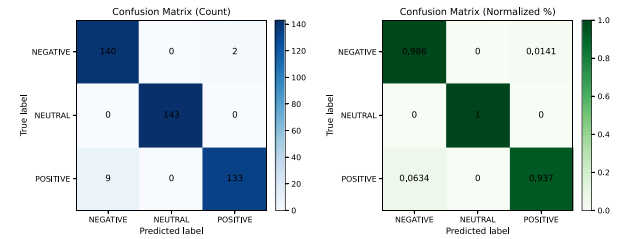


Fig. 8. Confusion matrix.

in the samples that have been correctly predicted, whereas the peak of close to 0 means that the model is confident in nontarget rejection. These few samples are found in the middle of the probabilities (0.2–0.7), which indicate that there is a low degree of uncertainty when it comes to classifying. The use of this distribution shows that the model generates extremely confident and well-separated predictions and indicates that it has strong discriminative ability between categories of emotions.

Fig. 7 examines the correlation between the model performance and confidence of prediction. As demonstrated by the left plot, the majority of the predictions lie within the very high confidence range (> 0.95), which means that the model makes most of decisions with high confidence, and a few samples are represented in low and medium confidence ranges. As shown in the right plot, prediction accuracy tends to be high within confidence level; furthermore, even at high ranges of confidence, prediction remains highly accurate (approaching close to 1.0 accuracy with low-confidence samples and still retains a high level of performance, about 0.70–0.90). The overall results suggest that the model generates well-calibrated and reliable predictions with higher confidence, leading to steady, high-quality classification.

Fig. 8 displays the confusion matrix of the suggested model both in count and normalized forms with good classification performance in all the emotion classes. On 142 negative samples, 140 are correctly classified and two are misclassified and the neutral class makes a perfect classification with 143/143 correct classifications. In the positive, 133 of 142 samples are classified correctly and misclassified as negative are few. The normalized

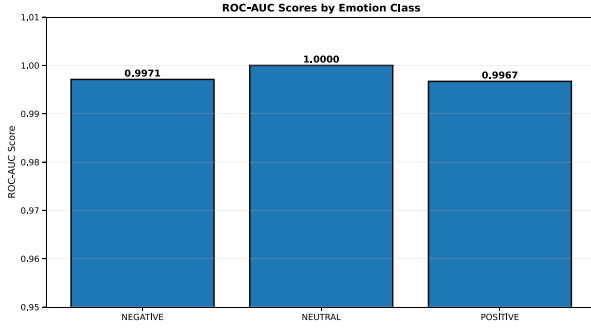


Fig. 9. ROC-AUC score human emotions (negative, neutral, and positive).

TABLE VII
CROSS-DATASET GENERALIZATION RESULTS

Train	Test	Acc. (%)	F1	AUC
SEED	SEED	96.5	0.96	0.99
SEED	DEAP	88.2	0.87	0.92
DEAP	DEAP	91.4	0.90	0.95

matrix also proves high levels of classwise accuracy, with the approximation of 98.6% of negative, 100% of neutral, and 93.7% of positive emotions, which points to high levels of reliability and balanced performance of emotion recognition.

Fig. 9 indicates ROC-AUC of the model proposed in each of the emotion classes and indicates almost perfect classification. The neutral category has the highest score 1.0000 as the best separability, whilst the negative and positive categories score very high of 0.9971 and 0.9967, respectively. These high values are consistently high indicating that the model can discriminate well between emotional states with very low false-positive and false-negative levels, and therefore, it has a good level of discriminative and generalization.

The cross-dataset test in Table VII indicates that ahvEEGNet can be kept at a high level of functioning even with considerable domain variations. Despite a moderate accuracy decrease after moving the SEED to the DEAP, the model keeps the discriminative ability steady, meaning that hierarchical variational regularization is able to limit dataset-specific overfitting and enhance representation generalization. The performance improvement was statistically validated for all the LOSO folds to test the robustness of the claims. The mean performance, standard deviation (SD) and 95% confidence interval (CI) were calculated for each evaluation metric

$$CI = \bar{x} \pm 1.96 \frac{s}{\sqrt{n}} \quad (24)$$

Mean metric value, denoted by $\bar{x}$, represents the average value across the folds, while s stands for the standard deviation value across the folds and n is the number of folds used in the LOSO procedure. In addition, the performance scores were tested for statistical significance between ahvEEGNet and the top baseline methods, one-by-one. The resulting p -values in Table VIII suggest the improvements in performance that the ahvEEGNet provides statistically significant, and therefore not a random occurrence ($p < 0.05$).

