Methodological Perspectives on EEG-Based Emotion and Mental State Recognition Using Signal Processing and Deep Learning

Edgie Bajuyo
Department of Computer Engineering
University of Science and Technology
of Southern Philippines
Cagayan de Oro City, Philippines
bajuyo.edgieroszel21@gmail.com

Adriane Loquinte
Department of Computer Engineering
University of Science and Technology
of Southern Philippines
Cagayan de Oro City, Philippines
adrianeloquinte@gmail.com

Uriel Dionsay
Department of Computer Engineering
University of Science and Technology
of Southern Philippines
Cagayan de Oro City, Philippines
urieldionsay@gmail.com

Abstract—This review examines developments in electroencephalography (EEG)-based emotion and mental state recognition, focusing on the integration of digital signal processing and machine learning methodologies. The reviewed studies identify four dominant methodological patterns: time–frequency image conversion with convolutional neural network (CNN) feature extraction; hybrid CNN–LSTM architectures for joint spatial–temporal modeling; entropy-based and block-structured convolutional networks for continuous state decoding; and neuromorphic systems utilizing spiking neural networks (SNNs) for efficient spike-encoded classification.

Across these approaches, persistent challenges arise, including high inter-subject variability, signal noise and non-stationarity, methodological inconsistencies that hinder comparability, and the computational complexity of deep learning architectures. These limitations affect generalization, reproducibility, and the feasibility of real-time deployment, highlighting the need for standardized preprocessing workflows, robust evaluation protocols, and compact models suitable for embedded or wearable devices.

Solutions identified in the reviewed literature include data augmentation, transfer learning, domain adaptation, and multimodal fusion to enhance generalization and efficiency. Future research should emphasize the integration of these strategies into unified, interpretable pipelines that balance model performance with transparency and computational efficiency, enabling reliable and reproducible EEG-based emotion recognition systems for real-world applications.

Index Terms—Electroencephalogram (EEG), Emotion Recognition, Digital Signal Processing (DSP), Machine Learning, Feature Extraction, Deep Learning, CNN, LSTM.

I. Introduction

The analysis of Electroencephalogram (EEG) signals has emerged as a significant area of research, offering a non-invasive window into the complexities of human brain function. In recent years, emotion and mental state recognition from electroencephalography is an increasingly important direction in human–computer interaction, mental health assessment, and adaptive multimedia systems as it enables machine learning systems to infer internal cognitive–affective dynamics from noninvasive neural signals.

There has been a considerable convergence of Digital Signal Processing (DSP) and Machine Learning (ML) to decode these signals for a variety of applications, Traditional pipelines have relied on band-limited preprocessing and handcrafted features, but their sensitivity to artifacts, subject variability, and non-stationarity often limits generalization across individuals and recording sessions [1], [2]

We are observing a clear trend where DSP-based feature extraction is not a separate preliminary step but is instead intricately woven into the architecture of advanced models. This includes the usage of Wavelet Convolutional Neural Network (WCNN) and Support Vector Machine (SVM) [3] where it basically converts electroencephalography (EEG) windows into continuous wavelet transform scalograms. This approach also uses several references where it also makes use of explicit time–frequency images plus a convolutional neural network feature extractor and finishes with a classical margin-based classifier [1], [2]. Similarly, [4] makes use of EEG Recognition through the usage of hybrid CNN and Long short-term Memory Classification (LSTM) which combines convolutional neural network with a long short-term memory module after standard preprocessing and feature formation, aligning with the hybrid convolutional neural network–recurrent neural network designs highlighted and contextualized in the deep-learning survey [5].

As of recent, there are numerous amounts of features for CNN that can coincide with the recognition of emotion and mental states. Such as that of [6], [7] which focuses on the usage of entropy and other multiple features that can determine and decode a person's mental state. Specifically, the former [6] uses mathematically defined entropy descriptors as the primary digital signal processing representation and evaluates with standard supervised classifiers [2], [8]. Whereas, the latter [7] uses a tailored convolutional neural network directly on electroencephalography (with minimal handcrafted features) and targets continuous, sliding-window inference rather than only trial-wise emotion labels [1]. In a way, [9] is also an approach where it uses spiking neural networks and basically extracts engineered features per window which then converts them into spike trains. In itself, it's a rather distinct approach that keeps a conventional digital signal processing front-end but replaces standard classifiers with a spiking neural network

to exploit spike-based computation [10].

Through these advances, this paper proposes the usage of several processing techniques that work alongside with the training of appropriate models. This review paper aims to introduce several approaches to realizing EGG-Recognition for both emotion and mental spectrum in order to be used alongside machine learning or basically the training of appropriate models.

