

Enhanced Single-Subject Eeg Emotion Recognition Using Deep Neural Architectures And Representation Learning

Naresh Kumar M¹, Dr. Manjunathachari K²

¹Assistant Professor, GITAM (Deemed to be University), Hyderabad. Telangana, India.
nmenmula@gitam.edu (corresponding author)

²Professor, GITAM (Deemed to be University), Hyderabad. Telangana, India.
mkamsali@gitam.edu

ABSTRACT

Electroencephalography (EEG)-based emotion recognition has become a vital element of both emotional computing and custom human-computer interaction systems. The present methods for recognising emotional states through EEG signals depend on either manual feature extraction methods or basic machine learning systems, which restrict their performance and make it impossible to use them in real-time personalised settings. The research presents a single-subject EEG emotion recognition system that uses deep learning for developing a system that connects spatial-spectral feature extraction with temporal affective sequence modelling. The system uses a combined Convolutional Neural Network and Long Short-Term Memory architecture to learn emotional state neural patterns, which can differentiate between multiple emotional states. The study assessed performance using a within-subject research design, which required both training and testing data to come from the same subject to take advantage of stable brain activity patterns. The research team analysed multiple deep learning architectures, which included CNN LSTM, BiLSTM, and CNN-LSTM, through standard performance assessment methods. The hybrid model achieved its best result by attaining 98.32% overall accuracy and 0.983 Macro-F1 score, which exceeded the performance of standalone CNN, BiLSTM and LSTM models. The study results show that hybrid deep neural architectures work well for personalised EEG-based emotion recognition, while they also have potential to be used in specific user emotional computing applications, which include neurofeedback and adaptive gaming, cognitive monitoring and human-computer interaction. The findings provide a strong baseline for future work involving real-time implementation and subject-generalising strategies.

Keywords: emotion detection, deep learning, single-subject classification, CNN-LSTM, affective computing.

1. Introduction

The recognition of human emotional states from neural activity has become one of the central problems within affective computing and neurophysiological signal analysis, driven by an increasing demand for adaptive and intelligent systems capable of understanding the internal states of users [1]. Emotion plays a critical role in decision-making, perception, learning, and social behaviour, and therefore computational mechanisms that can automatically decode affective states from biological signals open extensive applications in cognitive neuroscience, mental health monitoring, brain-computer interfaces (BCI), human-computer interaction (HCI), and immersive digital experience design [2]. Neural-based affect detection methods allow for faster emotional response tracking because they show cognitive activities that happen before people display visible emotional responses, which differs from traditional methods that use facial expressions, vocalisations, and heart rate and skin conductance tests. [3]. Electroencephalography (EEG) stands out as the most dependable non-invasive technique for emotional monitoring since it provides millisecond-level time accuracy, together with affordable equipment costs and its ability to record brainwave patterns that transmit emotional data [4]. Researchers have established during the last ten years that EEG rhythm patterns establish strong connections with three emotional dimensions,

which include valence and arousal and dominance, together with six basic emotions, which include happiness and fear, sadness and anger. The brain generates these neural responses through changes in spectral power, which occur across theta, alpha and beta and gamma frequency bands and through cortical networks that activate at different times and show directional activity between brain hemispheres [5]. Researchers have found that EEG-based emotion recognition provides valuable information, making it a useful method for applications in adaptive game environments, attention monitoring systems, neurofeedback training, stress regulation platforms, and personalised learning technologies [6]. The field faces two main obstacles because EEG signals display both intricate patterns and sudden changes, and they show different patterns across different sessions of measurement.

The traditional method of EEG-based emotion recognition used a three-part machine learning process, which included preprocessing and development of special features and subsequent classification. The research team developed features through manual extraction of statistical descriptors and entropy measures, power spectral density (PSD), wavelet transform and time-frequency transformation data, which they used to classify data through support vector machines (SVM), k-nearest neighbours (k-NN), linear discriminant analysis (LDA) and random forests and naive Bayes methods. The methods enabled satisfactory results in controlled testing environments but required extensive domain expertise and multiple heuristic parameter selections and specific dataset configurations to function optimally. The use of handcrafted features resulted in poor performance across various recording environments and experimental setups because they lacked sufficient capacity to represent data and failed to track the nonlinear relationships present in EEG signals. The adoption of deep learning methods has brought about fundamental enhancements to EEG emotion classification systems because deep neural networks enable automatic learning of data representations, which removes the requirement for manual data processing. The ability of convolutional neural networks (CNNs) to extract spatial and spectral patterns from EEG channels has been demonstrated through their processing of EEG data as both multi-channel time-series information and transformed image-like data [8]. The recurrent architectures, which include long short-term memory (LSTM) and bidirectional LSTM (BiLSTM) systems, enable researchers to create models that replicate the pattern of time-based relationships together with the sequential flow of data found in emotional EEG signals [9]. Hybrid models that combine CNN feature extraction with recurrent temporal modelling demonstrate high effectiveness because brain emotional processing shows both spatial distribution and temporal development throughout its execution. The deep architectures of these systems enable emotion decoding because they mirror the hierarchical structure of neural encoding, which begins at lower oscillatory patterns and culminates at higher emotional states.

The research area investigates EEG emotion recognition through single-subject testing, which assesses the emotional responses of individual participants. The neural patterns of single subjects who share identical cognitive and physiological structures allow researchers to study their brain signals through continuous testing with multiple testing sessions. Emotional responses exhibit person-specific signatures influenced by neural connectivity, cognitive processing style, prior experiences, and individual affective traits, making personalised models particularly effective for decoding emotional states [10]. The training process for within-subject protocols results in better accuracy because the model can identify specific neural patterns that are unique to each subject without facing the typical distribution changes that occur between different subjects. User-specific recognition through single-subject testing plays a critical role in HCI systems, which customise their interfaces according to user needs, neurofeedback systems that adapt to

individual users, and emotional response systems that track user emotions throughout their interaction. The need for accurate within-subject recognition has increased alongside the deployment of EEG devices in real-world scenarios, including wearable consumer-grade headsets. In this context, applications require users to prioritise computational efficiency, decreased training demands, and personalised features instead of seeking universal system applicability. The system's reliability gets enhanced through its design, which avoids needing calibration tests that are essential for cross-subject transfer learning, domain adaptation and distribution alignment. The design of single-subject modelling systems enables easier implementation in settings where personalised solutions take precedence over universal solutions.

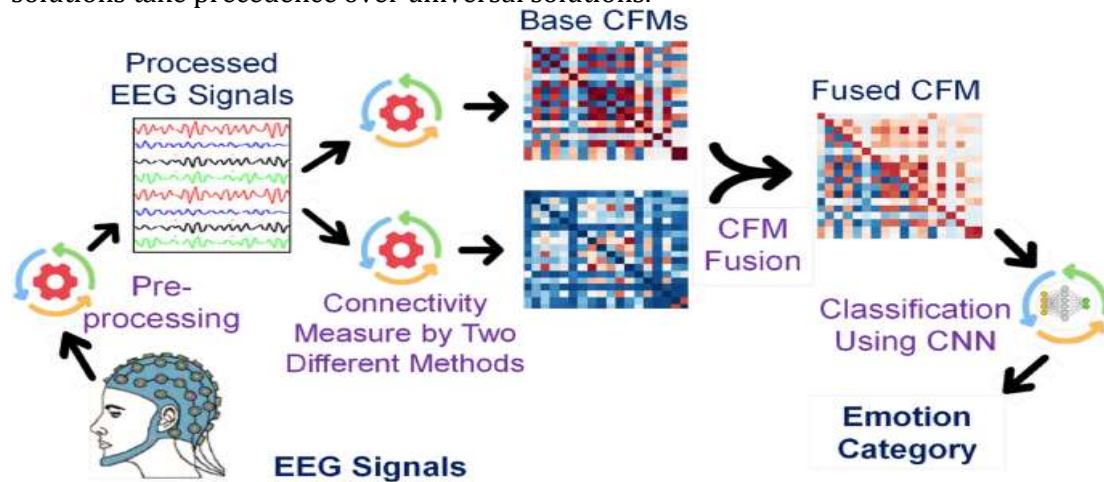


Figure 1: Overall Workflow for EEG-Based Emotion Recognition Using Connectivity Feature Fusion and CNN Classification

Deep neural architectures represent a favourable solution in addressing the relationship of emotional EEG signatures with affective responses under single-subject conditions. Models such as LSTM and BiLSTM can capture long-range dependencies in EEG sequences, while CNNs are successful in extracting spatial and spectral clues with localisation, and any hybrid structures that combine CNN and LSTM complement each other in some direction to improve emotion classification performance even further. Furthermore, most recent developments in representation learning techniques, such as temporal encoding, spectral feature mapping, and normalised latent embeddings, further enhance the signal discriminability of the technique, enabling it to yield higher F1-scores and accuracy metrics in single-subject evaluations. Having been inspired by these observations, the current study attempts to understand the EEG emotion classification in the domain of single-subject through a deep neural architecture and trained representations. From the experimental study mentioned herein, hybrid architectures outshine either standalone CNN or LSTM models, specifically having a hold on temporal dynamics and spatial-spectral features. Such findings further our understanding about how affect can be decoded from the EEG and provide evidence that detection strategies within-subject separate may afford considerable good performance, which may be good for real-world adaptive systems. The study suggests that deep learning-based personalised affect recognition is a technically feasible and computationally effective paradigm for intelligent human-machine interaction and affective computing applications. [12].