The experimental result shows that deterministic CNN-based models cannot generalize subjects because of overfitting to

TABLE VIII
STATISTICAL VALIDATION OF CLASSIFICATION PERFORMANCE

Method	Accuracy (%)	95% CI	p -value
EEGNet	88.4 ± 1.8	[86.9, 89.9]	–
Attention-EEGNet	93.1 ± 1.4	[91.9, 94.3]	0.018
VAE-EEGNet	92.0 ± 1.5	[90.7, 93.3]	0.011
ahvEEGNet	96.5 ± 1.1	[95.6, 97.4]	–

the subject-specific EEG properties. Uncertainty-aware embeddings obtained through incorporation of variational learning are known to smoothen latent distributions and increase stability. The attention mechanism also boosts performance by providing performance by dynamically laying stress on emotion-relevant temporal-spatial patterns. A combination of these elements allows the ahvEEGNet to attain similar levels of improvements with a similar computational efficiency of lightweight EEGNet models. To further examine the influence of attention location, we created additional analyses of the impact of alternative architectural configurations, such as attention prior to variational. The two components are: 1) encoding and 2) parallel attention and variational branches. Experimental results suggest that a preattention application of variational modeling results in There was a reduction in the generalization of the task, presumably because of attention to noisy or subject-specific patterns. Likewise, parallel configurations exhibit less because there was not much interaction between uncertainty modeling and performance, feature weighting. Contrary to that, the proposed postvariational attention achieves the best performance, showing that operating on uncertainty-aware latent The robustness and discriminative capacity of the representations are enhanced.

A. Potential Application Scenarios

The proposed ahvEEGNet architecture can be utilized in several real-world EEG-based systems. For wearable BCI, strong emotion recognition has the potential to support adaptive interaction between the user and the intelligent device. The framework could have applications in health care settings, including continuous monitoring of emotional states for stress management, depression screening, cognitive assessment, and neurofeedback-based treatment. The model can also be used in affective computing systems to create emotion-aware learning platforms, smart virtual assistants, and emotion-aware entertainment applications. The lightweight design and superior cross-subject performance of ahvEEGNet enable it to serve as a practical building block for the development of future emotion-aware systems in real-world scenarios.

The experimental results always show that the combination of hierarchical variational learning and attention-based feature aggregation significantly boosts the classification accuracy and generalization performance. The conclusion summarizes the most important contributions, implications, limitations, and future research directions based on these observations.

VII. CONCLUSION

This article introduced ahvEEGNet, a hierarchical variational EEGNet scheme based on attention to achieve sound EEG-based emotion recognition. The proposed model can

solve the problems of intersubject variation and noisy EEG data by combining lightweight spatial–temporal feature extraction methods with uncertainty-aware variational encoding and multiple-heads attention. Strict cross-subject validation protocol experimental assessment shows that a representative result of a strict cross-subject validation protocol indeed indicates a better performance of a representative baseline model than that of the ahvEEGNet, as demonstrated by better performance metrics. This leads to the conclusion that probabilistic representation learning together with adaptive attention mechanisms is a credible and computationally optimal approach to emotion recognition based on EEG signals.

VIII. LIMITATION AND FUTURE SCOPE

Although the proposed ahvEEGNet framework was able to perform quite well, there are still some shortcomings. The model is mainly tested on publicly available benchmark datasets recorded in controlled experimental environments, which might not be representative of the variability of real-world EEG signals. The hierarchical variational learning is effective against intersubject variability, while the generalization over unseen datasets, recording devices, and environmental conditions is not guaranteed. Moreover, the model might be affected by the distribution shift and noise properties that are not in the training dataset. Future work will be dedicated to improving generalization by evaluating on more different datasets, incorporating domain adaptation methods and self-supervised learning methods.

