DECS: A Deep Learning Approach for EEG Channel Selection in Emotion Classification

Muhammad Saad Hussain[1], Areebah Mubarak[2], Amna Arshad[3]
[1]Software Engineering Department (Army Public College of Management Sciences, APCOMS).
[2]Department of Cyberspace Science and Technology (Beijing Institute of Technology, Haidan District, Beijing).
[3]Computer Science and Technology Department (Beijing Institute of Technology, Haidan District, Beijing).
Correspondence: amna.arshad@bit.edu.cn

Citation | Hussain. M. S, Mubarak. A, Arshad. A, "DECS: A Deep Learning Approach for EEG Channel Selection in Emotion Classification", IJIST, Volume 07. Issue 04. pp 2290-2310, October 2025
DOI: https://doi.org/10.33411/IJIST/1591
Received| August 24, 2025 Revised| October 01, 2025 Accepted| October 03, 2025 Published| October 05, 2025.

The non-stationary nature of Electroencephalogram (EEG) signals often leads to high computational complexity in emotion recognition systems. To address this, we propose a novel framework that integrates optimal channel selection with efficient feature extraction. Our method begins by converting preprocessed EEG signals into two-dimensional spectrograms using a Continuous Wavelet Transform (CWT). These spectrograms are then processed by a Google Net model for deep feature extraction. A key contribution is the Differential Entropy-based Channel Selection (DECS) technique, which identifies and retains the most informative channels. To manage dimensionality, the extracted features are encoded using the Bag-of-Deep-Features (BoDF) method, which employs k-means clustering to create a visual vocabulary and represents features as histograms. Finally, these histogram features are classified using a Support Vector Machine (SVM). Evaluated on the SJTU SEED and DEAP datasets, the proposed model achieves state-of-the-art classification accuracies of 95.1% and 81.1%, respectively, demonstrating its effectiveness and efficiency.

Keywords: Bag-of-Deep Features; Continuous Wavelet Transform; Differential Entropy-based channel selection; Support Vector Machine.

Introduction:

Biomedical engineers have long employed Brain-Computer Interfaces (BCIs) to control devices using brain signals[1]. To recognize and classify human emotions, Electroencephalographic (EEG) signals generated by the brain are captured using electrodes placed on the scalp[2]. Although extensive research has been conducted on recognizing and classifying human emotions, machines still find the process difficult to accomplish. The demand for human emotion assessment is at an all-time high due to the advancements in machine learning tools[3] [4].

Perceptual experience and apprehensiveness are of prime importance when it comes to the emotional states of humans. The major fields of affective computing, Emotion detection, play a fundamental role when it comes to providing emotional care to people. Human emotions can be detected through several approaches, such as analyzing visual cues, speech variations, and video data. However, these systems are often complex and costly to implement. Detection and classification of emotions via EEG signals is a particularly challenging task because of insufficient time boundaries. The subjects might depict differing emotional reactions because emotional states have no definite margins[5]. While emotions have traditionally been detected through speech, images, and videos, Brain-Computer Interface (BCI) technology now offers a more direct gateway. BCIs track brain signals to help us understand emotional responses. However, a significant challenge remains that we are still largely unable to fully decode this complex neural information[6] [7]. The bio-signals acquired from these techniques can help uncover insights into the psychological and emotional state of the user[8]. The signal acquisition process is done through the 10-20 International System of electrode placement[9] [10].

In previous work, researchers have typically either reduced the computational overhead by limiting the number of features or improved decision precision by utilizing all features separately. When applied independently, these techniques either lower the overall computational overhead or enhance decision accuracy. The challenge emerges in real-time applications, where limited time is available to detect emotions from EEG signals and high decision precision is required. To address this, the proposed algorithm aims to extract quality features while simultaneously reducing computational overhead. The proposed work is not as good as needed for real-time, but a sufficient improvement from the previous work.

The remainder of this paper is organized as follows: Section II reviews the related literature, while Section III describes the datasets employed in this study. Section IV outlines the methodology, detailing the stages of TFR, FE, DECS, BoDF, and classification. Section V presents the results of the proposed algorithm along with a comparative analysis against previous work. Finally, Section VI concludes the paper.

Existing methods address channel selection and feature reduction in isolation, often relying on shallow statistical measures or simplistic aggregation. This creates a clear research gap: the lack of an integrated framework that leverages the power of deep learning to simultaneously identify emotionally salient brain channels and compactly represent their high-dimensional features.

To bridge this gap, this paper proposes a novel methodology that synergistically combines a Differential Entropy-based Channel Selection (DECS) technique with a Bag-of-Deep-Features (BoDF) model. The novelty of our work is threefold:

We introduce DECS, which uses differential entropy robust measures of signal complexity—derived from deep feature maps to systematically identify and select the most discriminative EEG channels for emotion recognition, thereby reducing computational overhead and enhancing model focus.

We adapt the BoDF model, inspired by computer vision, to effectively reduce the dimensionality of deep features. By creating a 'visual vocabulary' of feature clusters, BoDF

transforms the high-dimensional feature set into a compact histogram representation, mitigating the curse of dimensionality without significant information loss.

The primary novelty lies in the integrated use of DECS and BoDF. DECS ensures that only the most relevant channels are processed, and BoDF provides an efficient, fixed-length representation of their complex features. This combined approach is specifically designed to enhance both the accuracy and efficiency of EEG-based emotion classification.

The proposed model, when evaluated on the SJTU SEED and DEAP datasets, achieves state-of-the-art performance, demonstrating the effectiveness of this integrated approach.

Related Work:

This section reviews previously reported studies on human emotion recognition. In[11], the authors discuss methods for acquiring EEG signal data, along with its preprocessing and enhancement. They evaluate various feature extraction techniques and classification algorithms, with classification accuracy serving as the primary performance metric.

In [12], an EEG-based multi-modal model for the purpose of emotion recognition employs a multi-layered deep neural network for performing classification. Raw EEG signals are transformed into 2-D spectrograms that retain spatial and temporal information. Feature extraction from these spectrograms is then performed for each channel using a pre-trained AlexNet model. In[13], a study with ten participants analyzes seven emotions. The signals are filtered using a Savitzky-Golay (SG) filter to identify the channels that deliver maximum accuracy. Employing the Differential Entropy-based Channel Selection algorithm (DECS) increases the classification accuracy for Linear Discriminant Analysis (LDA) to 86.85%, up from 80%. In[14], the authors propose a technique that converts 32-lead EEG signals into 2-D images using the Azimuthal Equidistant Projection (AEP) method. These images are then processed using Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs), achieving an accuracy of up to 96.09%. In[15], the authors propose an emotion detection system based on an advanced convolutional neural network, implemented on a very large-scale integration (VLSI) hardware design. The designed model was validated from a set of 32 subjects that were taken from the DEAP dataset, and the average classification accuracy of 83.88% was attained. The[16] offered an emotion recognition that extracts features by application of Compressed Sensing (CS), Local Binary Patterns (LBP), and Wavelet Transformation (WT). Furthermore, Support Vector Machine (SVM) and Fuzzy Cognitive Maps (FCM) are unified to formulate an algorithm that provides an accuracy of up to 75.64% from the feature patterns. In [17], the authors discuss the techniques that are used for emotion recognition, also survey, and examine several parameters that include accuracy, specificity, and sensitivity. The proposed technique achieves 89% accuracy in emotion recognition and classification.

In[18], the authors introduce a learning algorithm designed to identify the most distinct EEG channels suitable for recognizing internal emotions. The approach employs kernel-based representations, which are computationally derived from EEG signals. The algorithm is used to reduce the data with an improvement in computational efficacy and classification accuracy carried out alongside. In[4], the authors proposed a model for emotion assessment in which the perceptual experience and apprehensiveness of the subject played an important role. They used the MFM and CapsNet classifiers on the DEAP dataset, achieving an accuracy of 68.2%. Several interfacing techniques have been reported for extracting EEG signals from the brain to facilitate emotion detection. However, there remains considerable scope for improvement, particularly in enhancing accuracy and interpretability while extracting spatial features. The normalization of common spatial patterns (CSP) is among the most consistent methods[19], as it ensures a reduction in noise along with accompanying artifacts that are known to exist in

EEG signals of the unprocessed kind. In [20], the authors combined Virtual Reality (VR) and the Internet of Things (IoT) to enable real-time control over both the virtual and physical environments of a user. They utilized the CSP features in virtual reality for the process of motor rehabilitation. In[21], the authors have suggested a framework that applies the Fisher ratio criterion and estimates the optimal weights of filters using the available data in an automatic fashion. When evaluated experimentally, the performance of the proposed model based on dynamic features is comparable with that obtained upon using CSP and AR features.