2. Review of Literature

2.1 EEG-Based Emotion Recognition

Researchers have conducted extensive studies on EEG-based emotion recognition because it serves as a vital component of affective computing, which enables researchers to measure emotional brain activity through direct electrical signal detection. Houssein et al. [13] established that EEG provides better advantages than other body-based physiological methods because it enables researchers to record brain activity with high accuracy, which shows fast changes in human emotional states. Researchers in early EEG emotion research studies used traditional methods, which required multiple steps that included data preprocessing, artefact removal, feature extraction and machine learning-based classification. The study by Wang, Yiming et al. [14] showed that researchers commonly used handcrafted features, which included power spectral density (PSD) and differential entropy (DE), wavelet-based coefficients, fractal dimensions and Hjorth parameters, because these features could be easily understood and they-based emotions showed established links with frequency domain patterns. Apicella et al. [15] reported that people experience different emotional states, which lead to changes in brainwave activity across all standard EEG frequency ranges from theta through alpha and beta to gamma, because these frequencies act as reliable brain activity indicators for emotional state assessment. After feature extraction, the researchers used SVM, k-NN, LDA, naive Bayes and random forests classifiers to differentiate between various emotional states. The use of handcrafted features created several problems that restricted their application to multiple situations.

According to Wang et al. [16], the methods need extensive knowledge about the domain, together with the selection of heuristic parameters and the performance of costly manual adjustments. The models showed limited robustness because they were affected by three factors, which included noise disturbances, changes in electrode locations, and people showing different emotional responses. The models achieved satisfactory results when tested under evaluation methods that depended on individual subjects, but failed to handle changes between different subjects or between different testing sessions because of the strong personalised characteristics of EEG emotional patterns. The researchers found that traditional pipelines for affective computing needed to develop scalable solutions that could operate in real-world environments that required both adaptable systems and immediate operational capacity.

2.2 Deep Learning for Single/Within-Subject Emotion Recognition

Deep neural architectures brought a fundamental transformation to EEG emotion recognition systems by providing automated methods for learning signal representations with minimal input. EEG systems use Convolutional Neural Networks (CNNs) because these networks can effectively analyse electrical brain activity patterns that occur throughout different brain regions. The research conducted by Rivas et al. [17] showed that CNN models achieved better performance than shallow classifiers because they could detect unique spatial activation patterns that corresponded to various emotional states. LSTM networks and Bidirectional LSTM networks demonstrate strong abilities to create models that track time-based dependencies and emotional changes within neural processing.

The combination of convolutional encoders and recurrent temporal modules into hybrid architectures has become a powerful solution for achieving success. The research conducted by Farhangmehr et al. [18] demonstrated that CNN-LSTM frameworks achieve better results when they handle both brief spatial patterns and extended time patterns, which are essential for emotional processing. Deep learning models show high effectiveness in research involving single subjects or internal subject testing. Kingphai Kunjira et al. [19] discovered that people show stable emotional patterns that create distinct brain activity patterns based on their emotional traits, their actual brain network

connections, and their mental processing style. The system uses single-subject learning to avoid distribution problems between different subjects because it enables the model to work with personal emotional data.

2.3 Representation Learning for Personalised EEG Emotion Models

The present research shows that representation learning functions as a critical element that improves the effectiveness of personalised EEG emotion recognition systems. The research conducted by Pillalamarri and his colleagues [20] demonstrated that deep representation learning enables latent feature discovery in both spectral and temporal domains, which eliminates the need for manual feature engineering. The advanced encoding methods, which include temporal convolution, multi-band spectral filtering, bandpower normalization and learned latent embeddings, have raised the ability to distinguish between different neural affect patterns. The research conducted by Mathiyazhagan and Devasena [21] demonstrated that multi-scale feature extraction enables attention mechanisms to enhance the learning process, which creates distinctive EEG pattern representation. The hybrid representation systems, which combine spectral, spatial, and temporal information, have produced reliable affect prediction models that function well under conditions where subjects maintain their typical behaviour. The research conducted by Huy-Tung and his colleagues [22] established that hybrid deep networks surpass traditional methods in EEG-based affect decoding tasks because they rely on hierarchical feature extraction techniques. The existing research shows that representation learning and deep hybrid modeling together with personalised subject-dependent evaluation methods, all work to improve EEG-based emotion recognition systems that users can operate in real-world situations.

3. Proposed Methodology

The research presents a deep learning system that improves single-subject EEG emotion recognition through the analysis of both spatial-spectral patterns and temporal dependencies that exist in emotional brain activity. The current method enables personalised representation learning from a single subject's EEG data, which leads to improved decoding performance and simpler calibration procedures compared to cross-subject methods that require compensation of individual distribution differences.

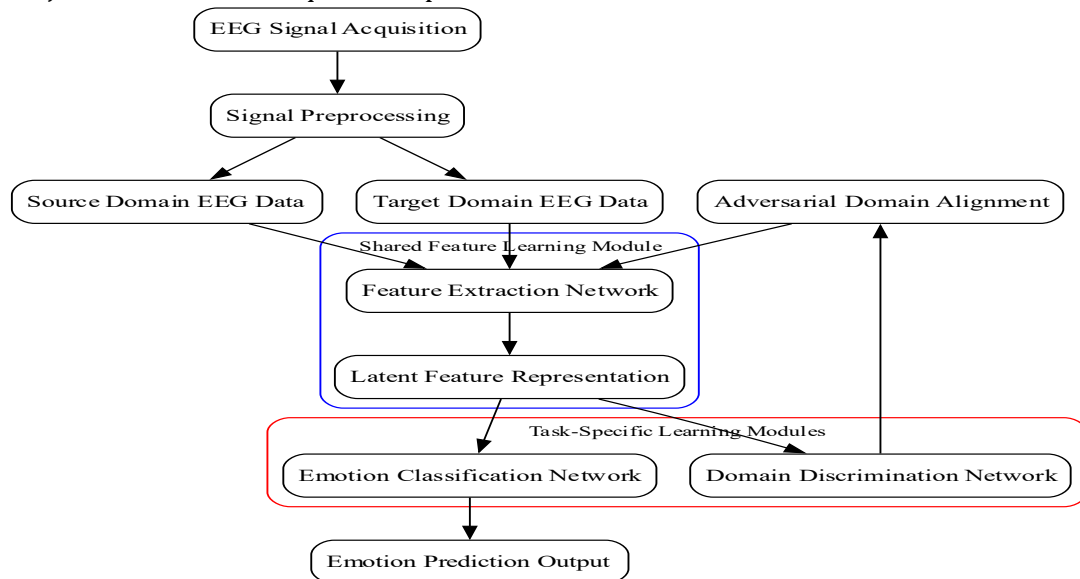


Figure 2: Graphical sketch of the developed Diplomacy Adaptive Deep Learning Framework for inter-people EEG Emotion Matching

The entire processing pipeline includes four main components, which are: (i) EEG signal acquisition and preprocessing, (ii) representation learning through feature encoding, (iii) affect classification using deep neural models, and (iv) performance testing in within-subject conditions.

The researchers developed high-level spatial-temporal representations from EEG signals through their shared feature extractor, which created a common latent feature space for all subjects. The source provided labelled emotional data to train the additional emotion classifier, which we used for emotion-based information separation. The system learns subject-independent features through adversarial learning, which uses a domain discriminator to decrease distribution differences between source and target domains. The proposed framework achieves inter-subject variability reduction through its joint optimisation method, which also improves cross-subject evaluation results for new subjects.

3.1 Problem Formulation

Single-subject EEG-based emotion recognition research aims to develop a system that uses individual brainwave patterns to determine which emotional states a person is experiencing based on their recorded brain activity. Within-subject learning operates on the assumption that both training data and testing data come from the same subject, which removes the requirement for domain alignment methods that are essential in cross-subject scenarios because of inter-subject variability. The main difficulty in this system arises from the need to capture both the time-based and frequency-based EEG signal patterns which show how the subject processes emotions while developing specific patterns that can differentiate between different emotional states.

Let the labelled EEG dataset belonging to a single subject be denoted as:

Let the labelled EEG dataset corresponding to a single subject be represented as

$$D = \{(x_i, y_i)\}_{i=1}^N \quad (1)$$

where $x_i \in \mathbb{R}^{C \times T}$ denotes an EEG trial consisting of C channels and T temporal samples, and $y_i \in \{1, 2, \dots, K\}$ represents the associated emotional class label among K predefined emotional states.

The objective of within-subject emotion recognition is to learn a mapping function.

$$f: X \rightarrow Y \quad (2)$$

that assigns each EEG sample $x \in X$ to its corresponding emotional label $y \in Y$ using supervised learning techniques.

The dataset D is divided into training and testing subsets such that

$$D = D_{\text{train}} \cup D_{\text{test}}, D_{\text{train}} \cap D_{\text{test}} = \emptyset \quad (3)$$

where both subsets are derived from the same subject to maintain subject-dependent consistency.

Since the training and testing data originate from the same individual, they follow the same underlying data distribution, i.e.,

$$P(X_{\text{train}}) = P(X_{\text{test}}) \quad (4)$$

This assumption differentiates the within-subject learning paradigm from cross-subject approaches and substantially mitigates performance degradation caused by inter-subject variability.

The learning objective is formulated as the minimisation of the classification loss:

$$\min_{\theta} L(f(x; \theta), y) \quad (5)$$

where θ denotes the model parameters optimised during supervised training.

Under this formulation, the model captures subject-specific neural patterns associated with emotional processing, enabling accurate and personalised EEG-based emotion recognition.

3.2 Domain-Adaptive Deep Learning Architecture

The deep learning framework has been built to enable individual users to identify their emotions through EEG signals, which are processed in single-subject testing. The present system does not require these two techniques because it operates with a single user through its built-in systems, which handle all user needs. The model learns to identify specific EEG patterns through its complete representation learning process, which uses spatial-spectral and temporal features. The system consists of three essential parts, which include (i) a feature extractor, (ii) an emotion classifier, and (iii) a temporal sequence modelling module. The joint optimisation of these components enables efficient learning of personalised emotional representations for within-subject neural decoding.