Incorporating domain adaptation and transfer learning techniques for improving the generalization ability of ahvEEGNet for heterogeneous EEG datasets, acquisition devices, and recording environments will be the subject of future research. Furthermore, representation learning methods, such as self-supervised and contrastive representation learning, will be explored to decrease reliance on large labeled EEG datasets and enhance the robustness against distribution shifts. The other direction of grand potential research that is also promising is to couple the proposed hierarchical variational framework with GNNs to explicitly model dynamic functional connectivity among EEG channels. In addition, explainable artificial intelligence techniques will be integrated to make the attention maps more neurophysiologically interpretable and enhance the transparency of emotion classification decisions. Lastly, compression techniques, federated learning, and deployment of edge-AI in compressing the model for real-time emotion recognition on wearable and resource-limited EEG devices will be discussed.

REFERENCES

- [1] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1995.
- [2] T. S. Rached and A. Perkusich, "Emotion recognition based on brain computer interface systems," in *Brain–Computer Interface Systems: Recent Progress and Future Prospects*, 2013, pp. 253–270.
- [3] P. Bhattacharya, R. K. Gupta, and Y. Yang, "Exploring the contextual factors affecting multimodal emotion recognition in videos," *IEEE Trans. Affect. Comput.*, vol. 14, no. 2, pp. 1547–1557, 2021.
- [4] K. Boehner, R. DePaula, P. Dourish, and P. Sengers, "How emotion is made and measured," *Int. J. Hum.–Comput. Stud.*, vol. 65, no. 4, pp. 275–291, 2007.
- [5] P. Samal and M. F. Hashmi, "Role of machine learning and deep learning techniques in EEG-based BCI emotion recognition system: A review," *Artif. Intell. Rev.*, vol. 57, no. 3, 2024, Art. no. 50.
- [6] A. Jain and A. Jain, "AI-based emotion detection system in healthcare for patient monitoring," in *Proc. Generative Artif. Intell. Biomed. Smart Health Informat.*, 2025, pp. 455–470.
- [7] S. Machado et al., "EEG-based brain–computer interfaces: An overview of basic concepts and clinical applications in neurorehabilitation," *Rev. Neurosci.*, vol. 21, no. 6, pp. 451–468, 2010.
- [8] X. Li et al., "EEG-based emotion recognition: A tutorial and review," *ACM Comput. Surv.*, vol. 55, no. 4, 2022, Art. no. 79.
- [9] E. P. Torres et al., "Emotion recognition related to stock trading using machine learning algorithms with feature selection," *IEEE Access*, vol. 8, pp. 199719–199732, 2020.
- [10] L. Shu et al., "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, 2018, Art. no. 2074.
- [11] A. Khosla, P. Khandnor, and T. Chand, "A comparative analysis of signal processing and classification methods based on EEG signals," *Biocybern. Biomed. Eng.*, vol. 40, no. 2, pp. 649–690, 2020.
- [12] T. B. Alakuş and I. Türkoğlu, "EEG-based emotion estimation with different deep learning models," in *Proc. Int. Conf. Comput. Sci. Eng. (UBMK)*, 2019, pp. 33–37.
- [13] M. Nankya, R. Chataut, and R. Akl, "Security and privacy in E-health systems: A review of AI and machine learning techniques," *IEEE Access*, vol. 12, pp. 148796–148816, 2024.
- [14] N. Ayari et al., "Hybrid model-based emotion contextual recognition for cognitive assistance services," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3567–3576, 2022.
- [15] J. Li et al., "Multisource transfer learning for cross-subject EEG emotion recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3281–3293, Jul. 2019.
- [16] L. Chen et al., "Adaptive feature selection-based AdaBoost-KNN for dynamic emotion recognition," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 5, no. 2, pp. 205–213, Feb. 2019.
- [17] M.-P. Hosseini, A. Hosseini, and K. Ahi, "Machine learning for EEG signal processing in bioengineering: A review," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 204–218, 2020.
- [18] V. J. Lawhern et al., "EEGNet: A compact convolutional network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, 2018, Art. no. 056013.
- [19] A. Zancanaro et al., "CNN-based approaches for cross-subject classification in motor imagery," in *Proc. IEEE CIBCB*, 2021, pp. 1–7.
- [20] N. M. N. Leite et al., "Deep convolutional autoencoder for EEG noise filtering," in *Proc. IEEE BIBM*, 2018, pp. 2605–2612.
- [21] X. Du et al., "Fast reconstruction of EEG signal compression sensing based on deep learning," *Sci. Rep.*, vol. 14, 2024, Art. no. 5087.
- [22] N. Chen et al., "iVAE: An improved deep learning structure for EEG signal characterization," in *Proc. IEEE BIBM*, 2020, pp. 1909–1913.
- [23] A. Z. Al-Marridi et al., "Convolutional autoencoder approach for EEG compression," in *Proc. IEEE IWCMC*, 2018, pp. 370–375.
- [24] F. Topic and M. Russo, "Emotion recognition based on EEG feature maps through deep learning," *Eng. Sci. Technol.*, vol. 24, no. 6, pp. 1442–1450, 2021.
- [25] G. Bao, B. Yang, and L. Tong, "Linking multi-layer dynamical GCN with style-based recalibration CNN for EEG-based emotion recognition," *Front. Neurobot.*, vol. 16, 2022, Art. no. 834952.
- [26] D. Kuang and C. Michoski, "Attention with kernels for EEG-based emotion classification," *Neural Comput. Appl.*, vol. 36, no. 10, pp. 5251–5266, 2024.
- [27] Y. Geng, S. Shi, and X. Hao, "Deep learning-based EEG emotion recognition: A comprehensive review," *Neural Comput. Appl.*, vol. 37, no. 4, pp. 1919–1950, 2025. [Online]. Available: <https://doi.org/10.1007/s00521-024-10821-y>
- [28] E. H. Houssein, A. Hammad, and A. A. Ali, "Human emotion recognition from EEG-based brain–computer interface using machine learning: A comprehensive review," *Neural Comput. Appl.*, vol. 34, no. 15, pp. 12527–12557, 2022.
- [29] Y. H. Choi et al., "A deep learning approach for classifying rapid eye movement sleep behavior disorder using EEGNet," *J. Clin. Neurol.*, vol. 22, no. 3, pp. 303–312, 2026. [Online]. Available: <https://doi.org/10.3988/jcn.2025.0552>
- [30] Y. Wang, X. Ju, J. Sun, Y. Yu, M. Li, and D. Hu, "Improved EEGNet with a multilevel spatial feature extraction module for EEG decoding," *IEEE Trans. Instrum. Meas.*, 74, 2025, Art. no. 2519512, doi: [10.1109/TIM.2025.3538466](https://doi.org/10.1109/TIM.2025.3538466).