Efforts to achieve robust emotion recognition from EEG signals have traditionally relied on signal processing techniques that assume stationarity, such as the Fourier Transform and Discrete Wavelet Transform (DWT)[22] [23]. However, the inherently subjective and complex nature of emotional states presents a significant obstacle to developing a universal analytical framework. EEG data is inherently non-stationary and non-linear, with statistical properties that vary over time, making the accurate characterization of transient emotional cues more challenging. This complexity underscores the critical importance of channel and feature selection, as the inclusion of irrelevant or redundant data can severely compromise detection accuracy. Furthermore, the high dimensionality of multi-channel EEG recordings often leads to substantial computational costs, especially when dealing with large-scale datasets required for training sophisticated models. To address this, prior research has explored signal decomposition methods such as Empirical Mode Decomposition (EMD) and Wavelet Packet Decomposition (WPD)[24] [25] [26]. These techniques aim to break down the raw signal into a series of constituent components, facilitating the handling of vast data volumes by extracting a more manageable set of informative features.

Parallel research in speech emotion recognition (SER) has also seen notable advancements, driven by the shared objective of improving human-computer interaction (HCI). One such study presents an Artificial Intelligence-assisted Deep Stride Convolutional Neural Network (DSCNN) as a novel architecture[27]. This model leverages a 'plains nets' strategy within the DSCNN to identify distinctive patterns directly from speech spectrograms. When tested on the IEMOCAP and RAVDESS datasets with a SoftMax classifier, this approach achieved performance improvements of 7.85% and 4.5%, respectively, while also reducing the model size by 34.5 MB. Such cross-modal innovations highlight the potential of advanced neural architectures for complex pattern recognition in affective computing.

Feature selection is a fundamental process for obtaining valuable features, as it helps reduce the dimensionality of the feature set. This technique helps identify the most relevant features in the classification domain. Broadly, feature scoring methods are categorized into four classes: statistical-based, information-theoretic, sparse learning-based, and agreement-based approaches[28]. Although there is are variety of procedures for scoring features that are proposed by many scholars in [29] [27]. Non-negative Laplacian is one of the estimation techniques to identify the contribution of a feature in feature selection using unsupervised learning[30]. Principal Component Analysis (PCA) is a technique used by several scholars in the field of emotion detection using EEG signals to reduce the size of features by picking procedures depending on their exclusive value. Ignoring features with uncorrelated quantities is a conventional clipping procedure. For optimization of statistical features, a dynamic search strategy is proposed in[31]. The Receiver Operating Characteristics (ROC) is used for feature selection to find valuable features. This technique helps to decrease the size of features while providing high accuracy for classification in the case of an electrocardiography (ECG) signal.

The authors in[25] employed a multi-scale PCA along with WPD for fragmentation and extermination of noise from the signal. A classification accuracy of 92.8% was obtained while classifying EEG signals for motor rehabilitation. In[32], the authors propose a flexible analytical wavelet-based decomposition (FAWT) method applied to the SEED dataset. Using an SVM classifier, they achieve a classification accuracy of 83.3%. FAWT is also a channel-

specific process. A limitation of this approach is that some features may appear irrelevant individually but can provide valuable information when considered in combination with others. Another demerit is the ability to choose functions that are individually related, which might cause duplication. The authors in [33] proposed a feature selection algorithm based on decomposition, which was evaluated on DEAP and MAHNOB datasets. The literature demonstrates various sophisticated approaches to optimizing feature selection and classification for EEG-based emotion recognition. For instance, one methodology integrated a Probabilistic Neural Network (PNN) classifier with a feature selection mechanism driven by Differential Evolution (DE). This hybrid model was reported to achieve classification accuracies of 77.8% on the MAHNOB dataset and 79.3% on the DEAP dataset, showcasing its effectiveness.

Further investigations have focused on the informational content of specific EEG rhythms. A notable study[28] quantified the differential entropy features from standard frequency bands, identifying beta and gamma rhythms as the most discriminative for emotional valence and arousal. To capture the complex dynamics of these signals, the authors extract a comprehensive set of 18 linear and non-linear features from the time-frequency domain. These features were then processed using a Spatial-Temporal Recurrent Neural Network (STRNN), an architecture specifically designed to model the inherent dependencies across both space (channels) and time. To address the challenge of generalizability, the study in[29] systematically examines key factors influencing cross-target emotion recognition, such as the dynamic properties of EEG signals and inconsistencies among datasets with varying channel montages. To enhance model efficiency and performance, the authors introduced a Recursive Feature Elimination (RFE) technique, which iteratively prunes the least important features. This method yielded an average accuracy of 60.5% on the DEAP dataset and a notably higher 90.33% on the SEED dataset, underscoring the impact of dataset characteristics on final performance.

Conventional methodologies for EEG-based emotion recognition have historically presented a trade-off between performance and practicality. One common strategy utilizes the full array of EEG channels, a method that typically yields high classification accuracy but at the cost of significant computational burden. The alternative approach involves pre-emptively reducing the number of channels to enhance processing speed; however, this often leads to a substantial decline in predictive performance due to the loss of critical neurophysiological information.

This study introduces a novel framework designed to transcend this traditional compromise. The proposed model integrates two key components to achieve both high accuracy and computational efficiency simultaneously. First, a Differential Entropy-based Channel Selection (DECS) algorithm identifies and preserves the most informative EEG channels, discarding redundant ones. Subsequently, a Bag-of-Deep-Features (BoDF) method is applied to further condense the feature space. This two-stage process ensures that the model operates on a compact, high-fidelity representation of the original EEG signal, enabling robust emotion classification without prohibitive computational demands.

Material and Methods:

Dataset 1 (SEED). The experimental data for this study are obtained from the publicly available SJTU Emotion EEG Dataset (SEED), developed by Professor Bao-Liang Lu and his team at the Brain-Like Computing and Machine Intelligence (BCMI) Laboratory [27]. This dataset was specifically designed to elicit and record neural correlations of distinct emotional states.

During the data acquisition phase, fifteen subjects (seven male and eight female) participated in experiments where they viewed a series of fifteen carefully selected Chinese film clips. As illustrated in Figure 1, these clips were intended to provoke one of three target

emotional responses: positive, negative, or neutral. Stimulus selection followed strict criteria to ensure data quality and ecological validity: the experiment duration was controlled to prevent participant fatigue, video clips were non-verbal or easily understandable to minimize cognitive bias, and each was designed to elicit a single, dominant emotional state. The experimental protocol was structured into discrete trials. Each trial began with a 5-second preparatory cue, followed by the 4-minute emotional stimulus. Subsequently, participants were given a 45-second period for subjective emotional self-assessment, concluding with a 15-second rest interval. To prevent emotional carry-over effects, clips targeting the same category were not presented consecutively. Furthermore, participant feedback was systematically collected through post-experiment questionnaires to validate the emotional induction [27].

Dataset 2 (DEAP). The Database for Emotional Analysis using Physiological Signals (DEAP) dataset was collected by young researchers at Queen Mary University. An analysis of effective human states is presented in the dataset. The stimuli from 32 participants (and frontal face videos from 22 participants) were collected by exposing them to 40 selected videos, each one minute in length. Familiarity, dominance, like or dislike, valence, and arousal were the parameters on which the participants ranked each video. Affective Tags from last.FM website, video highlight detection, and an online assessment tool were incorporated for the selection of the generated stimuli. The video selection process began with 120 YouTube videos, of which half were selected manually and the remainder semi-automatically.

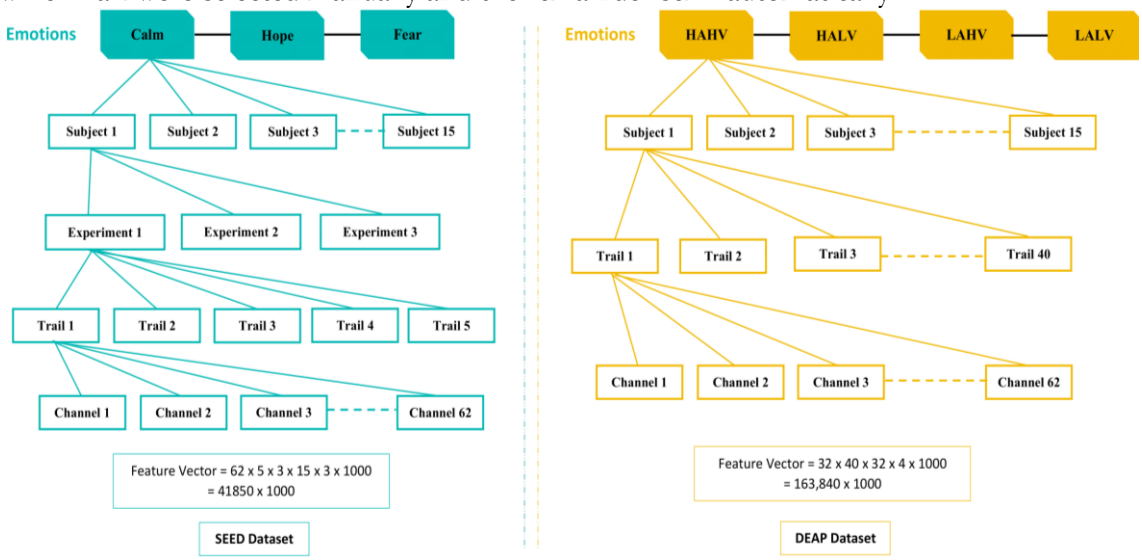


Figure 1. Raw Feature Dimension of SEED and DEAP Datasets.