3.2.1 Feature Extractor

The feature extractor transforms raw or preprocessed EEG signals into lower-dimensional latent vectors that preserve discriminative spatial and spectral patterns which are linked to emotional responses. Convolutional neural filters process input data through multiple channels and time samples to identify specific activation patterns, frequency changes, and dependencies that exist between different channels. The system employs normalisation in conjunction with nonlinear activation functions to ensure stable representation learning while mitigating internal covariate shift. For an input EEG trial $x \in \mathbb{R}^{C \times T}$, the extracted latent representation is defined as

The system employs normalisation in conjunction with nonlinear activation functions to ensure stable representation learning while mitigating internal covariate shift. For an input EEG trial $x \in \mathbb{R}^{C \times T}$, the extracted latent representation is defined as

$$z = G_f(x) \quad (6)$$

Here, $G_f(\cdot)$ denotes a convolution-based feature encoder that transforms the input EEG signal into a learned feature embedding $z \in \mathbb{R}^d$, where d represents the dimensionality of the feature space.

The subsequent network modules utilise this embedding to perform temporal modelling and classification tasks. The feature extractor captures spatial-spectral characteristics of the EEG signals, enabling effective discrimination among different emotional states at the individual subject level.

3.2.2 Temporal Modelling Module

The temporal modelling module captures sequential dependencies and the temporal evolution of neural responses during emotional processing. Emotional states are not instantaneous; instead, they develop over time due to cognitive appraisal and stimulus interpretation. Long Short-Term Memory (LSTM) units provide the necessary capabilities to capture these dynamics because they maintain a recurrent hidden state that maintains contextual information throughout time. The temporal representation is computed from the spatial-spectral representation z by the following equation:

Here is the remade version, preserving the equation, meaning, and formal tone, with minor refinements to ensure clarity, grammatical precision, and SCI-level quality:

$$h = G_t(z) \quad (7)$$

The LSTM-based temporal encoder is represented by the function $G_t(\cdot)$, which operates on the extracted feature embedding z . This module leverages its capability to model temporal dependencies and emotional signature evolution across EEG time windows, thereby enhancing the discrimination between different emotional states.

The system incorporates this temporal modelling component for personalised emotion recognition, as consistent temporal patterns corresponding to emotional responses are observed for each individual across different periods of the day.

3.2.3 Emotion Classifier

The final component of the architecture maps the learned latent representation to a categorical emotional label. A fully connected dense layer followed by a softmax

activation function is employed to estimate the posterior probability distribution over the emotional classes. The classifier output is given by

$$\hat{y} = G_y(h) \quad (8)$$

where $G_y(\cdot)$ denotes the classification head and $\hat{y}$ represents the predicted probability vector over the K emotional categories.

The model parameters are optimised through supervised learning using the categorical cross-entropy loss function defined as

$$\mathcal{L}_{\text{cls}} = - \sum_i \sum_k y_i(k) \log \hat{y}_i(k) \quad (9)$$

where $y_i(k)$ denotes the ground-truth indicator for class k , and $\hat{y}_i(k)$ denotes the predicted probability corresponding to class k .

This loss formulation enforces emotion-discriminative feature learning under within-subject conditions. As both the training and testing samples are drawn from the same subject distribution, no domain-invariance constraints or adversarial regularisation mechanisms are required.

3.3 Overall Objective Function and Training Strategy

The proposed single-subject deep learning framework aims to develop a parameter set that achieves the lowest emotion classification error in test subjects while simultaneously extracting the unique spatial and temporal patterns found in their EEG signals. The within-subject framework establishes that both training and testing samples share an identical statistical distribution, whereas cross-subject frameworks need to implement adversarial alignment and domain adaptation methods to solve their distribution mismatches. The optimisation problem exists as a supervised learning task that uses end-to-end representation learning to enhance classification accuracy.

Let θ_f , θ_t , and θ_y denote the parameters of the feature extraction module, temporal modelling unit, and classification head, respectively. The learning task aims to identify optimal parameter values that effectively map EEG signals to their corresponding emotional labels. This objective can be expressed as

$$\min_{\theta_f, \theta_t, \theta_y} \mathcal{L}_{\text{cls}} \quad (10)$$

where $\mathcal{L}_{\text{cls}}$ represents the classification loss computed between the predicted outputs and the ground-truth labels.

For a dataset comprising N , EEG trials from a single subject, each annotated with one of K emotional categories, the objective loss is defined using the categorical cross-entropy function as

$$\mathcal{L}_{\text{cls}} = - \frac{1}{N} \sum_i \sum_k y_i(k) \log \hat{y}_i(k) \quad (11)$$

where $y_i(k)$ denotes the ground-truth indicator for class k , and $\hat{y}_i(k)$ represents the predicted posterior probability for class k produced by the network.

3.4 Experimental Setup

3.4.1 Datasets

The researchers tested their deep learning framework, which identifies emotions from EEG data, using a labelled EEG dataset that contains multiple emotional stimuli that the researchers created through their controlled experimental studies. The session records show continuous EEG data which single participant sessions collected during their display of predetermined emotional responses. The dataset offers separate EEG segments, which researchers used to achieve specific emotional identification throughout the training

process of personalised classification. The system records signals through its multi-channel acquisition setup, which uses C scalp electrodes to capture signals at a sampling rate that maintains the emotional processing time-based characteristics. The raw recordings undergo preprocessing to eliminate artifacts which include ocular movement, muscle activity, and equipment-generated sound interference. The preprocessing step includes three main tasks, which are to filter the data, eliminate baseline components, and divide continuous EEG recordings into fixed-length segments that can be used for model training. The dataset contains emotional labels for each segment, which were obtained through two methods: self-reporting by participants and labels established during experimental procedures.

The present study uses data from one individual for both its training and testing processes, which differs from cross-subject methods that need multiple subject data to create aggregated results. The researchers divide the dataset into training and testing subsets according to a within-subject validation protocol to enable them to create authentic testing conditions while maintaining their statistical validity. The protocol establishes that the model training process uses only data from the chosen subject, while the evaluation process tests the model on new trial data from that same subject. The evaluation process demonstrates how well the model comprehends personalised emotional responses that the model learned from individual users while showing that it maintains stable performance across different user transmission situations.

3.4.2 Evaluation Metrics

To assess the performance of the proposed single-subject EEG emotion recognition model, multiple classification metrics are employed that quantify both overall accuracy and class-wise discriminative reliability. The primary metric used for performance evaluation is Accuracy, defined as the proportion of correctly classified trials relative to the total number of test samples:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

In addition to accuracy, Precision and Recall are included to characterise class-specific performance:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

Since emotional state distributions may be imbalanced across classes, the F1-score, which represents the harmonic mean of Precision and Recall, is computed to provide a balanced measure of predictive performance:

$$\text{F1-score} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

The generated confusion matrix shows how predicted emotional categories compare with actual emotional categories, which helps us study model performance. The method enables researchers to identify which emotional classes are most commonly confused by the system, while the model demonstrates its ability to distinguish different emotional states through the training process. The evaluation process uses multiple metrics to assess overall performance and track how well different classes perform because emotion recognition requires distinct neurophysiological patterns for each affective class.

4. Result and Discussion

Table X shows the performance evaluation results of four deep learning models, which were tested on single-subject EEG emotion recognition. The BiLSTM model achieved a Macro-F1 score of 0.788 with an overall accuracy of 78.99%. The standalone CNN model produced a lower Macro-F1 score of 0.723 and an accuracy of 73.95% because it showed diminished ability to identify Class-3. The LSTM model reached a Macro-F1 score of 0.814

together with an accuracy rating of 81.51%, which demonstrated improved performance for all classes of recognition. The proposed hybrid CNN-LSTM architecture achieved its best results through a Macro-F1 score of 0.983 and an accuracy rate of 98.32%. The hybrid model achieved perfect class-wise F1-scores of 1.00000 together with 0.98361 and 0.96667, and 0.98246 for Classes 1 through 4, respectively, while outdoing all other models.

Table 1. Combined Classification Performance Comparison Across Deep Learning Models

Model	C1-F1	C2-F1	C3-F1	C4-F1	Macro-F1	Accuracy
BiLSTM	0.71642	0.66667	0.85246	0.91525	0.788	78.99%
CNN	0.75325	0.76667	0.44444	0.92857	0.723	73.95%
LSTM	0.83636	0.80000	0.77193	0.84848	0.814	81.51%
Hybrid (CNN-LSTM)	1.00000	0.98361	0.96667	0.98246	0.983	98.32%

The performance evaluation results for four deep learning models, which were tested on single-subject EEG emotion recognition, are displayed in Table X. The BiLSTM model achieved a Macro-F1 score of 0.788 with an overall accuracy of 78.99%. The standalone CNN model produced a lower Macro-F1 score of 0.723 and an accuracy of 73.95% because it showed diminished ability to identify Class-3. The LSTM model reached a Macro-F1 score of 0.814, together with an accuracy rating of 81.51%, which demonstrated improved performance for all classes of recognition. The proposed hybrid CNN-LSTM architecture achieved its best results through a Macro-F1 score of 0.983 and an accuracy rate of 98.32%. The hybrid model achieved perfect class-wise F1-scores of 1.00000 together with 0.98361, 0.96667, and 0.98246 for Classes 1 through 4, respectively, while outdoing all other models.