- [31] M. M. Hassan et al., "Efficient prediction of coronary artery disease using machine learning algorithms with feature selection techniques," *Comput. Elect. Eng.*, vol. 115, 2024, Art. no. 109130.
- [32] R. Priyadarshi et al., "Exploring the frontiers of unsupervised learning techniques for diagnosis of cardiovascular disorder: A systematic review," *IEEE Access*, vol. 12, pp. 139253–139272, 2024.
- [33] A. D. Patel, R. H. Jhaveri, and R. S. Rathore, "Healthcare on smart wheels—A review of intelligent vehicles assisted healthcare systems," in *Proc. 4th Int. Conf. Distrib. Sens. Intell. Syst.*, 2023, pp. 83–89.
- [34] A. Alghamdi and I. Keshta, "Blockchain consensus mechanisms and enhancement techniques for federated learning-based intrusion detection systems in IoT smart homes," *J. Reliable Secure Comput.*, vol. 2, no. 1, pp. 1–26, 2026.
- [35] P. Malik et al., "Enhancing intracranial hemorrhage diagnosis through deep learning models," *Procedia Comput. Sci.*, vol. 235, pp. 1664–1673, 2024.
- [36] J. Pan et al., "ST-SCGNN: A spatio-temporal self-constructing graph neural network for cross-subject EEG-based emotion recognition and consciousness detection," *IEEE J. Biomed. Health Informat.*, vol. 28, no. 2, pp. 777–788, Feb. 2024, doi: [10.1109/JBHI.2023.3335854](https://doi.org/10.1109/JBHI.2023.3335854).
- [37] A. Nalwaya, K. Das, and R. B. Pachori, "An automated framework for human emotion detection from multichannel EEG signals," *IEEE Sensors J.*, vol. 24, no. 13, pp. 20920–20927, Jul. 2024, doi: [10.1109/JSEN.2024.3396739](https://doi.org/10.1109/JSEN.2024.3396739).
- [38] Z. Jia, Y. Ouyang, X. Kong, Y. Guo, Z. Li, and H. Zeng, "A novel dual-task model for EEG-based emotion and cognition recognition," *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art. no. 2503314, doi: [10.1109/TIM.2024.3521360](https://doi.org/10.1109/TIM.2024.3521360).
- [39] T. D. T. Mai, D.-Q. Vu, H. N. Phuong, and T.-N. Phung, "EEG-SKNet: A self-knowledge distillation model with scaled weights for emotion recognition from EEG signals," *IEEE Access*, vol. 13, pp. 140239–140250, 2025, doi: [10.1109/ACCESS.2025.3528292](https://doi.org/10.1109/ACCESS.2025.3528292).
- [40] S. Gong et al., "LEL: Lipschitz continuity constrained ensemble learning for efficient EEG-based intra-subject emotion recognition," *IEEE Sensors J.*, vol. 26, no. 9, pp. 13446–13456, May 2026, doi: [10.1109/JSEN.2026.3672658](https://doi.org/10.1109/JSEN.2026.3672658).