These 120 videos were ranked subjectively, and the top 40 videos were selected to be presented to 32 subjects. The subjects were 50% female and 50% male; the average age was 26.9, ranging from 19 to 37 years. After exposure to each video, the participants were asked to fill out a questionnaire for self-assessment of their arousal, liking, and dominance on a scale of 1 to 9 (1 represents a low state while 9 represents a high state). Valence, arousal, and liking provide information about sad/happy, calm/excited, dislike/like. The EEG signal generated by stimuli was collected over the 10-20 international system. The signals were collected by utilizing a 32-channel array at 512 cycles/second. The outliers were removed using pre-processing and down-sampled to 128Hz.

Methodology:

In this section, we described the methods that were employed to implement the proposed model. The following Figure 2 shows the architecture of our framework. The frames were discussed step by step in the following sections. In the Preprocessing step, one-dimensional EEG signals are represented as a two-dimensional spectrogram of time and frequency, known as Time-frequency representation (TFR). DNN is used to extract features.

To reduce the feature dimension and select the high-quality features. The DECS (Channel Selection) technique is proposed along with BoDF (Feature Selection).

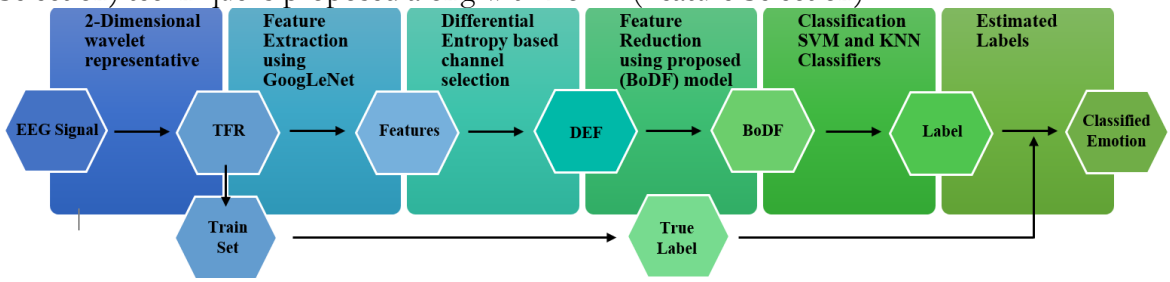


Figure 2. The proposed framework for the classification of Emotions from EEG signals of the SEED and DEAP datasets.

The proposed framework for emotion classification initiates by transforming the preprocessed EEG signals into a two-dimensional time-frequency representation using the Continuous Wavelet Transform (CWT). This process yields a spectrogram for each signal, which effectively captures its non-stationary characteristics. These spectrograms are subsequently fed into the GoogLeNet architecture, pre-trained on image recognition tasks, to extract a comprehensive set of high-level, deep features for all subjects.

To refine this feature set, the Differential Entropy-based Channel Selection (DECS) algorithm is applied. This step functions as an intelligent filter, identifying and retaining only those EEG channels that contribute the most discriminative information for emotion recognition, thereby enhancing the signal-to-noise ratio of the data. Following channel selection, the Bag-of-Deep-Features (BoDF) model was employed for further dimensionality reduction. This involves using k-means clustering to generate a compact "visual vocabulary" from the deep features, where each cluster centroid represents a fundamental pattern.

Finally, this learned vocabulary is used to encode the feature set into a standardized histogram representation. These histograms are then presented to multi-class classifiers for the final emotion assessment. The model is evaluated on its ability to discriminate between the three emotional states—neutral, positive, and negative—present in the SEED dataset, with performance being benchmarked across various kernel functions within the classifiers.

Time Frequency Representation. The time-frequency representation of the signals was done by utilizing the preprocessed EEG signals. Poor and low accuracy results were obtained from the traditional emotion recognition model, which takes out features straight from the preprocessed EEG signals. In our work, to visualize the EEG signal and extract the desired features, the one-dimensional EEG signal was represented as a two-dimensional time-frequency representation (TFR) using the Continuous Wavelet Transform (CWT). The reason for using CWT over other techniques was to acquire full knowledge of signal frequency in the temporal and spatial domains.

Continuous Wavelet Transformation. The preprocessed EEG signals were mapped into a joint time-frequency domain using the Continuous Wavelet Transform (CWT) to construct a Time-Frequency Representation (TFR)[31]. The CWT decomposes a signal by convolving it with a set of wavelets that were scaled and shifted versions of a fundamental mother wavelet.

Let x(ξ) represent the continuous-time EEG signal. The CWT is defined as a two-dimensional function, T_x (σ,ν), which quantifies the correlation between x(ξ) and a wavelet basis function. $\psi_{\sigma,\nu}(\xi)$

$$T_x(\sigma,\nu) = \int_{-\infty}^{\infty} x(\xi)\psi^*_{\sigma,\nu}(\xi)\, d\xi \quad \#(1)$$

In this formulation:

$\sigma \in \mathbb{R}^+ +$ is the dilation parameter, inversely proportional to frequency.

$\nu \in \mathbb{R}$ is the translation parameter, indicating temporal location.

The family of wavelets $\psi_{\sigma,\nu}(\xi)$ is generated from the mother wavelet $\psi(\xi)$ through:

$$\psi_{\sigma,\nu}(\xi) = \frac{1}{\sqrt{\sigma}} \psi\left(\frac{\xi - \nu}{\sigma}\right) \#(2)$$

This factor $1/\sqrt{\sigma}$ ensures energy preservation across different scales.

For the transformation to be invertible, the mother wavelet must satisfy the admissibility condition, which requires it to have zero mean and finite energy. The corresponding stability constant,K_ψ is given by:

$$K_\psi = \int_0^\infty \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < \infty \#(3)$$

Where ψ^ (ω) is the Fourier transform of ψ (ξ). This condition guarantees that the original signal can be reconstructed from its CWT coefficients.

In practice, a filter-bank approach was used to compute the CWT for each EEG channel, generating a distinct TFR. A visual analysis confirms discernible differences in the TFR patterns across different experimental classes. These TFR images are subsequently converted to grayscale scalograms and resampled to a uniform size of 227×227 pixels. A convolutional neural network is then trained on this dataset, leveraging a 20-fold cross-validation protocol to optimize performance before proceeding with feature extraction.

Feature Extraction:

In our working model, the process of feature extraction was done by the DNN termed as GoogLeNet. GoogLeNet is a multi-layer DNN comprised of 22 layers[34] as shown in Figure 3. GoogLeNet utilizes 5 million parameters than its forerunner, AlexNet, which had 60 million parameters. So, it can deliver even under the primary confinements of memory and computational cost[35]. GoogLeNet won the ILSVRC 2014, and it used 12 fewer layers than the winning network from two years ago[34]. It is a personification of Inception architecture submitted at ILSCRC 2014 competition as it comprises recurring Inception layers in its model. The 2-D spectrogram images were fed to the 22 layers, which applied the process of convolution and max pooling.

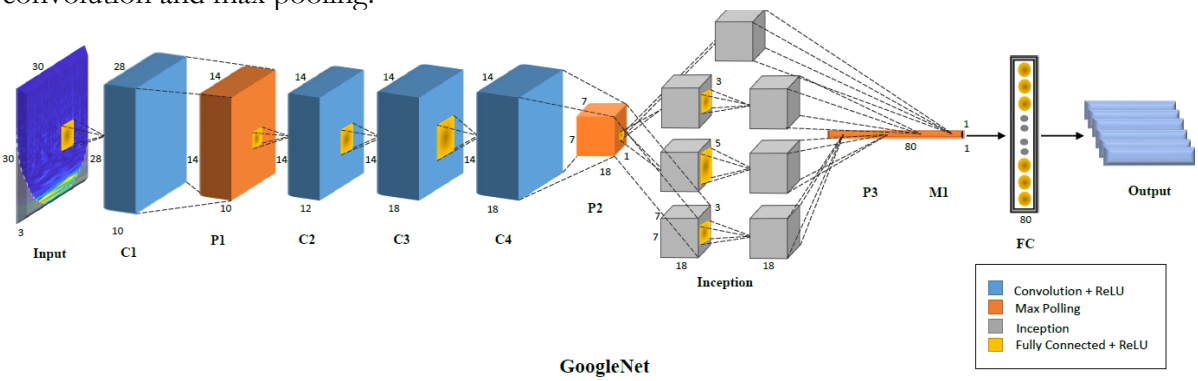


Figure 3. The GoogLeNet Architecture.