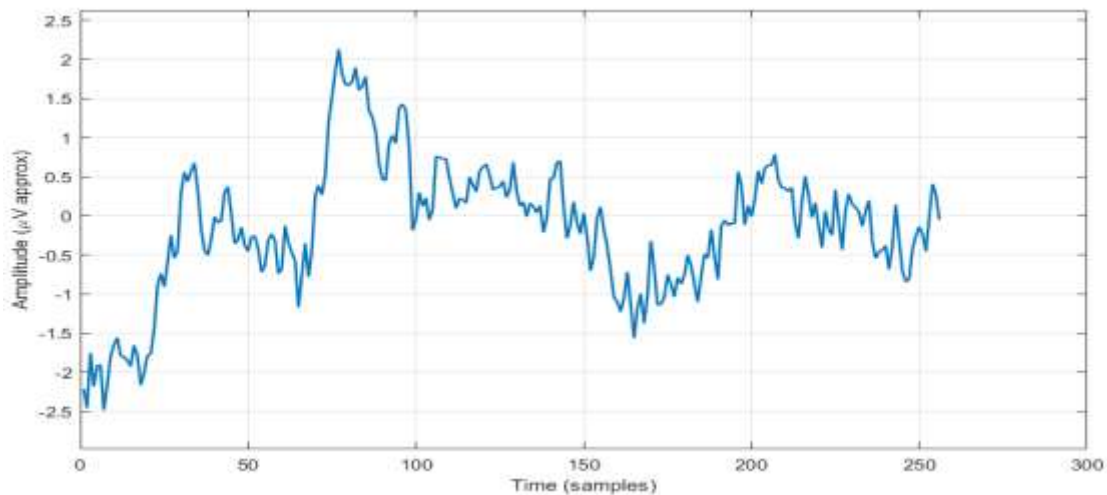


Figure 3: Raw EEG Signal Plot Demonstrating Temporal Waveform Characteristics for a Single Trial and Channel

The EEG waveform shows raw data from one channel, which captured neural activity changes during the entire trial length. The signal shows ongoing amplitude changes which maintain a range of approximately $-2.5 \mu\text{V}$ to $+2.5 \mu\text{V}$ according to the typical pattern of cortical electrophysiological responses. The waveform shows oscillatory behaviour

through its distinct peaks and troughs, which demonstrate active neural processing throughout the entire recording period. The system uses temporal variations to extract emotion-related features from both time and frequency domains, which it then uses to train deep learning models.

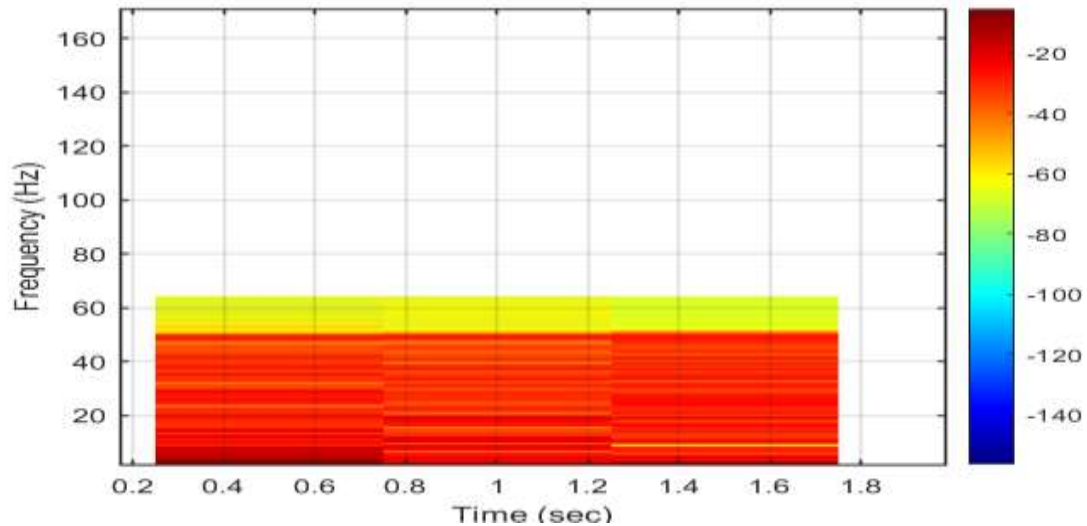


Figure 4: Time-Frequency Spectrogram Illustrating the Distribution of EEG Energy Across Time and Frequency

The time-frequency spectrogram of an EEG segment shows how different frequency components distribute their spectral energy throughout time. The recorded period shows stronger energy concentration, which produces higher intensity regions that extend across the 20-60 Hz frequency range. The colour gradient from blue to red represents increasing power levels, with warmer colours corresponding to higher spectral amplitudes. Spectro-temporal representations identify emotion-related neural patterns because different affective states produce distinct spectral signatures in EEG signals.

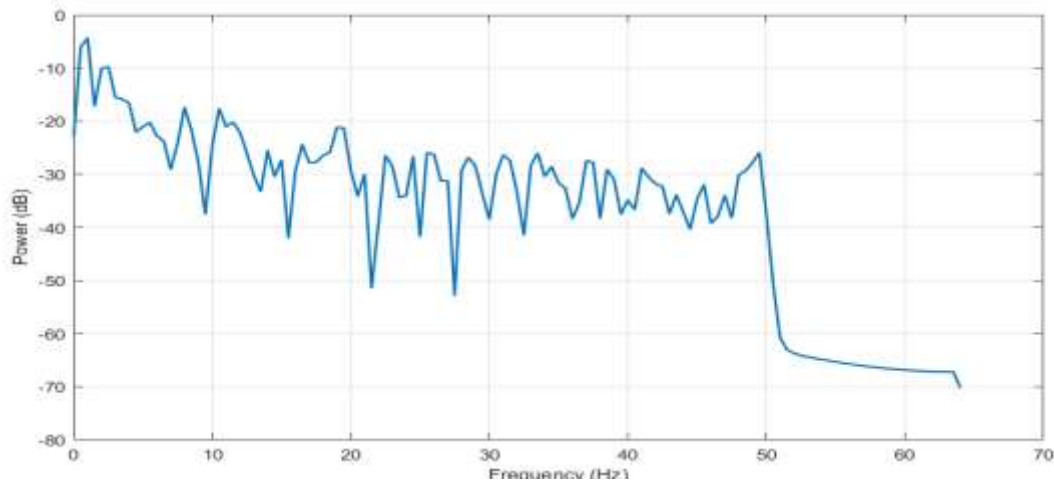


Figure 5: Power Spectral Density (PSD) Showing Frequency-Domain Energy Distribution of the EEG Signal

The EEG signal power spectral density graph shown in Figure 5 demonstrates how signal power distributes itself across different frequency bands. The low-frequency range below 20 Hz shows the highest power levels, which then decrease progressively as frequencies rise. The preprocessing stage produces a sharp power drop in the 50 to 60 Hz range because noise suppression and filtering techniques were applied. PSD analysis helps

detect emotional processing frequency components because different oscillatory bands contain essential neural signals that researchers use to extract features and classify data.

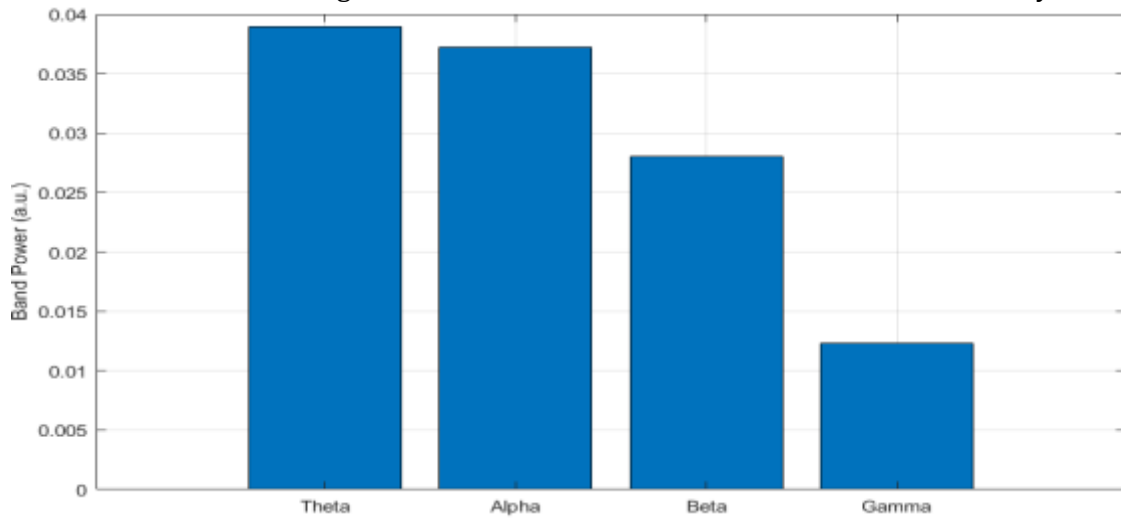


Figure 6: Comparative EEG Band Power Distribution Across Theta, Alpha, Beta, and Gamma Frequency Bands

Figure 6 displays the power distribution of four primary EEG frequency bands, which include Theta and Alpha, and Beta and Gamma. The plot shows that Theta and Alpha bands have the highest power levels, while the Beta band contributes moderate power, and the Gamma band displays lower power. The distribution shows common EEG spectral patterns because lower frequency bands show stronger neural activity during cognitive and emotional processing. Band power analysis provides essential numerical data that helps to extract features while it enables researchers to identify frequency components that are important for emotion classification.



Figure 7 Comparative Classification Accuracy of Deep Learning Models Including BiLSTM, CNN, LSTM, and Hybrid CNN-LSTM Architectures

Figure 7 shows how four deep learning models perform in their ability to classify single-subject EEG emotion recognition data. The hybrid CNN-LSTM model achieved the highest accuracy of 98.32%, which surpassed LSTM (81.51%), BiLSTM (78.99%), and CNN

(73.95%) model performance by a significant margin. The hybrid model shows substantial accuracy improvement because its combination of convolutional spatial feature extraction with temporal sequence modelling provides better discriminative power for personalised EEG-based emotion decoding.