Vinay Kumar Singh received the Master of Technology degree from the Indian Institute of Information Technology, Design, and Manufacturing Jabalpur, Jabalpur, India.

He is currently a Research Scholar with the University of Allahabad, Allahabad, India. His academic journey has been characterized by a strong commitment to exploring the intersections of technology and health, particularly through the application of signal processing techniques to biomedical data. His research fo-

cuses on various advanced fields, including signal processing, biomedical engineering, brain-computer interfaces, artificial intelligence, and machine learning.



Bharati Rathore is currently a Lecturer with the Faculty of Business and Creative Industries, University of South Wales, Cardiff, U.K. She was a Lecturer of Fashion Business and Promotions with the School of Fashion and Textiles, Birmingham City University, Birmingham, U.K., and also with the Faculty of Business and Creative Industries, University of West of Scotland (Paisley), Paisley, Scotland. She has several years of rich experience in quality of teaching, learning, and research excellence. She has authored or

coauthored several articles in top tier journals and participated in various flagship International Conference.

Ms. Rathore is the Fellow of the Higher Education Academy U.K.



Raja Saeed Iqbal received the postgraduation degree in business management, strategy, and consumer behavior from the University of Suffolk, Ipswich, U.K.

He is currently specializing in the psychological impact of artificial intelligence in marketing. He has an experience in retail management, including training new managers in operational and back-office systems within a supermarket retail project. He aims to engage in strategic consultancy, focusing on AI, consumer psychology, and digital transformation to bridge research and industry practice. His research interests include how AI-driven technologies shape consumer perception, trust, decision making, and brand relationships, and combining psychological theory with emerging marketing technologies.



Dinesh Prasad Sahu (Member, IEEE) received the Ph.D. degree in computer science and engineering from JNU, New Delhi, India, in 2019.

He is currently an Assistant Professor with the School of Computer Science, Engineering, and Technology, Bennett University, Greater Noida, India. He has maintained an outstanding academic record throughout his education. He has authored more than 20 articles. His research interests include cloud computing, fog computing, edge computing, mist computing, and generative AI.



Shiv Prakash received M.Tech. and Ph.D. degrees from Jawaharlal Nehru University, New Delhi, India.

He has an excellent educational record throughout. His current research interests include the Internet of Connected Vehicles, electronic vehicles, artificial intelligence, Big Data analytics, biometric security, cloud computing, computer networks, machine learning, network security, and IoT use cases of Sensor Networks.

He has authored or coauthored 75+ research papers in various peer-reviewed international journals and transactions (including IEEE, Taylor and Francis, Elsevier, Springer, Wiley, American Scientific Publishers, etc.) and around 45+ research papers in proceedings of various peer-reviewed conferences in India and abroad.

Dr. Prakash is currently a Reviewer of several referred International journals of repute (including ACM, IEEE, Taylor and Francis, Elsevier, Springer, Wiley, etc.).



Vishal Krishna Singh (Member, IEEE) received the bachelor's degree in information technology, the master's degree in computer technology and application, and the Ph.D. degree in information technology from the Indian Institute of Information Technology, Allahabad, India, in 2010, 2013, and 2018, respectively.

He is currently a Lecturer with the Networks and Communications Research Group, School of Computer Science and Electronics Engineering, University of Essex, Colchester, U.K. His research interests include the Internet of Things, wireless sensor networks, in-network inference, machine learning, and data analytics.