The architectural backbone of this work is derived from the Inception model. In this network, all convolutional layers employ the Rectified Linear Unit (ReLU) as their activation function. The model was designed to process input images of size 224×224 pixels within the RGB color space, which have been preprocessed by centering their mean to zero.

A key efficient feature of the Inception architecture is its use of dimensionality reduction. This is achieved through specialized layers that utilize 1×1 convolutional filters, often designated as '3×3 reduce' and '5×5 reduce' layers. These layers are strategically placed prior to the more computationally intensive 3×3 and 5×5 convolutions, serving to project the input into a lower-dimensional

EEG Channel Selection using Differential Entropy. The process of channel selection is of crucial importance when it comes to omitting irrelevant and redundant features while maintaining the quality of the selected features. By employing an effective method to extract quality features, the total amount of the features is significantly reduced without degrading the standard of classification. In this paper, a decision-making algorithm was proposed to evaluate the relevance of feature subsets. It enabled efficient representation of ambiguous data at the boundaries of these subsets, such as in rough sets or fuzzy-rough sets. The recommended algorithm uses differential entropy to obtain novel features by evaluating the feature subsets.

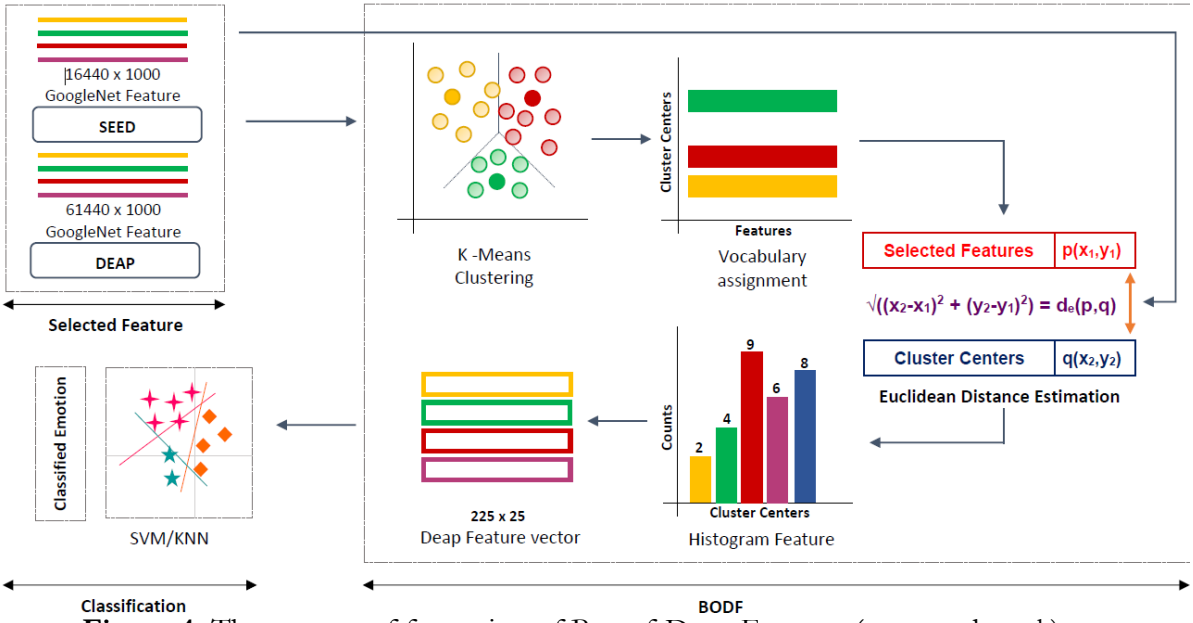


Figure 4. The process of formation of Bag-of-Deep Features (proposed work).

Effective channel selection is a critical step in building a robust BCI system, as it aims to eliminate non-informative and redundant data streams while preserving the core discriminatory information. A well-designed selection strategy can drastically reduce the computational burden and model complexity without compromising, and sometimes even enhancing, classification performance. This paper introduces a feature evaluation algorithm grounded in information-theoretic principles, which quantifies the informational sufficiency of a candidate feature subset. This method is particularly adept at handling the inherent uncertainty and imprecision found in the decision boundaries of neurophysiological data, a challenge often addressed by techniques like rough sets.

The proposed framework is versatile, capable of processing both categorical and continuous-valued data[36]. At its core, we leveraged the concept of conditional entropy to accurately measure the discrepancy between the knowledge represented by a subset of channels and that of the complete set. This measure provides a powerful and efficient means to assess a feature subset's quality, yielding several desirable theoretical properties.

Consider a decision system represented by the tuple (U, F∪D), where U is the universe of instances, F is the full set of features (EEG channels), and D is the decision attribute (e.g., emotion class). For any feature subset S⊆F, we define its informational divergence $D(S \mid U, F$ as

$$D(S \mid U, F) = \frac{-1}{|U|} \sum_{x \in U}^{x} log_2 \left(\frac{[x]F \cap [x]S}{[x]S}\right) \#(4)$$

Here, $[x]S$ denotes the equivalence class of instance x induced by the feature subset S, i.e., the set of all instances in U indistinguishable from x based on the features in S. Since S is a subset of F, it follows that $[x]F \subseteq [x]S$, guaranteeing the non-negativity of the measure.

An alternative, equivalent formulation of the divergence is:

$$D(S \mid U,F) = -\frac{1}{U}\sum_{x\in U}^{x} log_2 \frac{|[x]F|}{|[x]S|} \#(5)$$

This measure exhibits key monotonicity and equivalence properties. For nested subsets $S1 \subset S2 \subset$F, the following holds:

$$D(S2 \mid U,F) \leq D(S1 \mid U,F) \#(6)$$

Furthermore, if the partitions induced by $S1$ and $S2$ are identical (i.e., $/S1 = U/S2$), then their divergent values are equal:

$$D(S2 \mid U,F) = D(S1 \mid U,F), if\ U/S1 = U/S2 \#(7)$$

Interpretation and Significance:

The divergence D(S|U, F) serves as a direct metric for the representational power lost by using the subset S instead of the full feature set F. A higher value indicates a greater loss of information and a larger discrepancy between the granularity of S and F. Crucially, when D(S|U, F)=0, it implies U/S=U/F. This means the subset S induces the same partition over the data as the full set F, proving that S is informationally equivalent to F and is therefore a minimal and sufficient feature set for the task[36].

Bag-of-Deep Features (BoDF). In computer vision, this procedure was used to reduce the dimensionality of feature sets for classification[37]. Specifically, it was applied to limit the number of features to a desired value. This method helps in reducing the feature size and the time required to train the dataset, in contrast to processing a massive number of features, which results in low accuracy classification and excessive training time. In our proposed method, all channels were used for emotion identification, resulting in high-dimensional features that demanded significant computational resources for dataset training. In the previous studies, only 8 to 12 channels were utilized, which resulted in a reduction in accuracy and the number of channels employed for classification[38].

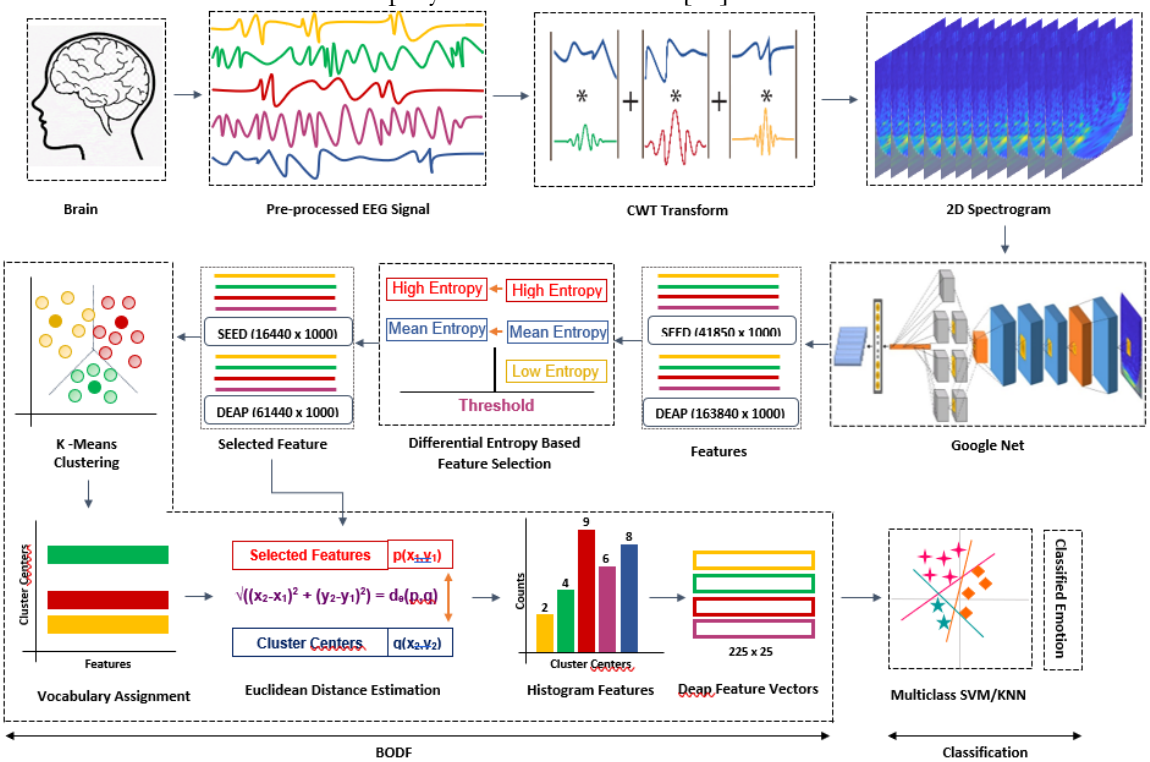


Figure 5. The overall flow diagram of the proposed model.