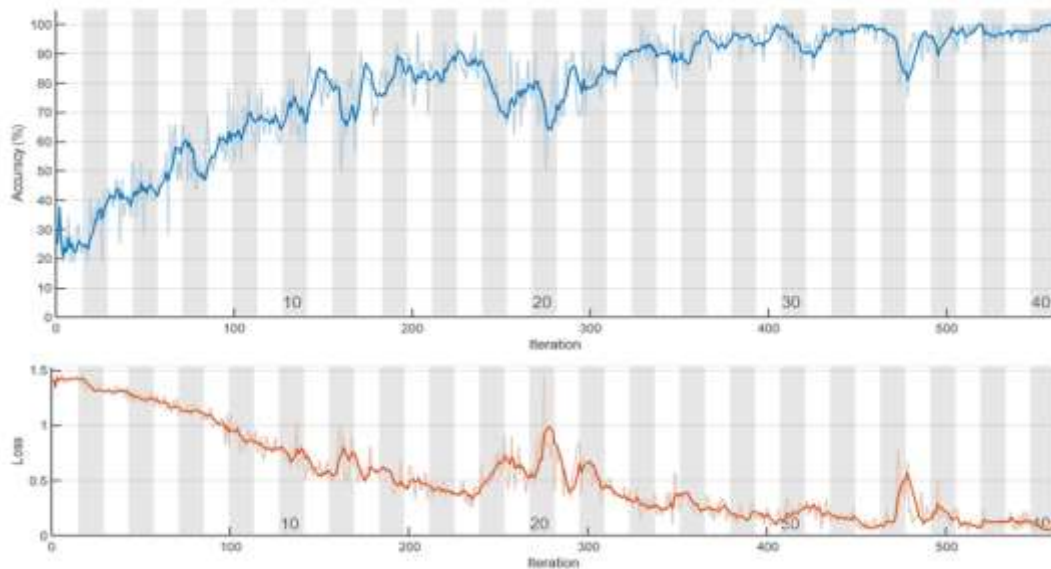


Figure 8: Training Accuracy and Loss Curves for the BiLSTM Model

The BiLSTM model training process shows its accuracy and loss progression through multiple iterations, as Figure 8 displays. The accuracy curve exhibits a steady upward trend, which shows that the model develops its ability to recognise different time-based patterns in EEG data. The loss curve shows a decreasing pattern, which proves that optimisation efforts succeeded and reached a point of stability. The combination of increasing accuracy together with decreasing loss demonstrates that the BiLSTM model training process maintains a stable training pattern.

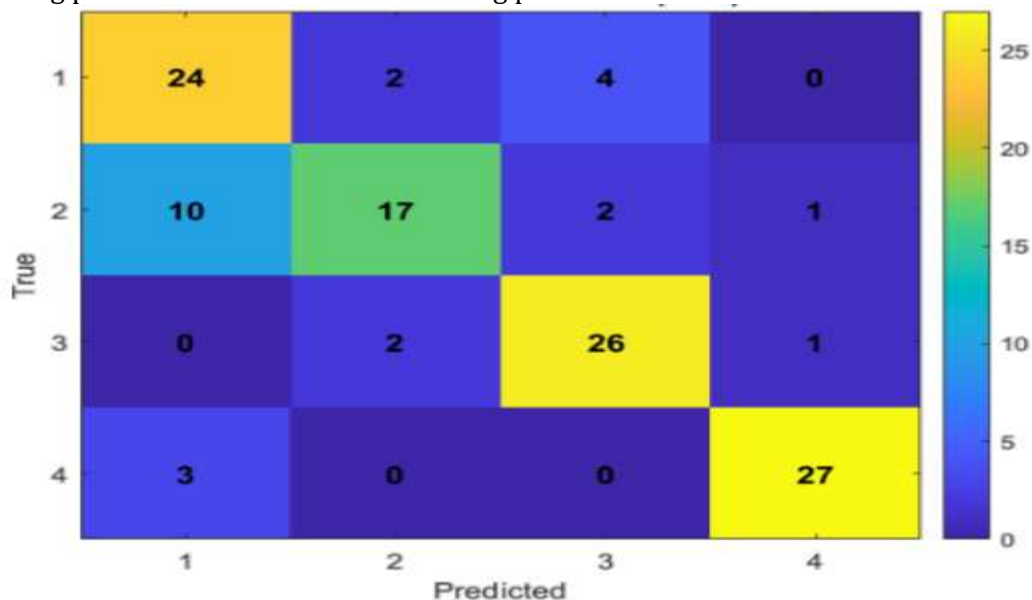


Figure 9: Raw Confusion Matrix for the BiLSTM Model Showing Class-Wise Prediction Counts Across Emotional Categories

The raw confusion matrix for the BiLSTM model is presented in Figure 9, which shows diagonal entries for correct sample classification and off-diagonal entries for misclassification. The model achieves higher correct prediction counts for Class-3 and Class-4 because their temporal EEG patterns exhibit more distinctive characteristics, which make them easier to learn. The misclassification pattern shows that Class-1 and Class-2 display more distributed misclassification to their adjoining classes. The two classes show similar emotional signatures, which lead to their temporal domain patterns becoming less separable. The raw representation enables users to see complete prediction counts, which remain unaltered through normalisation methods.

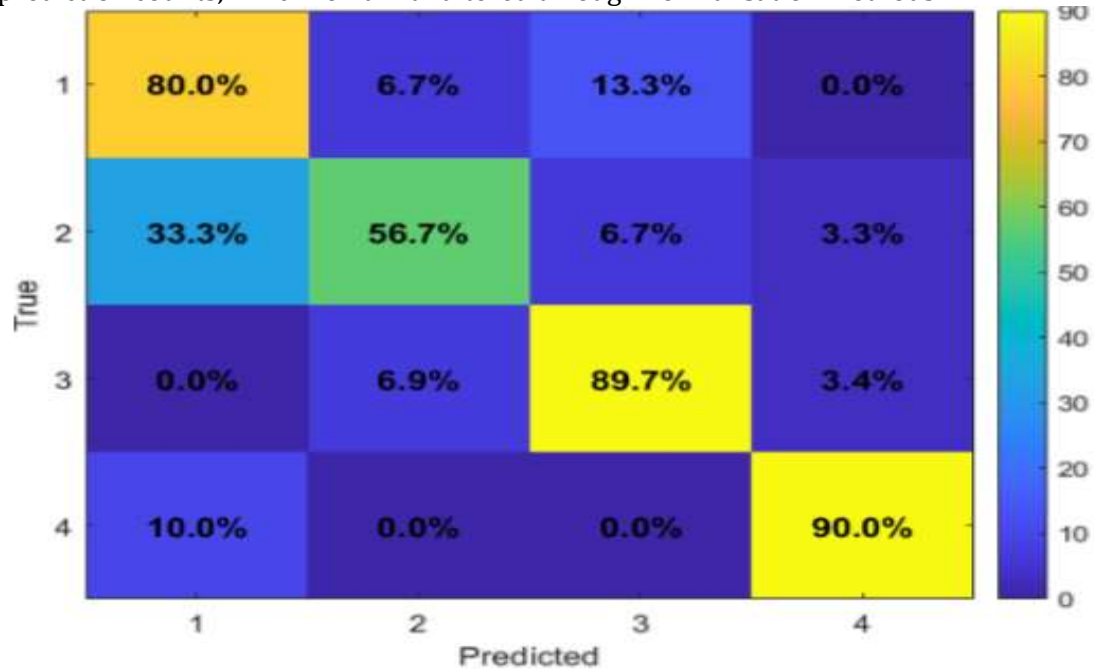


Figure 10. Row-Normalised Confusion Matrix for the BiLSTM Model Illustrating Class-Wise Prediction Percentages Across Emotional Categories

Figure 10 shows the row-normalised confusion matrix, which displays class prediction results in all rows that add up to 100 per cent. The diagonal values show true-positive recognition rates which reach 80.0 per cent for Class-1, 56.7 per cent for Class-2, 89.7 per cent for Class-3 and 90.0 per cent for Class-4. The lower recognition rate for Class-2 shows that the model confuses this class with other classes more often, while high recognition percentages for Class-3 and Class-4 show that the model excels at detecting these emotional states. Normalisation enables better class comparisons, which become especially important when sample sizes between classes vary. The current view demonstrates how different emotional categories show different classification results when assessed through single-subject evaluation methods.

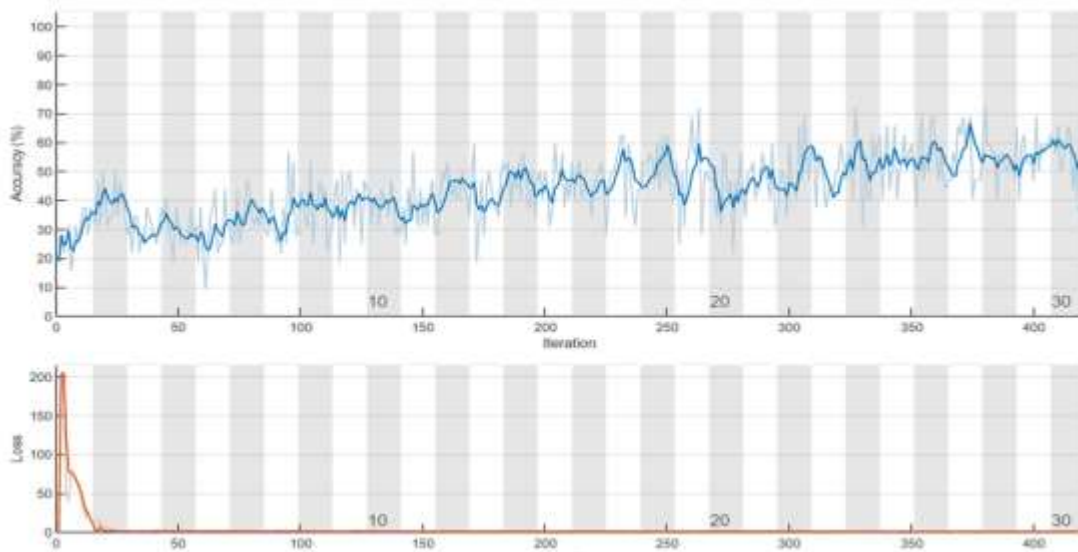


Figure 11. Training Accuracy and Loss Curves for the CNN Model During Learning Iterations

The study displays digraphs that illustrate how the CNN system developed its accuracy and lost performance through multiple learning sessions. The accuracy curve shows continuous upward movement, which demonstrates that the model is becoming better at identifying emotional EEG samples. The loss curve starts with sharp declines during its first iterations before reaching a stable point, which shows that the optimisation process has successfully achieved its goal. The CNN curve shows more variation than recurrent architectures because pure convolutional systems cannot extract temporal features through their design.