II. METHODOLOGY

The intellectual challenge of decoding emotions from EEG signals lies not in a single algorithm, but in the thoughtful construction of a multi-stage processing pipeline. This section dissects this symbiotic relationship.

A. Wavelet Convolutional Neural Networks and Support Vector Machine

A pivotal insight in recent research is that the representation of data is as crucial as the classification model itself. One dominant trend is the transmutation of 1D EEG time-series data into 2D time-frequency representations, effectively turning a signal processing problem into a computer vision one. Time-frequency (TF) methods [11], such as the Continuous Wavelet Transform (CWT), can convert a 1D EEG signal into a 2D representation that captures the spectral variation of the signal over time. One dimension corresponds to time while the other represents frequency, forming an image that reflects EEG power fluctuations across both domains. This process models the signal as a linear combination of basic functions known as wavelets, mathematically expressed as:

$$W_{(a,b)}[x(t)] = \frac{1}{|a|^{1/2}} \int_{-\infty}^{+\infty} x(t)\, \varphi^* \left(\frac{t-b}{a} \right) dt \quad (1)$$

where a is the scale (a real and positive value), b is the translational parameter (a real number), and φ denotes the mother wavelet, which localizes the signal in both time and frequency domains.

This image representation of EEG power changes in time and frequency effectively captures the transient, non-stationary nature of neural oscillations. Similarly, [3] provides a compelling demonstration of this, employing the CWT to generate scalograms. These 2D images allow researchers to leverage the immense power of pre-trained Convolutional Neural Networks (CNNs) [3]. In the study, the deep features learned by a pre-trained ResNet-18 from the CWT scalograms were subsequently fed into a Multiclass Support Vector Machine (MSVM) [3]. This hybrid strategy, which separates the task of feature representation from classification, significantly boosted performance on the MAHNOB-HCI dataset.

Four pre-trained CNNs were evaluated using three performance measures—accuracy, precision, and recall—through the leave-one-subject-out cross-validation (LOSO-CV) [12] approach. In this approach, data from 31 subjects were used to fine-tune existing CNNs, while data from another subject were used as the test set. This ensures subject independence, as no

samples from a single subject appear in both the training and testing sets. The process repeats until all subjects have been tested once, and the final reported results are the mean and standard deviation across all iterations.

Accuracy, precision, and recall [13] are calculated as follows:

$$Accuracy = \frac{1}{l} \sum_{i=1}^{l} \frac{tp_i + tn_i}{tp_i + tn_i + fp_i + fn_i} \quad (2)$$

$$Precision = \frac{1}{l} \sum_{i=1}^{l} \frac{tp_i}{tp_i + fp_i} \quad (3)$$

$$Recall = \frac{1}{l} \sum_{i=1}^{l} \frac{tp_i}{tp_i + fn_i} \quad (4)$$

where tp_i, tn_i, fp_i, and fn_i denote the true positive, true negative, false positive, and false negative counts, respectively, for each emotional class derived from the confusion matrix of the classifier.

B. Hybrid CNN and LSTM Classification

Another powerful approach involves creating hybrid architectures that combine different neural network types to process EEG data. The CNN-LSTM architecture is a prime example, leveraging the spatial feature learning of CNNs with the temporal sequence modeling of LSTMs. [4] exemplifies this with a model that integrates a CNN, LSTM, and a ResNet-152 backbone [14].
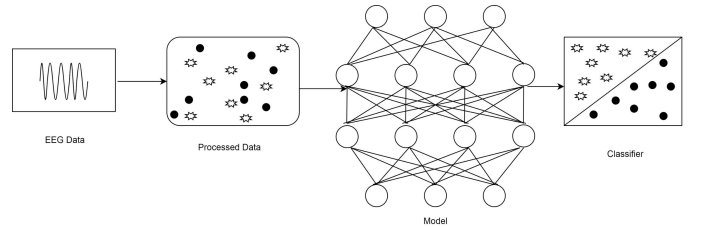

Fig. 1. The flow of electroencephalography (EEG) data processing, showing sequential stages from raw EEG input, preprocessing, feature modeling, and emotion classification.