In this model, we utilized the SEED dataset, which has a dimension of 41850×1000 number of features when reduced to 17550×1000 from 26 channels out of 62, which are fed

to BoDF, as shown in Figure 5. The BoDF model comprises two stages by which features are taken out from Google-Net and are reduced by utilizing k-means clustering. K-means clustering accumulates similar features into one vector. So, in the first stage, the features are lowered by using mean clustering. In the second stage, further reduction is done by computing the histogram features. This process chooses a specific number of features that are closer to the centroid, while the other is neglected.

K-means Clustering. The BoDF stage, in which selected channels are grouped using the DECS technique for each class. Grouping is employed to accumulate similar features in the same class. This algorithm works efficiently when dealing with large datasets, whilst other methods face the over-fitting problem[39]. This technique also helps in grouping unnecessary features by comparing the distances. At the start, the value of the number of clusters 'k^' that will be used in the gathering of the provided features is selected at random. We experimented with the value of k; the features were organized into clusters correctly at the value of k = 8. The value was increased up to 14, but values higher than 14 and lower than 8 had a large sum of squared error, so they were omitted. So, the features of the SEED dataset are organized into eight groups compared to each other.

The displacement for each number was measured by utilizing Euclidean distance. The feature is classified with that cluster, which provides the minimum range. The average of the quantities in that cluster is needed to get the average value for each given cluster. The collected mean values are then re-evaluated until the meaning of each cluster centroid converges. For the SEED dataset, the attributes of 17550×1000 are clustered in a 24×1000 feature vector. The 24×1000 for each class function vector is called the visual vocabulary. The decision process of cluster number k is an important task because it directly affects the results obtained. The hit-and-trial method algorithm is applied to obtain the optimal value. In this method, the difference was measured between the sum of quadratic error results at different values of k.
Histogram Features. The second phase of the pipeline focuses on transforming the localized channel features into a global image representation using a codebook-based approach, often conceptualized as building a "visual vocabulary."

In this model, each feature vector from a channel was treated as a local "descriptor." A codebook (or visual vocabulary) was previously generated, typically via clustering, where each cluster centre represents a "codeword." The core of this phase is to generate a histogram that aggregates these local descriptors. This histogram essentially counts the frequency with which descriptors from a given input sample are assigned to each codeword in the vocabulary. The computational process involves a comparison operation. For the SEED dataset, the feature matrix from a single sample, with dimensions of 225 × 24, is compared against a codebook derived from 26 representative channels. This comparison, typically performed using a distance metric such as Euclidean distance, assigns each of the 225 feature vectors to its closest codeword in the vocabulary. The final representation is a fixed-length histogram feature vector, H, where each bin H_i contains the count of feature vectors assigned to the i-th codeword. For the SEED dataset, this results in a histogram of size 225 × 24, where the bin values (the frequency counts) are observed to fall within a range of 0 to 30.

To optimize the representation for the three-class problem, a selective pooling strategy is employed. Instead of using the entire histogram, the ten most discriminative codewords (histogram bins) are identified for each class, leading to the selection of a compact feature subset of size 10×3=30. This process achieves a significant reduction in feature dimensionality while preserving the most salient frequency patterns for classification.

Classification Framework and Model Selection:

For feature extraction, the GoogLeNet (Inception) architecture was leveraged. The model was truncated at the "loss3-classifier" fully connected layer, discarding the final two layers responsible for the original 1000-class classification. This adaptation allows the network

to serve as a powerful feature extractor for the three-class problem inherent to the SEED dataset (Happy, Sad, Neutral). The high-dimensional features produced by this modified GoogLeNet, subsequently refined by the DECS-BoDF feature selection technique, were fed into two distinct classifiers: Support Vector Machine (SVM) and k-Nearest Neighbours (k-NN). This dual-classifier approach was employed to ensure robust and generalizable performance across different classification paradigms.

Support Vector Machine (SVM) Implementation:

The SVM classifier operates by constructing an optimal hyperplane that maximizes the separation margin between different classes in a high-dimensional feature space. During training, the algorithm learns these decision boundaries, effectively creating a clear delineation between the three emotional states. A key strength of SVM is its ability to handle non-linear classification tasks using kernel functions. These functions implicitly map the original features into a higher-dimensional space where the classes become linearly separable. In this study, we evaluated several kernel types, including linear, polynomial, and Gaussian Radial Basis Function (RBF), to identify the most effective mapping for the emotional EEG data.

k-Nearest Neighbours (k-NN) Implementation:

As a complementary approach, the k-NN classifier was implemented. This instance-based, or lazy-learning, algorithm classifies a test sample based on the majority vote of its 'k' closest neighbours in the feature space. The proximity between data points is quantified using a distance metric. For our model, the Euclidean distance was employed, which for two feature vectors p and q, each of dimensionality NN, is calculated as:

$$d(\boldsymbol{p},\boldsymbol{q}) = \sqrt{\sum_{i=1}^{N} (pi - qi)^2} \quad \#(8)$$

Where N=45 for the feature vectors derived from the SEED dataset. The parameter 'k', representing the number of neighbors considered, was tuned to optimize classification performance.

To ensure a statistically rigorous evaluation of both classifiers, a 20-fold cross-validation protocol was strictly followed. This process involves repeatedly partitioning the data into training and testing sets, providing a reliable estimate of model generalizability and mitigating the risk of overfitting. The complete procedural workflow of the proposed model is summarized in Algorithm 1.

Algorithm 1: Feature and Channel Selection Algorithm

Inputs:

A set of pre-processed, multi-channel EEG trials: X=\ {X_1, X_2,…, X_N \},
where N is the total number of trials.
A corresponding set of emotion labels: Y=\{y_1,y_2,…,y_N \}.
Pre-trained GoogLeNet model.
Number of optimal channels to select, K.
Size of the visual vocabulary, V.
SVM kernel type and hyperparameters.

Outputs:

A trained Support Vector Machine (SVM) classifier model, 〖SVM〗 _model.
Classification Accuracy on the test set.

Procedure:

Step 1: Time-Frequency Representation via Continuous Wavelet Transform (CWT)
for each trial $X_i \in \boldsymbol{X}$ **do**
for each EEG channel c in trial X_i do

Let $s_{i,c}(t)$ Be the raw temporal signal from channel c of trial X_i.
Apply the Continuous Wavelet Transform (CWT) to $s_{i,c}(t)$ to generate a 2D spectrogram image $I_{i,c}$.
end
end
Step 2: Deep Feature Extraction using GoogLeNet
for each trial $X_i \in \boldsymbol{X}$ **do**
for each channel spectrogram $I_{i,c}$ in trial X_i do
Pre-process $I_{i,c}$ (resize to 224×224 pixels, normalize pixel values).
Feed $I_{i,c}$ Into the pre-trained GoogLeNet.
Extract the feature activation vector. $\boldsymbol{f}_{i,c}$ from a pre-final layer ('$pool5 - 7 \times 7_s1$' layer).
end
end
Step 3: Differential Entropy-based Channel Selection (DECS)
Let $\boldsymbol{S}_c$ be the set of all feature vectors $\boldsymbol{f}_{i,c}$ From all trials for a specific channel c.
For each channel c **do**
Calculate the channel selection score $Score(c)$. This can be implemented in one of two ways:
Option A (From Raw EEG):
Compute the average Differential Entropy:

$$\text{Score}(c) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\log\left(2\pi e.\sigma_{i,c}^2\right)$$

where $\sigma_{i,c}^2$ is the variance of the raw signal $s_{i,c}(t)$.
Option B (From Deep Features):
Compute the average magnitude of the deep features: $Score(c) = \frac{1}{|S_c|}\sum_{f\in S_c}||f||_2$.
end for
Rank all channels based on $Score(c)$ In descending order.
Select the top K channels to form the optimal channel set. $C_{optimal}$.
Step 4: Bag-of-Deep Features (BoDF) Vocabulary Construction
Construct a global feature pool using only the selected optimal channels:

$$\mathbf{F}_{\text{pool}} = \{\mathbf{f}_{i,c} \mid \forall i, \forall c \in \mathcal{C}_{\text{optimal}}\}$$