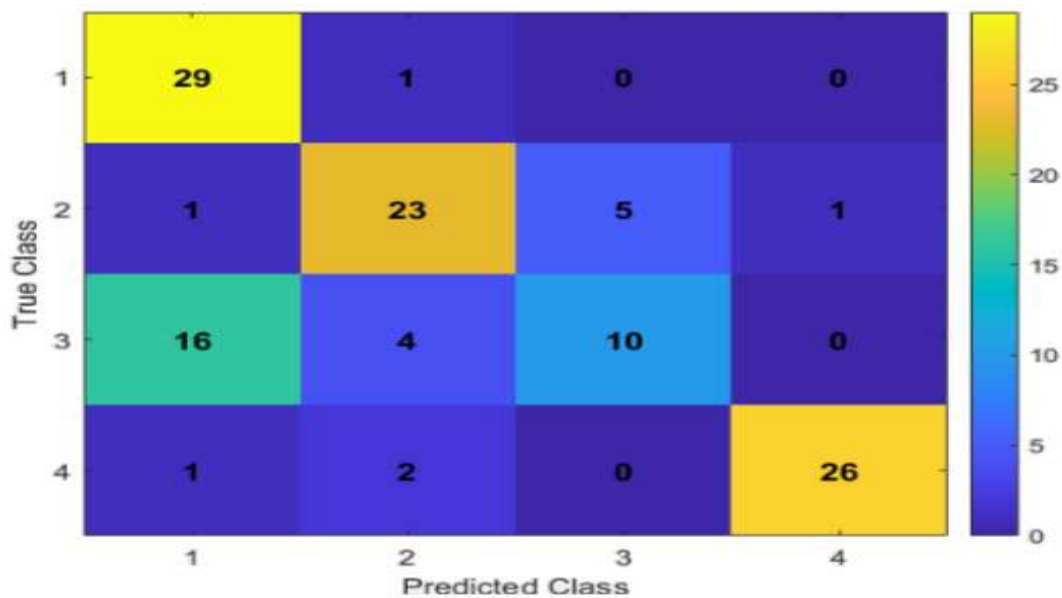


Figure 12. CNN Confusion Matrix Represented Using Raw Classification Counts Across Four Classes

The CNN model shows its prediction results for four emotional classes through its raw confusion matrix, which is displayed in Figure 12. The diagonal entries show the number of correctly identified samples, while Class-1 and Class-4 show higher accuracy results.

The distribution of Class-3 misclassified samples between Class-1 and Class-2 shows that the emotional category has decreased ability to separate different emotions. The raw representation shows how CNN makes predictions through its unscaled prediction distribution.

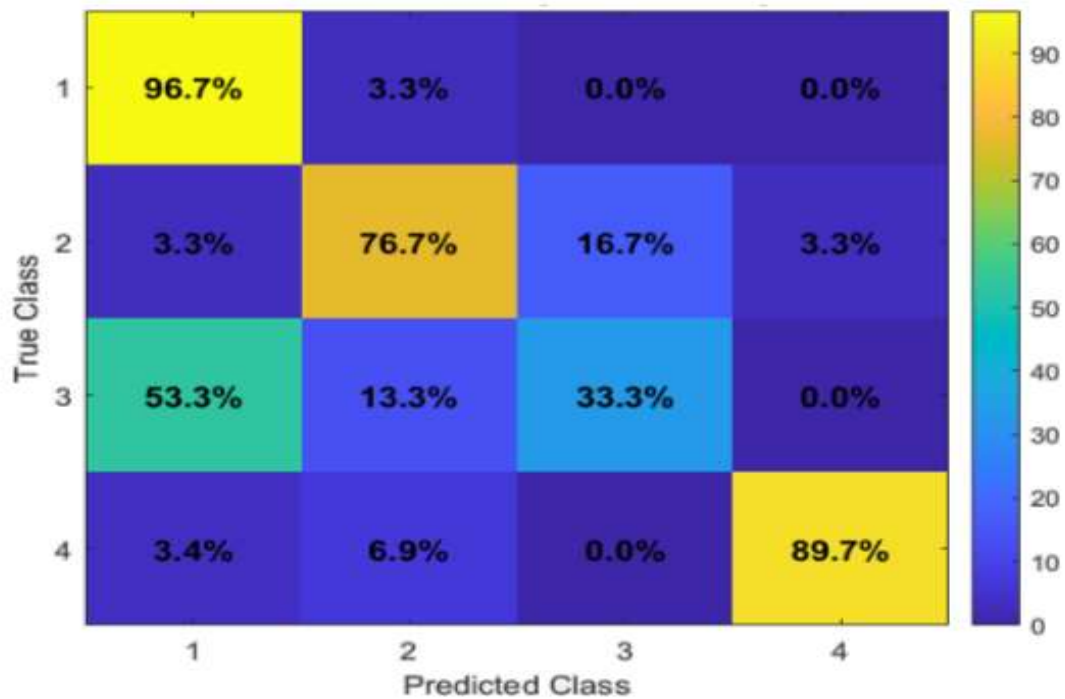


Figure 13. CNN Confusion Matrix Presented Using Row-Normalised Percentages for Classification Performance Comparison

The row-normalised confusion matrix of the CNN model appears in Figure 13 because all row elements add up to 100 percent which demonstrates the system's ability to recognise different classes. The diagonal percentages show that Class-1 recognition accuracy reached 96.7 per cent, while Class-4 recognition accuracy achieved 89.7 per cent, but Class-2 and Class-3 showed reduced recognition performance because their members were incorrectly identified as other classes. The normalised view enables a better understanding of class imbalance, which shows that CNN fails to identify the necessary temporal patterns needed for emotional state differentiation from EEG data.

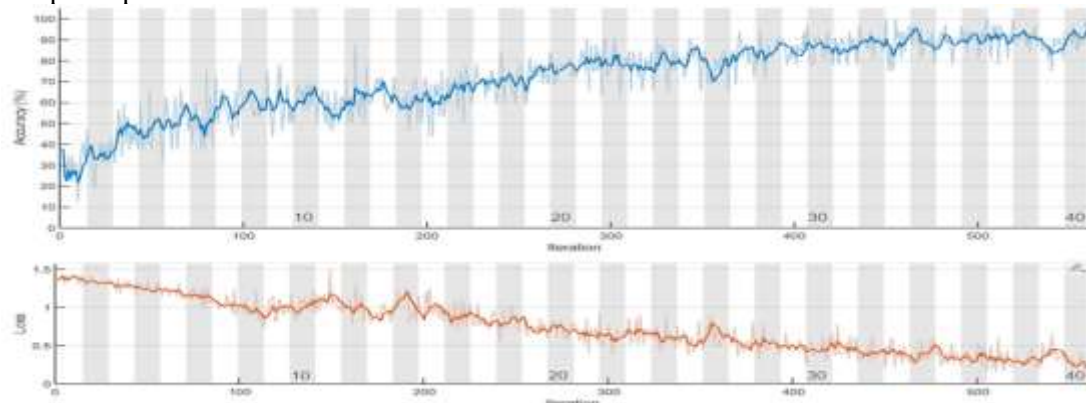


Figure 14: Comprehensive Training Performance of LSTM Model Showing Accuracy Improvement and Loss Convergence Across Iterations

The training process of the LSTM model shows its accuracy and loss values, which change throughout multiple learning iterations. The accuracy curve shows continuous upward movement, which shows the model improves its emotional EEG sample classification skills throughout the testing period. The loss curve shows a fast initial decline during the first training epochs, which progresses to a slower decline as the learning process reaches stabilisation, which demonstrates that the system achieves both optimisation and convergence. The LSTM training curve shows better performance stability than the CNN training curve because it can capture EEG signal temporal dependencies, which serve as essential indicators for detecting emotional states.