The general EEG-based emotion identification process begins with data acquisition and preprocessing, as shown in Figure 1. EEG data are collected and normalized to ensure consistency across sessions and subjects. Noise and artifacts are then mitigated using band-pass filters, typically in the range of 1–75 Hz, to preserve relevant frequency components. After preprocessing, the signals are transformed into structured representations such as Mel Frequency Cepstral Coefficients (MFCCs) [4], entropy parameters, or topographic feature maps. These representations serve as the input to the hybrid CNN-LSTM network, which learns spatial features through convolutional layers and temporal dependencies through recurrent layers. This combined approach enhances the model's ability to recognize emotion-related EEG patterns and classify them accurately.
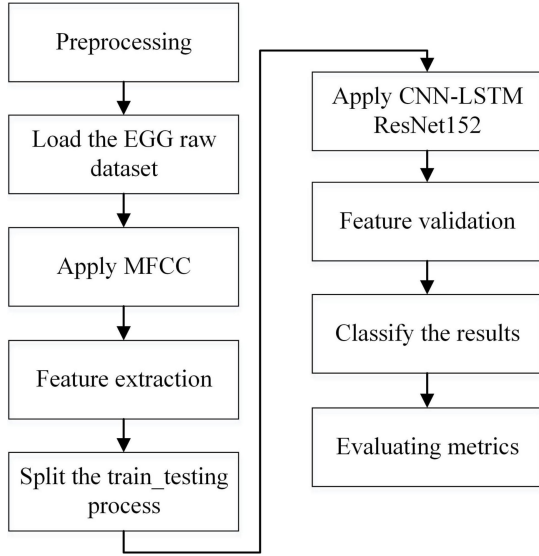
Fig. 2. Overview of the CNN-LSTM methodology, showing stages from pre-processing, feature extraction, model application, and performance evaluation.

Figure 2 presents the complete methodological framework of the CNN-LSTM model used for EEG-based emotion classification. The model merges a homogeneous CNN and LSTM classifier with the ResNet-152 backbone to maximize both spatial and temporal feature representation. The SEED-V EEG dataset—comprising emotional states such as happiness, disgust, fear, neutral, and sadness—was used to evaluate this model. Preprocessing was conducted to prepare EEG data channels (FP1, FP2, FC6, and F3), followed by feature extraction through MFCC [4] computation and entropy analysis. The entropy was calculated using the sample entropy method, while the Hurst exponent was determined via R/S analysis to capture signal complexity and variability.

The extracted features were converted into topographic maps that served as CNN inputs, while the LSTM layers modeled the temporal dynamics of these representations. Both CNN and CNN-LSTM configurations employing ResNet-152 were trained and compared. The models classified emotional states and conditions such as PTSD and non-PTSD, with performance evaluated using standard metrics including accuracy, precision, recall, and F1-score. This end-to-end framework demonstrates the hybrid model's capacity to fuse spatial-spectral and temporal information for robust EEG-based emotion recognition.

C. Multiple Feature Block-Based Extraction Measure

In EEG-based emotion and mental state recognition, one emerging approach focuses on enhancing feature representation by integrating domain knowledge directly into the network design. Rather than relying purely on end-to-end deep learning, these methods combine handcrafted feature engineering with specialized convolutional architectures. This strategy bridges the interpretability of traditional DSP-based features with the adaptability of deep neural models. The goal is to capture complex temporal and spatial patterns from EEG

signals while preserving meaningful relationships within the data, leading to more reliable cognitive state recognition.
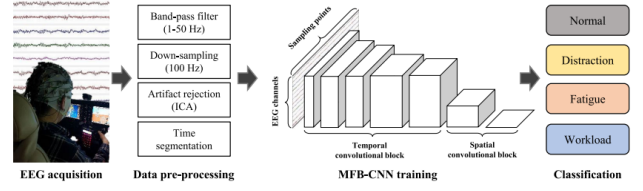

Fig. 3. Overview of the MFB-CNN workflow for EEG-based mental state classification, illustrating stages from EEG acquisition and preprocessing to temporal–spatial feature extraction and final classification into cognitive states.

A particularly salient feature in this domain is entropy, which quantifies the signal's complexity. [6] explored how various entropy-based measures—including Differential Entropy (DE), Sample Entropy (SampEn), and Approximate Entropy (ApEn) [6]—can serve as potent indicators of emotional states. These handcrafted features are then classified using traditional machine learning models such as SVM, k-NN, and Decision Trees.

Building upon this foundation, [7] proposed the Multiple Feature Block-based Convolutional Neural Network (MFB-CNN), a hybrid model designed to decode mental states in high-interference environments such as simulated flight. Because EEG data in such settings are susceptible to motion artifacts, preprocessing plays a vital role. Artifacts were removed using Independent Component Analysis (ICA) [7], where four EOG channels were designated as contamination references. Signal processing and experimental control were implemented using MATLAB 2019a with the BBCI toolbox.

After preprocessing, the EEG dataset consisted of all recorded channels and sampling points (300×300) as input matrices. The MFB-CNN was structured into seven convolutional blocks organized hierarchically to enable progressive feature abstraction. Each convolutional block contained four convolutional layers and one batch-normalization layer with a filter size of 32. Temporal feature extraction used filter sizes of 1×5, while spatial feature extraction in later blocks used 3×1 filters to capture inter-channel dependencies. Dropout layers with a rate of 0.5 were applied to prevent overfitting, and max- and average-pooling layers reduced redundant information while retaining essential activation patterns.