Apply k-means clustering to $\mathbf{F}_{\text{pool}}$to learn the visual vocabulary: $\{v1, v2, \ldots, vV\} = k - means(\boldsymbol{F}_{\text{pool}}, V)$
where $\boldsymbol{v}_j$ is the $j - th$ cluster centroid (codeword).
Step 5: Trial-Level Histogram Encoding
for each trial $X_i \in \boldsymbol{X}$ **do**
Initialize a histogram vector. $\boldsymbol{h}_i = [h_i^1, h_i^2, h_i^3 \ldots, h_i^v]$ with zeros.
for each feature vector $\boldsymbol{f}i, c$ from the selected channels $c \in \mathcal{C}_{\text{optimal}}$ in trial X_i **do**
Find the index j^* Of the nearest codeword:

$$j^* = argmin_{j\in\{1,\ldots,V\}} || \boldsymbol{f}_{i,c} - \boldsymbol{v}_j ||_2$$

Increment the corresponding histogram bin: $h_i^{j^*} = h_i^{j^*} + 1$.
end for
Normalize the histogram $\boldsymbol{h}_i$ by the total number of feature vectors in the trial X_i to form the final trial representation.
end for

Step 6: Model Training and Evaluation

Split the dataset of histogram representations $\{\boldsymbol{h}_i\}$ and labels $\boldsymbol{Y}$ into the training set $(\boldsymbol{H}_{train}, Y_{train})$ and testing set $(\boldsymbol{H}_{test}, Y_{test})$.

Train an SVM classifier on the training set:

$$SVM_{model} = SVM.train(\boldsymbol{H}_{\text{train}}, \boldsymbol{Y}_{\text{train}})$$

Predict labels for the test set:

$$\boldsymbol{Y}_{pred} = SVM_{model}.predict(\boldsymbol{H}_{\text{test}})$$

Calculate the classification accuracy:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{|\ Y_{test}\ |} \times 100\%$$

return $SVM_{model}, Accuracy$

Result and Discussion:

This section presents the results of the proposed model. During model design, we evaluated five deep neural networks—GoogLeNet, AlexNet, ResNet-50, ResNet-101, and Inception-ResNetV2—to identify the network that delivers optimal performance on the SEED dataset. The histogram features were classified using SVM, k-NN, Tree, and Ensemble classifiers by all kernels. The kernels that are shown in Table 1 are the ones that provide optimal accuracy results.

Figure 6 shows the improvement in accuracy by using the BoDF technique in the algorithm. This algorithm has been applied to both datasets (SEED and DEAP). The BoDF technique reduces the dimension of the extracted feature significantly, which in turn increases the accuracy of the algorithm. The results obtained using the DECS technique are omitted from the figures, as they only slightly improve algorithm accuracy. The primary purpose of employing DECS, as explained in the Introduction, is to reduce computational overhead.

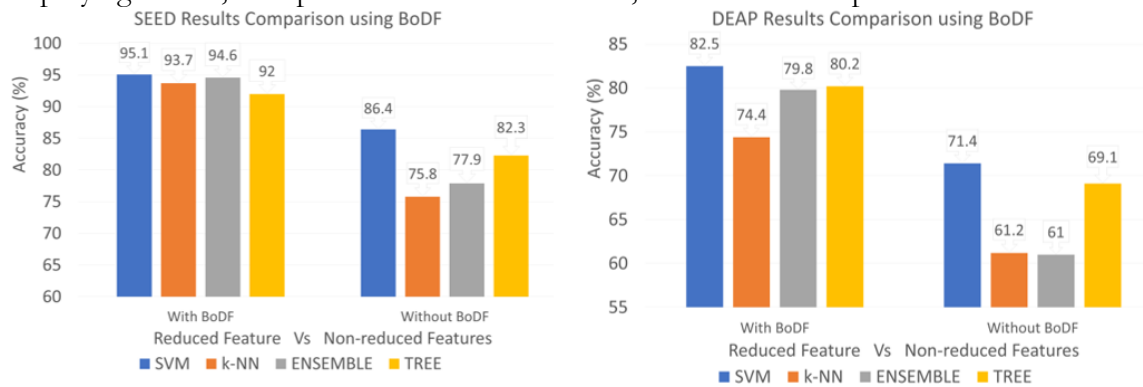


Figure 6. The improvement in accuracy by using the BoDF technique for SEED and DEAP datasets.

A comparison was made between the previous studies and these results to benchmark them. The classification techniques, along with the classifiers that were used to classify the feature vectors, are also mentioned in Table 2. The datasets on which the models were designed and the number of channels that were used are also highlighted. In Table 2, the feature vectors acquired from GoogLeNet were 41850 × 1000 lessened to a dimension of 225 × 24 vectors for the SEED dataset.

In our research work, we used the state-of-the-art DNNs, namely GoogLeNet, AlexNet, ResNet-50, ResNet-101, InceptionResNetV2, to extract the feature vectors. We also experimented with the number of clusters being made by varying the value of k. While selecting the value for the number of clusters 'k', we omitted the smaller ones because the sum of squared errors was more significant at these values and picked the larger ones, like k = 8,10,12,14, which had a low sum of squared errors.

K-NN Classifier:

For the SEAD dataset, the k-NN gave low classification results when combined with ResNet-101, with the highest only up to 92.8% with k = 8, as shown in Figure 7 (LEFT). Similarly, InceptionResNetV2 provided only up to 92.8% with the number of clusters k = 14, and the accuracy decreases with a decrease in the number of clusters.

The k-NN classifier can provide an accuracy of 93.7% using GoogLeNet, while the cluster value is at k=8, and provided relatively low accuracy as the cluster was increased. When combined with AlexNet, an accuracy of 94.6% is obtained when k=12, and the other values of k gave lower results. The k-NN classifier also provided the same results when integrated with ResNet-50 at k=14 and relatively low classification results at lower values of k.

In the case of the DEAP dataset, the highest accuracy achieved is 74.4% using Resnet-50 and AlexNet with k=12 and k=14. The accuracies obtained from different neural networks are shown in Figure 7 (RIGHT).

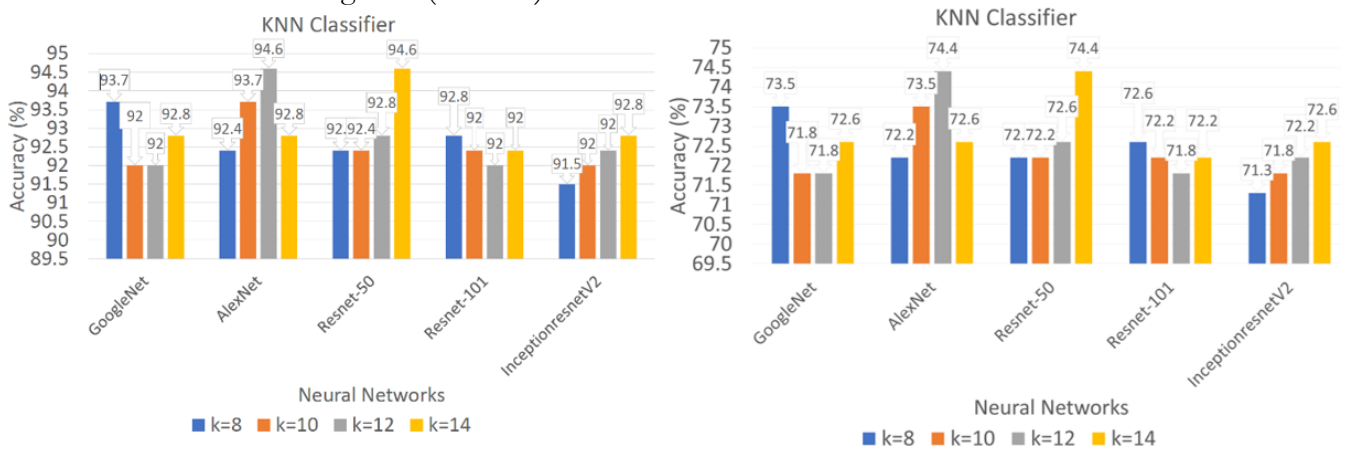


Figure 7. The results obtained from the k-NN Classifier on the SEED dataset (LEFT) and DEAP dataset (RIGHT).

Ensemble Classifier:

For the SEED dataset, the Ensemble classifier achieved an accuracy of 94.6% with GoogLeNet at a cluster value of k = 8, while higher numbers of clusters resulted in lower classification performance. AlexNet, ResNet-50, ResNet-101, InceptionresnetV2 gave an accuracy of 94.2% with k=8, 93.7% with k=12, 93.3% with k=12 respectively. The results are shown in Figure 8 (LEFT).

For the DEAP dataset, a maximum accuracy of 79.8% is achieved with GoogLeNet (k = 8). For a detailed comparison of accuracy across different neural networks, refer to Figure 8 (right).