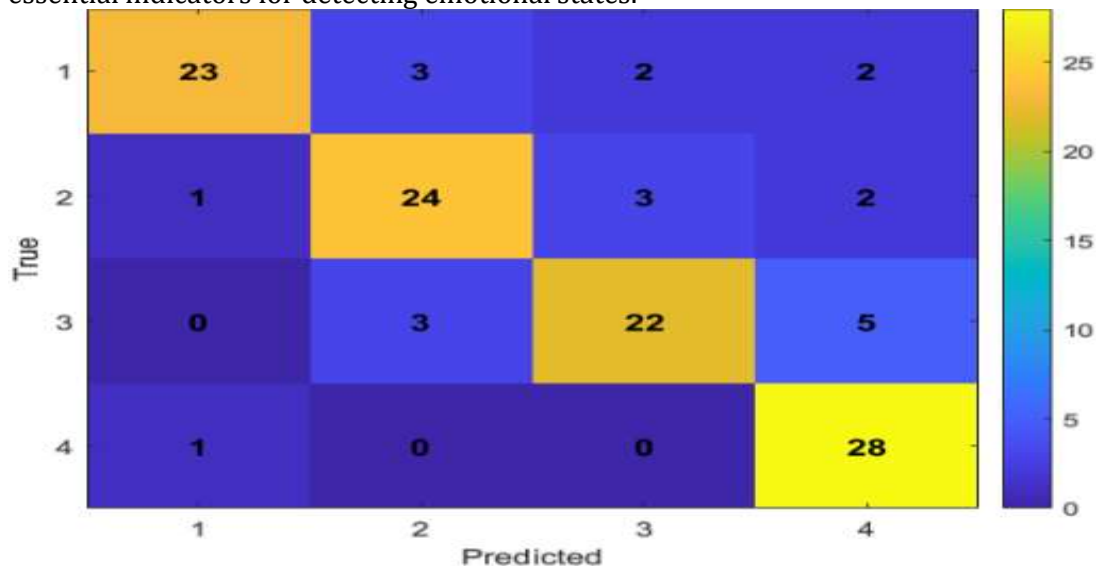


Figure 15: LSTM-Based Emotion Classification Confusion Matrix Depicting Raw Prediction Counts Across Four Emotional Classes

The LSTM classifier shows its unprocessed confusion matrix in Figure 15. The diagonal entries of the matrix show the number of samples that were correctly identified as belonging to each emotional category. The off-diagonal parts of the matrix show the instances of misidentification. The prediction results show that Class-2 and Class-4 have better performance because their correct predictions exceed those of other classes. The pattern of Class-3 confusion shows that it confuses with Class-4 because their EEG time patterning contains some elements in common. The unprocessed matrix allows users to view recognition accuracy for each class without applying any normalising processes.

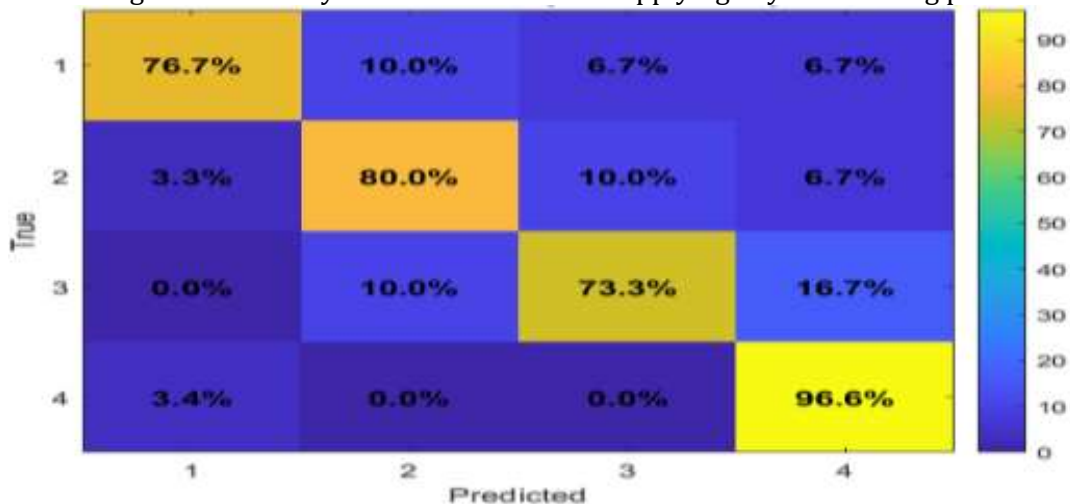


Figure 16: Normalised Confusion Matrix of LSTM Emotion Classifier Showing Percentage-Based Class-wise Prediction Distribution

The LSTM model's confusion matrix display in Figure 16, uses row-based normalisation to create a matrix that shows prediction results for each class. The classifier achieves a high true-positive rate for Class-4 (96.6 per cent) and Class-2 (80.0 per cent), while Class-3 and Class-1 show comparatively lower recognition rates due to distributed misclassification across neighbouring classes. The normalised visualisation shows how different emotional categories express varying levels of recognition difficulty, and it demonstrates how LSTM temporal feature encoding affects the final prediction results in single-subject testing.

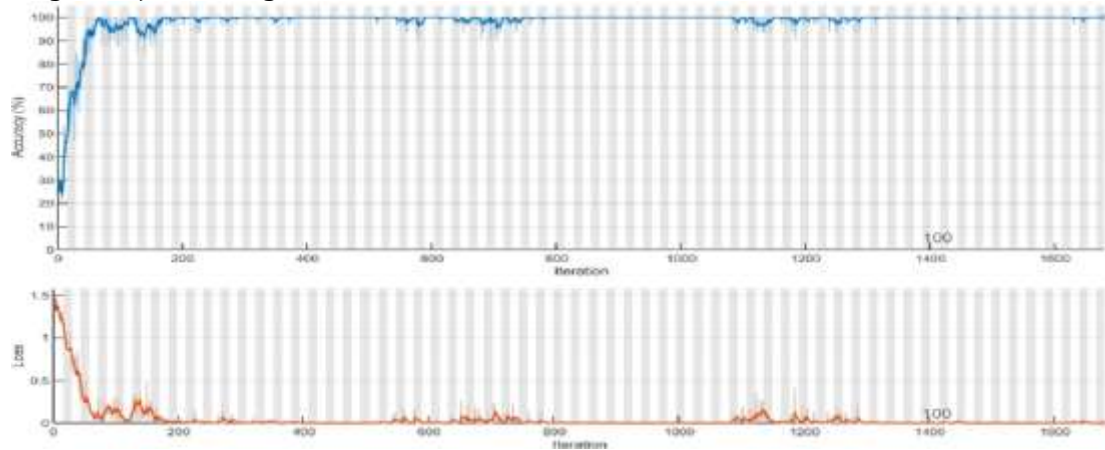


Figure 17: Training Accuracy and Loss Convergence Curve of the Hybrid CNN-LSTM Model

The hybrid CNN-LSTM model training results show accuracy and loss measurements across various training sessions, according to Figure 17. The accuracy curve shows a fast rise during the first training stage, which reaches a stable point above 95 per cent after several training iterations. The loss curve decreases sharply towards near-zero values during the learning process, which demonstrates high convergence rates and successful feature extraction performance. The hybrid system outperforms both CNN and LSTM systems by achieving better and more rapid convergence because it models EEG data through its spatial and temporal characteristics.

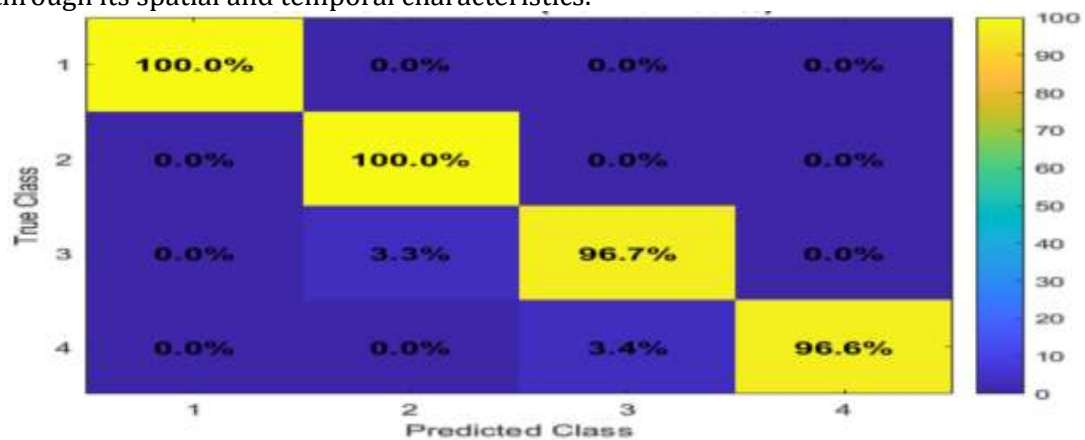


Figure 18: Confusion Matrix of the Hybrid CNN-LSTM Model (Raw Classification Counts)

The hybrid CNN-LSTM classifier confusion matrix raw data appears in Figure 18. The diagonal cells contain the number of correctly classified samples for each emotional class, while off-diagonal cells represent misclassification events. The matrix shows that the model achieves near-perfect classification performance across all four classes, with minimal misclassification observed only in Class-3 and Class-4. The strong diagonal dominance proves that the hybrid system successfully uses both spatial and temporal information, which leads to accurate predictions for individual subject testing.

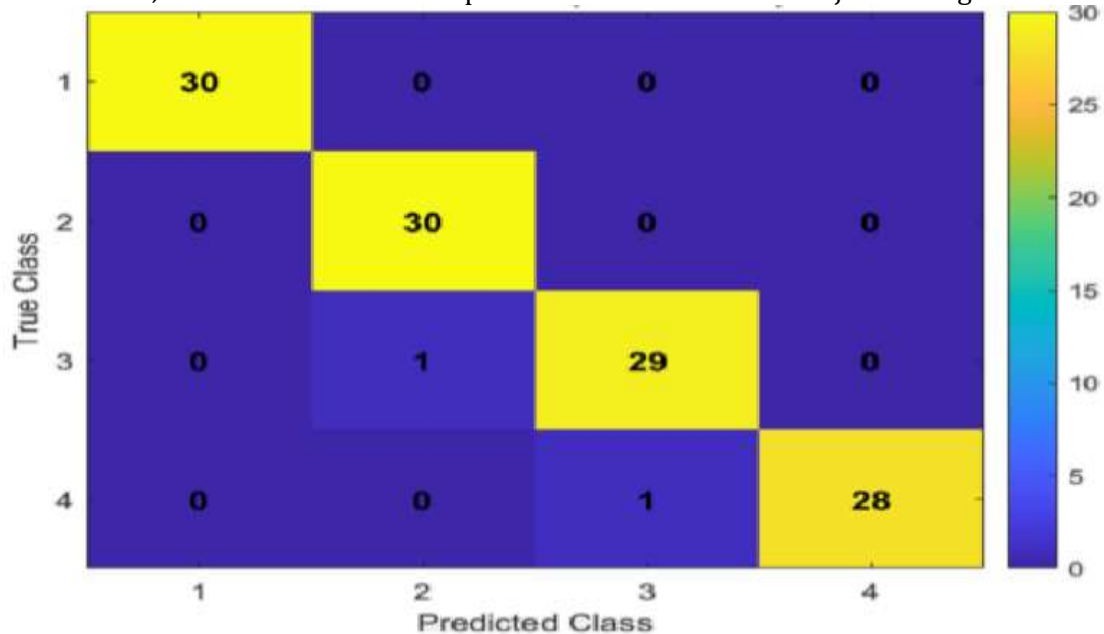


Figure 19: Confusion Matrix of the Hybrid CNN-LSTM Model (Normalised Classification Percentage)

The hybrid CNN-LSTM classifier shows its row-normalised confusion matrix through Figure 19, which displays Class prediction results that add up to 100 per cent across its entire matrix. The diagonal percentages show nearly perfect recognition rates for Class-1 and Class-2 (100.0 per cent each), along with high recognition rates for Class-3 (96.7 per cent) and Class-4 (96.6 per cent). The small off-diagonal percentages in Class-3 and Class-4 demonstrate that there exists very little confusion between these two emotional categories. The normalised view shows how the hybrid model establishes better class generalisation abilities together with its strong performance metrics.