An exponential linear unit (ELU) activation function was applied in the seventh convolutional block to improve nonlinear learning efficiency, defined as:

$$ELU(x) = \begin{cases} x, & \text{if } x > 0 \\ \alpha(e^x - 1), & \text{if } x \leq 0 \end{cases} \quad (5)$$

For model training, the network was iteratively trained over 30 epochs, and classification accuracies were evaluated on the test data. The MFB-CNN successfully classified EEG patterns into cognitive states such as normal, distraction, fatigue, and workload, achieving an offline accuracy of 0.75. This result

underscores the potential of the MFB-CNN framework to capture both temporal and spatial dependencies for real-time mental state monitoring.

D. Spiking Neural Networks

Venturing beyond conventional deep learning, some researchers are exploring neuromorphic approaches that emulate biological neural behavior more closely. Spiking Neural Networks (SNNs) [9] operate on event-driven computations, processing information as discrete spikes rather than continuous activations. This time-based mechanism enables SNNs to capture temporal dependencies within EEG signals efficiently while consuming significantly less power—making them ideal for real-time emotion recognition and wearable applications.
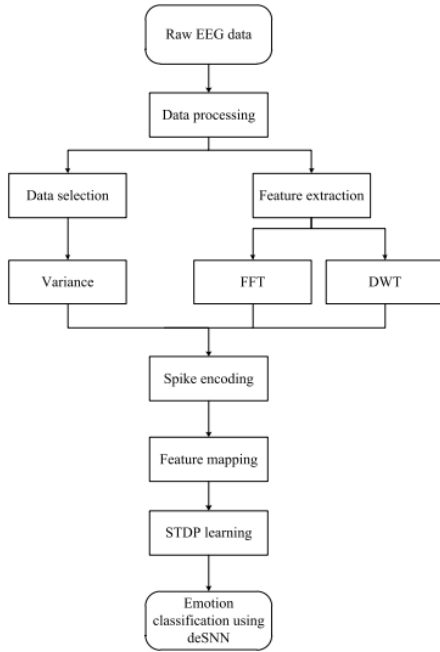

Fig. 4. Workflow of EEG signal processing and emotion classification using deSNN, showing the stages from raw EEG input through feature extraction, spike encoding, and STDP-based classification.

As shown in Figure 4, EEG-based emotion recognition using SNNs follows a structured pipeline: data processing, feature extraction, and classification. The raw EEG data undergo preprocessing to remove noise and artifacts. Datasets such as DEAP or SEED [15] are typically down-sampled to 128–200 Hz and filtered using a band-pass filter (0.75–45 Hz) via EEGLAB or similar toolboxes. Signals are separated into frequency bands (alpha, beta, gamma, and theta), corresponding to specific physiological and cognitive states.

Feature extraction then integrates both time-domain and frequency-domain analyses to generate spike-compatible features. In the time domain, the variance of EEG signals helps identify steady emotional states, computed as:

$$\mu_\xi = \frac{\xi(t_1) + \xi(t_2) + \cdots + \xi(t_T)}{T}, \quad v_\xi = \frac{1}{T}\sum_{t=1}^{T}[\xi(t) - \mu_\xi]^2 \tag{6}$$

In the frequency domain, the Fast Fourier Transform (FFT) and Discrete Wavelet Transform (DWT) [9] are applied to extract oscillatory and transient features. The DFT of the EEG signal $\xi(t)$ is expressed as:

$$\xi(k) = DFT[\xi(t)] = \sum_{t=0}^{N-1} \xi(t)W_N^{tk}, \quad W_N^k = e^{-j\frac{2\pi k}{N}} \tag{7}$$

and the maximum frequency measurable by the Nyquist theorem is $f_{max} = \frac{f_s}{2}$, where f_s is the sampling frequency. For the DWT, non-stationary EEG signals are decomposed into multiple scales, providing localized frequency information. It is given by:

$$W(a,b) = \frac{1}{\sqrt{|a|}} \int s(t)\psi^*\left(\frac{t-b}{a}\right) dt \tag{8}$$

where a and b represent the scaling and translation parameters, and $\psi(t)$ is the mother wavelet, often the "db4" wavelet for EEG applications.

After these transformations, features are spike-encoded, converting continuous amplitude information into discrete spike trains. These spike patterns are then used as input to a spiking network, where temporal learning mechanisms such as Spike-Timing-Dependent Plasticity (STDP) [9] dynamically adjust synaptic weights according to spike timing differences.
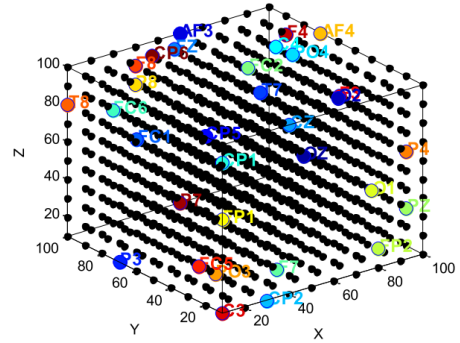

Fig. 5. Architecture of the NeuCube SNN framework showing the 3D spatio-temporal neuron cube and EEG electrode mappings used for encoding and classification.