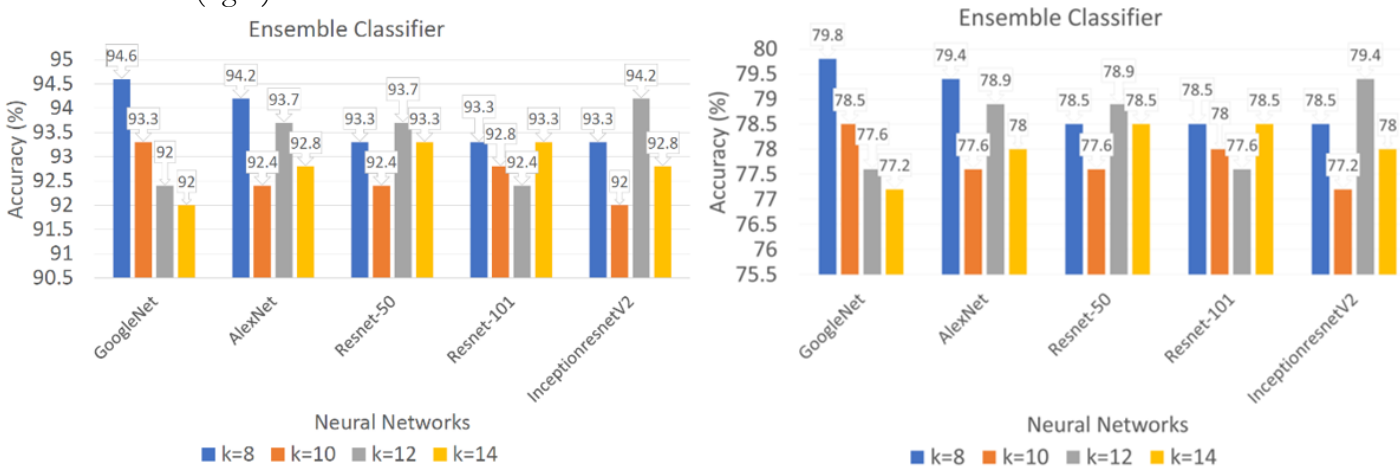


Figure 8. The results obtained from the Ensemble Classifier on the SEED dataset (LEFT) and DEAP dataset (RIGHT).

Tree Classifier:

For SEAD, the Tree classifier provided an accuracy highest up to 94.2% with the DNN, namely GoogleNet with clustering at k = 12, as shown in Figure 9 (LEFT), and InceptionResNetV2 with cluster value at k = 10. The other DNN can't provide high classification results even with changing the value of the number of clusters.

For DEAP, the tree classifier provides the accuracies in the range between 80.2% (GoogLeNet & InceptionsernetV2) and 76.2% (Resnet-50), see Figure 9 (RIGHT).

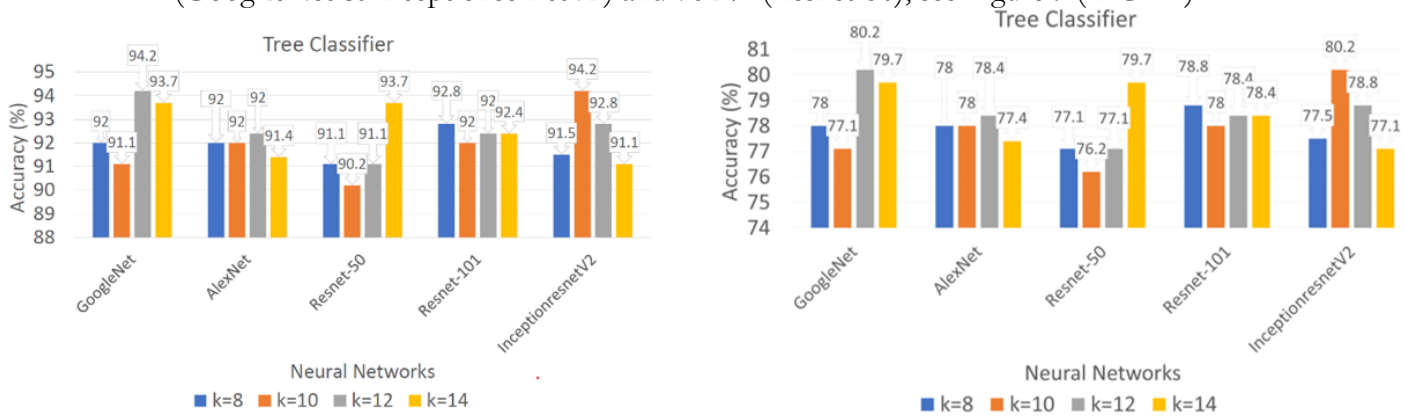


Figure 9. The results obtained from Tree Classifier on the SEED dataset (LEFT) and DEAP dataset (RIGHT).

SVM Classifier:

In the case of SEAD, the highest percentage was obtained with the SVM classifier, which gave an accuracy of 95.1% while applying GoogleNet for feature extraction with the value of k = 8. Meanwhile, none of the other DNNs achieved high-quality classification, even when the number of clusters k was varied, as shown in Figure 10 (left). An accuracy of 93.7% was obtained with AlexNet at the cluster value of k=14. The other DNNs like ResNet-50, ResNet-101, and InceptionResNetV2, could provide quality classification only up to 92.8% with k=8 and 12, 92.4% with k=12 and 14, 92.8% with k=8 and 12, respectively.

The optimal classification results were obtained from GoogleNet with the SVM classifier at a clustering value of k=8. The accuracy achieved in classifying emotions reached 95.1%, representing the highest performance obtained across all combinations of classifiers, cluster numbers, and DNNs.

In the case of DEAP, the maximum achieved accuracy is 82.5% using GoogLeNet (k=8). The accuracy obtained from the SVM classifier (using GoogLeNet) is the highest accuracy compared to other classifiers and neural networks, for a detailed comparison, see Figure 10 (RIGHT).

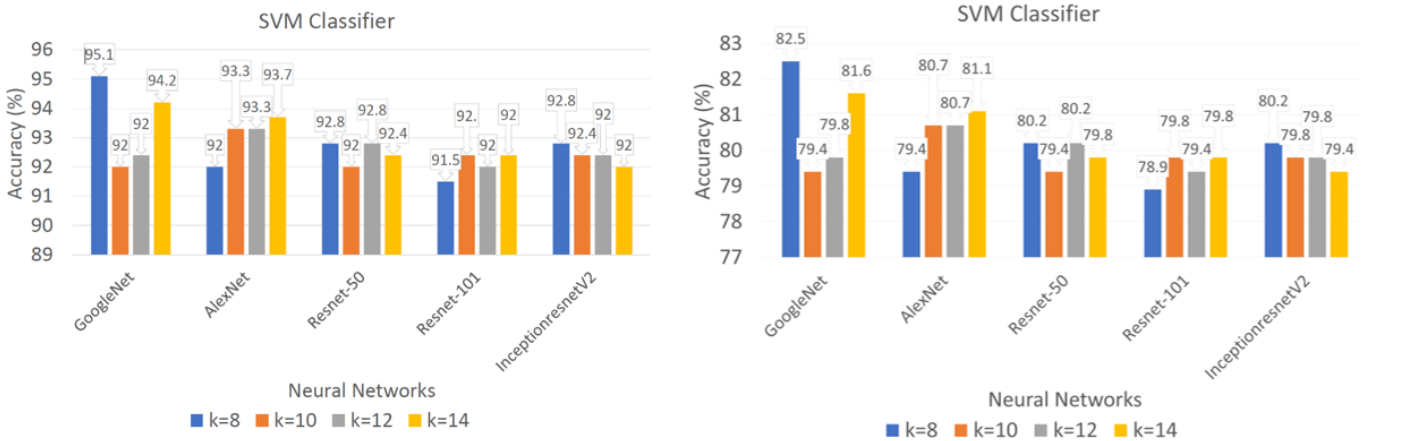


Figure 10. The results obtained from the SVM Classifier on the SEED dataset (LEFT) and DEAP dataset (RIGHT).

Table 1. Analysis of accuracy following channel reduction of various DNNs.

Neural Networks	SEAD, DEAP Channels	Classifiers	Kernal used	SEAD, DEAP Accuracies (%)
GoogleNet	26, 12	SVM	Cubic	95.1, 82.5
		kNN	Fine	93.7, 73.5
		Tree	Medium	94.2, 80.2
		Ensemble	Subspace KNN	94.6, 79.8
AlexNet	28, 13	SVM	Fine Gaussian	93.7, 81.1
		kNN	Weighted	94.6, 74.4
		Tree	Medium/Fine	92.4, 78.4
		Ensemble	Subspace KNN	94.2, 79.4
Resnet-50	40, 19	SVM	Fine Gaussian	92.8, 80.2
		kNN	Weighted	94.6, 74.4
		Tree	Medium/Fine	93.7, 79.7
		Ensemble	Subspace KNN	93.7, 78.9
Resnet-101	29, 14	SVM	Fine Gaussian	92.4, 79.8
		kNN	Weighted	92.8, 72.6
		Tree	Medium/Fine	92.8, 78.8
		Ensemble	Bagged Trees	93.3, 78.5
InceptionresnetV2	32, 16	SVM	Cubic	92.8, 80.2
		kNN	Weighted/Fine	92.8, 72.6
		Tree	Medium/Fine	94.2, 80.2
		Ensemble	Subspace KNN	94.2, 79.4

Table 2. Comparison of our proposed work with other studies.