The results demonstrate that the proposed hybrid CNN-LSTM model achieves superior performance for single-subject EEG emotion recognition when compared to standalone CNN, LSTM, and BiLSTM architectures. The model achieves better performance because it combines spatial-spectral feature extraction with temporal sequence modelling, which represents emotional neural response patterns. Previous research has shown that Rivas et al. [18] proved convolutional models deliver superior spatial feature encoding abilities for EEG signals, while Hilali et al. [19] demonstrated that combining convolutional with recurrent models enhances decoding accuracy through joint simulation of spatial and temporal affective patterns. The model achieves high accuracy in single-subject environments because research shows that people process emotions through their unique neural patterns. The study by Shen et al. [20] demonstrated that emotional EEG reactions show different patterns for each individual because their cognitive styles and neural connections differ, which results in learning from actual subjects maintaining their performance standards. The research by Baibulova et al. [21] demonstrated that

representation learning methods enhance emotional state identification by extracting hidden spectral-temporal features, which align with the improved results obtained from the hybrid model employed in this investigation. The study results demonstrate that hybrid representation learning architectures can effectively support personalised EEG-based emotion recognition in user-specific affective computing scenarios when domain uniformity remains intact.

5. Conclusion

This study established an emotion recognition system which utilizes single-subject EEG data and deep learning techniques through hybrid spatial-temporal representation learning. The hybrid CNN-LSTM architecture produced better results than the individual CNN, LSTM, and BiLSTM models according to the experimental tests. The hybrid model achieved 98.32% accuracy and a 0.983 Macro-F1 score, which demonstrated its ability to detect individual brain activity patterns that link to emotional reactions. The model achieved efficient performance because it utilised uniform brain activity patterns from the same subject throughout both training and testing phases without needing any domain adaptation techniques. The proposed method shows practical performance benefits, which make it suitable for use in real-world HCI systems, neurofeedback platforms, adaptive gaming systems and cognitive monitoring systems that require personalised emotional feedback from users. The current study has shown promising results, but there are still areas that need more research. The research team will attempt to expand the framework for testing across multiple subjects who will use domain adaptation methods, transfer learning techniques, and meta-learning methods to handle their unique individual differences. The model needs to extend its capabilities for continuous emotion prediction, which includes arousal and valence dimensions, and needs to integrate multimodal data from peripheral physiological signals and operate in real-time on embedded systems with low latency. The research results establish a base for upcoming studies, which will explore personalised affective computing and EEG-based emotion recognition systems that adjust to user preferences.

REFERENCE

1. Schultz, Tanja und Alexander Maedche. "Biosignals meet adaptive systems." *SN Applied Sciences* 5, no. 9 (2023): 234.
2. Nwagu, Chukwuemeka, Alaa AlSlaity, and Rita Orji. "EEG-based brain-computer interactions in immersive virtual and augmented reality: A systematic review." *Proceedings of the ACM on Human-Computer Interaction* 7, no. EICS (2023): 1-33.
3. Dhothar, Mehakdeep Kaur. "Multimodal Emotion Recognition Using Physiological Signals." PhD diss., Université d'Ottawa/University of Ottawa, 2025.
4. Gkintoni, Evgenia, Anthimos Aroutzidis, Hera Antonopoulou, and Constantinos Halkiopoulos. "From neural networks to emotional networks: A systematic review of EEG-based emotion recognition in cognitive neuroscience and real-world applications." *Brain Sciences* 15, no. 3 (2025): 220.
5. Angulo-Ruiz, Brenda Y., Elena I. Rodríguez-Martínez, Francisco J. Ruiz-Martínez, Ana Gómez-Treviño, Vanesa Muñoz, Sheyla Andalia Crespo y Carlos M. Gómez. "Brain Complexity and Parametrisation of Power Spectral Density in Children with Specific Language Impairment." *Entropy* 27, no. 6 (2025): 572.
6. Ma, Weizhi, Yujia Zheng, Tianhao Li, Zhengping Li, Ying Li, and Lijun Wang. "A comprehensive review of deep learning in EEG-based emotion recognition: classifications, trends, and practical implications." *PeerJ Computer Science* 10 (2024): e2065.

7. Madyatmadja, Evaristus D., Corinthias PM Sianipar, Cristofer Wijaya, and David JM Sembiring. "Classifying crowdsourced citizen complaints through data mining: Accuracy testing of k-nearest neighbours, random forest, support vector machine, and adaboost." In *Informatics*, vol. 10, no. 4, p. 84. MDPI, 2023.
8. Liu, Ziang, Kang Fan, Qin Gu, and Yaduan Ruan. "Channel-Dependent Multilayer EEG Time-Frequency Representations Combined with Transfer Learning-Based Deep CNN Framework for Few-Channel MI EEG Classification." *Bioengineering* 12, no. 6 (2025): 645.
9. Algarni, Mona, Faisal Saeed, Tawfik Al-Hadhrami, Fahad Ghabban, and Mohammed Al-Sarem. "Deep learning-based approach for emotion recognition using electroencephalography (EEG) signals using bi-directional long short-term memory (Bi-LSTM)." *Sensors* 22, no. 8 (2022): 2976.
10. Kannadasan, Kalidasan, Jainendra Shukla, Sridevi Veerasingham, B. Shameedha Begum, and Natarajan Ramasubramanian. "An EEG-based computational model for decoding emotional intelligence, personality, and emotions." *IEEE Transactions on Instrumentation and Measurement* 73 (2023): 1-13.
11. Makhmudov, Fazliddin, Alpamis Kutlimuratov, and Young-Im Cho. "Hybrid LSTM-attention and CNN model for enhanced speech emotion recognition." *Applied Sciences* 14, no. 23 (2024): 11342.
12. Rouast, Philipp V., Marc TP Adam, and Raymond Chiong. "Deep learning for human affect recognition: Insights and new developments." *IEEE Transactions on Affective Computing* 12, no. 2 (2019): 524-543.
13. Houssein, Essam H., Asmaa Hammad, and Abdelmgeid A. Ali. "Human emotion recognition from EEG-based brain-computer interface using machine learning: a comprehensive review." *Neural Computing and Applications* 34, no. 15 (2022): 12527-12557.
14. Wang, Yiming, Bin Zhang, and Lamei Di. "Research progress of EEG-based emotion recognition: a survey." *ACM Computing Surveys* 56, no. 11 (2024): 1-49.
15. Apicella, Andrea, Pasquale Arpaia, Giovanna Mastrati e Nicola Moccaldi. "EEG-based detection of emotional valence towards a reproducible measurement of emotions," *Scientific Reports* 11, no. 1 (2021): 21615.
16. Xu, Peng, Timothy M. Hospedales, Qiyue Yin, Yi-Zhe Song, Tao Xiang, and Liang Wang. "Deep learning for free-hand sketch: A survey." *IEEE transactions on pattern analysis and machine intelligence* 45, no. 1 (2022): 285-312.
17. Rivas, Jesús Joel, Felipe Orihuela-Espina, Luis Enrique Sucar y Nadia Bianchi-Berthouze. "Dealing with a missing sensor in a multilabel and multimodal automatic affective states recognition system." In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 1-8. IEEE, 2021.
18. Farhangmehr, V., Imanian, H., Mohammadian, A., Cobo, J.H., Shirkhani, H. and Payeur, P., 2025. A spatiotemporal CNN-LSTM deep learning model for predicting soil temperature in diverse large-scale regional climates. *Science of the Total Environment*, 968, p.178901.
19. Kingphai, Kunjira, and Yashar Moshfeghi. "Mental workload assessment using deep learning models from eeg signals: A systematic review." *IEEE Transactions on Cognitive and Developmental Systems* (2024).
20. Pillalamarri, Rajasekhar, and Udhayakumar Shanmugam. "A review on EEG-based multimodal learning for emotion recognition." *Artificial Intelligence Review* 58, no. 5 (2025): 131.
21. Mathiyazhagan, Sathish, and MS Geetha Devasena. "Cross-domain Spatio-temporal and Multi-scale Residual Attention Mechanisms for Robust EEG Motor Imagery

- Classification." *International Journal of Computational Intelligence Systems* 18, no. 1 (2025): 285.
22. Huy-Tung, Phuong, Im Eun-Tack, Oh Myeong-Seok, and Gim Gwang-Yong. "EEG-Based Emotion Recognition: A Review and Emerging Paths." *IEEE Access* (2025).