A well-known implementation of SNN for emotion classification is the NeuCube model, which employs a 3D spatio-temporal neural reservoir. As illustrated in Figure 5, the NeuCube architecture is composed of three principal components: (1) an input module for temporal encoding, (2) the 3D SNNcube for dynamic mapping of EEG signals, and (3) an evolving SNN classifier. Temporal coding is employed to represent EEG data where the timing of spikes carries information about the signal amplitude and frequency characteristics.

The NeuCube typically consists of 1000 neurons arranged in a 10×10×10 lattice structure based on the small-world connectivity principle, which ensures efficient information transfer among neurons. The neurons are interconnected through proximity-based weights, forming both short-range and long-range connections that simulate biological cortical networks. During initialization, input neurons corresponding to EEG electrode sites (e.g., FP1, FP2, FC2, and O1) are mapped spatially to specific nodes in the SNNcube.

The learning process in NeuCube utilizes a variant of the STDP rule to modify synaptic weights between presynaptic and postsynaptic neurons depending on their firing intervals. The STDP weight update function is defined as:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij},$$
$$\Delta w_{ij} = \begin{cases} A^+ e^{-\Delta t/\tau^+}, & \text{if } \Delta t > 0, \\ -A^- e^{\Delta t/\tau^-}, & \text{if } \Delta t \leq 0. \end{cases} \quad (9)$$

where Δt is the time difference between presynaptic and postsynaptic spikes, A^+ and A^- are learning rate constants, and τ^+ and τ^- are time constants for potentiation and depression, respectively. Positive Δt values increase synaptic strength (potentiation), while negative Δt values lead to depression.

Once trained, the NeuCube model classifies emotions by recognizing characteristic spatio-temporal spike patterns corresponding to different emotional states. This approach demonstrates how biologically inspired SNNs can bridge the gap between neuroscience and machine learning—achieving efficient, interpretable, and temporally precise emotion decoding from EEG signals.

E. Comparative Analysis of Reviewed Studies

To distill the diverse strategies into a coherent overview, this section presents a comparative analysis of the primary methodologies reviewed. Table I highlights the common threads and shared principles among the studies, while Table II provides a detailed summary of each paper.

TABLE I
COMPARATIVE ANALYSIS OF METHODOLOGICAL SIMILARITIES

Ref.	Approach	Similarities
[3]	Wavelet Convolutional Neural Networks and Support Vector Machine	The usage of wavelet convolutional neural networks along with support vector machines plays a fairly crucial role in this study as it provides a way to convert electroencephalography windows into continuous wavelet transform scalograms, which is proven necessary for further research due to the usage of convolutional neural networks
[4]	Hybrid Convolutional Neural Network and Long Short-term Memory Classification	Hybrid convolutional neural network and long short-term memory classification basically implements hybrid models (CNN+LSTM) to capitalize on the strengths of two-stage deep architecture that separates per-window representation from across-window dynamics.
[6]	Entropy as a Feature Extraction Measure	Entropy as a feature extraction measure uses mathematically defined entropy descriptors as the primary digital signal processing representation and evaluates with standard supervised classifiers. This works similarly with other features for convolutional neural network
[7]	Multiple Feature Block-Based Convolutional Neural Network	Block-based convolutional neural network features multiple Uses that are tailored to convolutional neural network which operates directly on electroencephalography and targets continuous, sliding-window inference rather than only trial-wise emotion labels.
[9]	Usage of Spiking Neural Networks	Spiking neural networks uses a distinct approach that keeps a conventional digital signal processing front-end but replaces standard classifiers with a spiking neural network to exploit spike-based computation.

The analysis across these studies reveals several key trends. As highlighted in Table I, The goal is to show, at a glance, which methods belong to the same family so that later comparisons are fair. This table also basically finds similarities between these main references in order to state the common idea these pipelines share (for example, time–frequency image formation, sequence modeling across windows, use of mathematically defined features, sliding-window inference, or engineered features with a nonstandard classifier). This table is present to justify why the paper groups methods the way it does, and to make the evaluation choices (preprocessing, windowing, feature formation, classifier) transparent and comparable across studies.

TABLE II
SUMMARY OF REVIEWED JOURNAL PAPERS

Paper Title	Focus	Mathematical Tools Used	Key Findings
A Hybrid EEG-Based Emotion Recognition Approach Using Wavelet Convolutional Neural Networks and Support Vector Machine [3]	Conversion of electroencephalography windows to time–frequency images and classify emotions with deep features plus a classical classifier.	Continuous Wavelet Transform to create scalograms; convolutional neural network intermediate feature maps; multiclass Support Vector Machine; leave-one-subject-out cross-validation.	Early-layer convolutional features combined with a multiclass Support Vector Machine improved accuracy versus end-to-end baselines; best results reported around 87.45% (DEAP) on the Database for Emotion Analysis using Physiological Signals.
EEG-based emotion recognition using hybrid CNN and LSTM networks [4]	Hybrid deep model that encodes per-window patterns and then models window-to-window evolution for emotion recognition.	Band-pass filtering and normalization; entropy measures such as sample entropy; band-power features in canonical frequency bands; convolutional neural network; long short-term memory recurrent module; softmax classification.	Reported accuracy of approximately 98% on the Shanghai Jiao Tong University five-class dataset, outperforming support vector machine and artificial neural network baselines provided by the authors.
EEG-based human emotion recognition using entropy as a feature extraction measure [6]	Survey and analysis of entropy-family features for emotion recognition from electroencephalography.	Approximate entropy; sample entropy; permutation entropy; differential entropy; multiscale fuzzy entropy; wavelet entropy; empirical mode decomposition–based entropies; classifiers such as support vector machines and k-nearest neighbors.	Shows that entropy features are effective descriptors of nonlinear dynamics; summarizes multiple studies with competitive results on common datasets, and positions entropy features as strong baselines and as inputs to modern models.
Continuous EEG Decoding of Pilots' Mental States Using Multiple Feature Block-Based Convolutional Neural Network [7]	Continuous decoding of mental states such as fatigue, workload, and distraction using a tailored convolutional architecture.	Multiple feature block–based convolutional neural network with temporal and spatial filtering; sliding-window inference; cross-subject validation protocols.	Offline accuracy reported around 0.75% with standard deviation near 0.04%; pseudo-online detection around 0.72% for fatigue and workload and around 0.61% for distraction, demonstrating feasibility for continuous monitoring.
EEG-Based Emotion Classification Using Spiking Neural Networks [9]	Neuromorphic classification of emotions using spike-encoded engineered features.	Discrete Wavelet Transform; Fast Fourier Transform; time-domain variance; spike encoding; spiking neural network classifier.	On the Database for Emotion Analysis using Physiological Signals, accuracies per affective dimension reported around 74% to 86%.

Table II presents a comparative summary of five journal papers, each addressing key challenges in electroencephalography-based emotion and mental-state recognition. Although the problem settings differ (trial-wise emotion classification versus continuous state decoding), the papers collectively advance practical digital signal processing pipelines for analyzing electroencephalography data.

Despite using different mathematical tools—ranging from time–frequency analysis with continuous wavelet transform and convolutional feature extraction, to hybrid convolutional neural network–long short-term memory models, entropy-based feature sets with classical classifiers, multiple-feature-block convolutional networks for continuous decoding, and spiking neural networks with engineered features—the studies share a common objective: improving accuracy and robustness under subject-wise evaluation while keeping computation feasible for deployment.

These shared methodologies underline the broader significance of structured preprocessing, well-defined feature construction, and clearly stated evaluation protocols in building reliable, reproducible electroencephalography emotion recognition systems.

III. CHALLENGES AND ISSUES

Challenges arise at multiple stages in developing and implementing EEG-based emotion recognition systems. One major challenge lies in managing the high inter-subject variability of EEG signals. Neural patterns associated with identical emotional states differ significantly across individuals, leading to inconsistencies in model generalization. To address this, studies such as the CWT-CNN approach [3] employ Leave-One-Subject-Out Cross-Validation (LOSO-CV) to evaluate inter-person robustness. Achieving consistent accuracy across subjects remains difficult due to variations in electrode placement, signal amplitude, and noise characteristics inherent to EEG recordings.

The noise-prone and non-stationary nature of EEG data also presents a considerable obstacle. Motion artifacts, muscle activity, and environmental interference often distort the signal, reducing its discriminative power. The MFB-CNN model [7], implemented in a simulated flight environment, required extensive preprocessing using Independent Component Analysis (ICA) to remove motion artifacts and eye-blink interference. Although such preprocessing improves signal quality, it significantly increases computational cost and limits real-time applicability.

Another persistent issue concerns the lack of methodological standardization across studies. Different research groups employ heterogeneous pipelines—ranging from CWT-CNN-SVM hybrids [3] and CNN-LSTM frameworks [4] to MFB-CNN architectures [7] and neuromorphic SNN models [9]. Each approach uses distinct preprocessing protocols, feature extraction strategies, and evaluation metrics, complicating the comparison of results across studies. This fragmentation hinders the establishment of consistent benchmarks and slows the development of reproducible EEG-based emotion recognition systems.

The computational complexity of deep learning models presents another difficulty. Architectures employing large backbones such as ResNet-152 [4] provide high accuracy but demand substantial computational resources for both training and inference. This constraint poses limitations for deployment in portable or embedded devices intended for continuous emotion monitoring. Spiking Neural Networks (SNNs) [9] have emerged as an energy-efficient alternative, yet they currently involve trade-offs between efficiency and precision, creating challenges in maintaining accuracy under low-power operation.

The scarcity and inconsistency of EEG datasets further constrain progress. Existing datasets differ in sampling rates, channel configurations, and labeling conventions, making cross-dataset evaluation and model transferability difficult. Data collection is resource-intensive and sensitive to subject-specific conditions, which restricts the availability of large, standardized datasets for robust training. Addressing these issues requires coordinated standardization of acquisition protocols, optimized preprocessing pipelines, and the design of computationally efficient models capable of operating under real-world variability.

IV. SOLUTION

To address high inter-subject variability and improve generalization across individuals, use subject-independent evaluation by default and report Leave-One-Subject-Out Cross-Validation, as done in the continuous wavelet transform plus convolutional neural network plus support vector machine study [3]. Stabilize inputs with per-recording normalization and fixed, overlapping windowing before feature construction. When a simple baseline is needed, include entropy-based features that depend on distributional properties rather than absolute amplitudes to reduce sensitivity to subject scale differences [6]. For time–frequency image pipelines, aggregate channels into consistent regional groupings and use early convolutional features with a lightweight classifier, which showed better cross-subject behavior than deeper end-to-end settings in [3].

To reduce the impact of noise and non-stationarity, keep a short, repeatable preprocessing chain: task-appropriate bandpass, mains notch, fixed segmentation, and artifact removal when present. Independent Component Analysis should be applied when ocular or motion artifacts are visible; it was necessary for reliable operation in the multiple feature block–based convolutional neural network pilot study [7]. For feature formation, use continuous wavelet transform scalograms to capture transient content for image-based models [3] and entropy measures such as sample entropy, permutation entropy, and differential entropy to summarize irregularity in engineered-feature pipelines [6].

To limit methodological fragmentation and make results comparable, lock a single protocol and report it completely: channel montage and reference, sampling rate, exact filter bands and notch frequency, window length and hop, whether Independent Component Analysis was used, the precise wavelet family and scales for continuous wavelet transform, and the exact definitions and parameters for each entropy measure. State clearly whether results are within-subject or cross-subject, and match the evaluation style of the main pipelines you are citing (for example, LOSO for [3] and [7], sequence evaluation for [4]).

To manage computational complexity, replace very large backbones with compact alternatives or tap early convolutional layers and classify with a lightweight model. The wavelet plus convolutional neural network plus support vector machine pipeline in [3] showed that early-layer features with a multiclass support vector machine can outperform deeper end-to-end models while reducing training and inference cost. In the convolutional neural network plus long short-term memory setting [4], keep the convolutional front-end and the recurrent depth minimal and limit input resolution and sequence length to what is needed for accuracy. For low-power targets, use engineered features such as discrete wavelet transform, fast Fourier transform, and time-domain variance with a spiking neural network; the neuromorphic approach in [9] reached competitive accuracy with a lower compute budget than conventional deep models.

To work around dataset scarcity and inconsistency, choose one primary dataset for all headline comparisons and resample any secondary datasets to a common sampling rate with a consistent reference montage before training. Map labels to a common scheme when definitions differ. Increase effective data volume using overlapping windows and balanced sampling across classes; these practices are standard in the hybrid pipelines of [3], [4]. For continuous monitoring scenarios, keep the same sliding-window and pseudo-online evaluation used during development so that deployment conditions match validation, as in [7].

V. CONCLUSION

This review analyzed EEG-based emotion and mental state recognition studies, identifying four distinct yet related methodological patterns: (1) time–frequency image conversion with convolutional neural network feature extraction combined with classical classifiers; (2) hybrid convolutional and recurrent architectures for joint spatial and temporal modeling; (3) feature-centric designs employing entropy families or specialized block-based convolutional networks for continuous decoding; and (4) neuromorphic systems that classify spike-encoded, engineered features.

Across these diverse methodologies, a consistent set of challenges emerges, including high inter-subject variability, signal noise and non-stationarity, methodological heterogeneity that impedes comparability, and the computational demands of deep learning models. In response, the reviewed literature highlights several practical solutions. Subject-independent evaluation protocols, such as Leave-One-Subject-Out Cross-Validation (LOSO-CV), are essential for assessing model generalization. Standardized preprocessing pipelines incorporating filtering, artifact removal using Independent Component Analysis (ICA), and fixed windowing are necessary to ensure reproducibility. To manage computational constraints, strategies such as utilizing early-layer convolutional features, designing compact architectures, and exploring low-power alternatives like spiking neural networks have demonstrated potential.

The overall trajectory of the field points toward the systematic integration of these approaches. Continued progress depends not on the discovery of a single dominant algorithm but on the principled construction of pipelines that balance the representational strength of deep learning with the interpretability and efficiency of engineered features. Future research should prioritize the development of standardized,

computationally efficient, and reproducible systems capable of transferring from controlled experimental settings to real-world emotion recognition applications.

REFERENCES

[1] A. Hag, D. Handayani, M. Altalhi, A. Almalki, and H. Alquhayz, "Enhancing EEG-based mental stress state recognition using an improved hybrid feature selection algorithm," *Sensors*, vol. 21, no. 24, p. 8370, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/24/8370

[2] H. A. Hamzah and K. K. Abdalla, "EEG-based emotion recognition systems; comprehensive study," *Heliyon*, vol. 10, no. 10, p. e31485, 2024. [Online]. Available: https://www.cell.com/heliyon/fulltext/S2405-8440(24)07516-9

[3] N. Bagherzadeh, S. S. Mirzapour, H. Golmohammadi, K. Massoudi, and A. Lotfi, "A hybrid EEG-based emotion recognition approach using wavelet convolutional neural networks and support vector machine," *Basic and Clinical Neuroscience*, vol. 14, no. 1, pp. 87–102, 2023. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10279985

[4] B. Chakravarthi, S. Ng, M. R. Ezilarasan, and M. Leung, "EEG-based emotion recognition using hybrid CNN and LSTM classification," *Frontiers in Computational Neuroscience*, vol. 16, p. 1019776, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fncom.2022.1019776/full

[5] D. Jung, J. Choi, J. Kim, S. Cho, and S. Han, "EEG-based identification of emotional neural state evoked by virtual environment interaction," *International Journal of Environmental Research and Public Health*, vol. 19, no. 4, p. 2158, 2022. [Online]. Available: https://www.mdpi.com/1660-4601/19/4/2158

[6] P. Patel, R. R, and R. N. Annavarapu, "EEG-based human emotion recognition using entropy as a feature extraction measure," *Brain Informatics*, vol. 8, no. 1, p. 20, 2021. [Online]. Available: https://braininformatics.springeropen.com/articles/10.1186/s40708-021-00141-5

[7] W. Lee, Y. Jung, S. Kim, H. Park, C. Park, and D. Kim, "Continuous EEG decoding of pilots' mental states using multiple feature block-based convolutional neural network," *IEEE Access*, vol. 8, pp. 184 681–184 692, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9133061

[8] J. Chen, Y. Cui, C. Wei, K. Polat, and F. S. Alenezi, "Advances in EEG-based emotion recognition: Challenges, methodologies, and future directions," *Applied Soft Computing*, p. 113478, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494625007896

[9] Y. Luo, M. Wu, F. Wang, J. Zhang, L. Ding, Y. Chen, D. Guo, S. Wan, and D. A. Clifton, "EEG-based emotion classification using spiking neural networks," *IEEE Access*, vol. 8, pp. 46 007–46 016, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9024211

[10] X. Wang, Y. Ren, Z. Luo, W. He, J. Hong, and Y. Huang, "Deep learning-based EEG emotion recognition: Current trends and future perspectives," *Frontiers in Psychology*, vol. 14, p. 1126994, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1126994/full

[11] R. B. Pachori, *Time-Frequency Analysis Techniques and Their Applications*, 1st ed. Boca Raton: CRC Press, 2023. [Online]. Available: https://www.taylorfrancis.com/books/mono/10.1201/9781003367987

[12] M. P. Pauli, C. Pohl, and M. Golz, "Balanced leave-one-subject-out cross-validation for microsleep classification," in *Current Directions in Biomedical Engineering*, vol. 7, no. 2, 2021, pp. 147–150. [Online]. Available: https://www.degruyter.com/document/doi/10.1515/cdbme-2021-2038/html

[13] D. M. W. Powers, "Evaluation: From precision, recall and f-measure to ROC, informedness, markedness & correlation," 2020. [Online]. Available: https://arxiv.org/abs/2010.16061

[14] S. Shanmugasundaram and N. Palaniappan, "Detection accuracy improvement on one-stage object detection using AP-loss-based ranking module and resnet-152 backbone," *International Journal of Image and Graphics*, vol. 24, no. 03, p. 2450030, 2024. [Online]. Available: https://www.worldscientific.com/doi/10.1142/S021946782450030X