Ref.	Technique	Dataset	Selected Channels	Classification model	Classification accuracy (%)
[4]	MFM	DEAP	18	CapsNet	68.2
[32]	MFCC	BEED	12	SVM	83.5
				Random Forest	72.07
		DEAP	6	Random Forest	72.07
[22]	MEMD	DEAP	12	ANN	75
				KNN	67
[28]	STRNN	SEED	62	CNN	89.5
[29]	RFE	SEED	18	SVM	90.4
		DEAP	12	SVM	60.5
[33]	DE	DEAP	32	PNN	79.3
[12]	BoDF	SEED	62	SVM	93.8
				KNN	91.4
		DEAP	32	SVM	77.4
				KNN	73.6
Our Work	DECS-BoDF	SEED	26	SVM	95.1
				KNN	94.6
				Tree	94.2
				Ensemble	94.6
		DEAP	12	SVM	82.5

				KNN	73.5
				Tree	78
				Ensemble	79.8
				SVM	95.1

Discussion:

Based on the presented results, a comparative analysis was conducted to evaluate the performance of the proposed DECS-BoDF method against recent state-of-the-art techniques in emotion recognition from EEG signals.

Our method demonstrates superior performance on the SEED dataset, achieving a peak accuracy of 95.1% with an SVM classifier. This outperforms all other cited studies, including the previous best of 93.8% by[12] using a Bag-of-DF (BoDF) model and 90.4% by[29] using Recursive Feature Elimination (RFE). Notably, our approach not only achieves a higher peak accuracy but also shows remarkable consistency across multiple classifiers (SVM, KNN, Tree, Ensemble), all yielding accuracies above 94%. This suggests that the DECS-BoDF features are highly robust and not dependent on a single classification algorithm.

The results on the more challenging DEAP dataset further validate the effectiveness of our method. Our model achieves a top accuracy of 82.5% with SVM, which is competitive with and, in most cases, superior to the existing literature. For instance, it substantially outperforms the 60.5% accuracy reported by[29] and the 77.4% by [12]. It is comparable to the 79.3% achieved by[33] using Differential Entropy (DE) and a Probabilistic Neural Network (PNN). The strong performance on DEAP is significant as it indicates that our feature extraction method (DECS-BoDF) generalizes well across different datasets, which is a key challenge in the field.

A critical advantage of our work is the achievement of high accuracy with a reduced number of EEG channels. On the DEAP dataset, we achieved 82.5% accuracy using only 12 channels, compared to [12] and [33], which used 32 channels to achieve 77.4% and 79.3%, respectively. This indicates that our channel selection strategy (likely implied by DECS) is highly effective at identifying the most informative brain regions for emotion recognition, which is a major step towards developing more practical and wearable BCI systems. Unlike many studies that report results for a single optimized classifier, our work demonstrates the robustness of the DECS-BoDF features across a diverse set of classifiers (SVM, KNN, Tree, Ensemble). The consistently high performance on the SEED dataset and competitive performance on DEAP regardless of the classifier used strongly suggest that the strength lies in the quality of the features themselves, rather than in a specific model's ability to fine-tune them.

Conclusion:

This work proposes a novel architecture that employs Differential Entropy-Based Channel Selection and Bag-of-Deep Feature technique to achieve appreciably higher classification accuracy for the SJTU SEED and DEAP datasets by reducing the computational cost of the model. The constituent techniques lead to a significant reduction in the number of feature vectors obtained initially via GoogLeNet, followed by SVM-based classification. Upon clustering the feature vectors into eight distinct groups, an accuracy of 95.1% is reported using the proposed method. Employing the proposed method also lowers the required number of EEG channels from 62 to 26, rendering the entire classification process less computationally intensive. The reduction in the number of features and, consequently, the number of channels paves the way for lowered computational cost and memory storage requirement for processing and storage of EEG signals with no substantial degradation in classification accuracy. The model can be extended in the future to allow for further reduction of required channels, paving the way for real-time emotion classification.

Acknowledgement. Acknowledgements are considered necessary.

Author's Contribution. Each Author contributed equally.
Conflict of interest. The authors declared no conflict of interest.
References:

[1] Y. Liu, R. Liang, S. Xu, and X. Guo, "Structural investigations of multi-reservoir Echo State Networks for EEG-based emotion classification," *Neurocomputing*, vol. 632, p. 129856, Jun. 2025, doi: 10.1016/J.NEUCOM.2025.129856.

[2] S. Abdumalikov, J. Kim, and Y. Yoon, "Performance Analysis and Improvement of Machine Learning with Various Feature Selection Methods for EEG-Based Emotion Classification," *Appl. Sci. 2024, Vol. 14, Page 10511*, vol. 14, no. 22, p. 10511, Nov. 2024, doi: 10.3390/APP142210511.

[3] R. Gajšek, V. Štruc, and F. Mihelič, "Multi-modal emotion recognition using canonical correlations and acoustic features," *Proc. - Int. Conf. Pattern Recognit.*, pp. 4133–4136, 2010, doi: 10.1109/ICPR.2010.1005.

[4] H. Chao, L. Dong, Y. Liu, and B. Lu, "Emotion Recognition from Multiband EEG Signals Using CapsNet," *Sensors 2019, Vol. 19, Page 2212*, vol. 19, no. 9, p. 2212, May 2019, doi: 10.3390/S19092212.

[5] W. Ma, Y. Zheng, T. Li, Z. Li, Y. Li, and L. Wang, "A comprehensive review of deep learning in EEG-based emotion recognition: classifications, trends, and practical implications," *PeerJ Comput. Sci.*, vol. 10, p. e2065, May 2024, doi: 10.7717/PEERJ-CS.2065.

[6] P. Tzirakis, J. Zhang, and B. W. Schuller, "End-to-end speech emotion recognition using deep neural networks," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2018-April, pp. 5089–5093, Sep. 2018, doi: 10.1109/ICASSP.2018.8462677.

[7] F. Aloise, P. Aricò, F. Schettini, S. Salinari, D. Mattia, and F. Cincotti, "Asynchronous gaze-independent event-related potential-based brain–computer interface," *Artif. Intell. Med.*, vol. 59, no. 2, pp. 61–69, Oct. 2013, doi: 10.1016/J.ARTMED.2013.07.006.

[8] P. Arico, G. Borghini, G. Di Flumeri, N. Sciaraffa, and F. Babiloni, "Passive BCI beyond the lab: current trends and future directions," *Physiol. Meas.*, vol. 39, no. 8, p. 08TR02, Aug. 2018, doi: 10.1088/1361-6579/AAD57E.

[9] G. Di Flumeri *et al.*, "EEG-based Approach-Withdrawal index for the pleasantness evaluation during taste experience in realistic settings," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 3228–3231, Sep. 2017, doi: 10.1109/EMBC.2017.8037544.

[10] P. Aricò *et al.*, "Human Factors and Neurophysiological Metrics in Air Traffic Control: A Critical Review," *IEEE Rev. Biomed. Eng.*, vol. 10, pp. 250–263, 2017, doi: 10.1109/RBME.2017.2694142.

[11] K. Markopoulos, C. Mavrokefalidis, K. Berberidis, and E. Daskalopoulou, "BCI-based approaches for real-time applications," *ACM Int. Conf. Proceeding Ser.*, Nov. 2016, doi: 10.1145/3003733.3003785.

[12] M. A. Asghar *et al.*, "EEG-Based Multi-Modal Emotion Recognition using Bag of Deep Features: An Optimal Feature Selection Approach," *Sensors 2019, Vol. 19, Page 5218*, vol. 19, no. 23, p. 5218, Nov. 2019, doi: 10.3390/S19235218.

[13] N. K. Al-Qazzaz, M. K. Sabir, S. Ali, S. A. Ahmad, and K. Grammer, "Effective EEG Channels for Emotion Identification over the Brain Regions using Differential Evolution Algorithm," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 4703–4706, Jul. 2019, doi: 10.1109/EMBC.2019.8856854.

[14] M. A. Ozdemir, M. Degirmenci, O. Guren, and A. Akan, "EEG based emotional state estimation using 2-D deep learning technique," *TIPTEKNO 2019 - Tip Teknol. Kongresi*, Oct. 2019, doi: 10.1109/TIPTEKNO.2019.8895158.

[15] K. Y. Wang, Y. L. Ho, Y. De Huang, and W. C. Fang, "Design of Intelligent EEG System for Human Emotion Recognition with Convolutional Neural Network," *Proc. 2019 IEEE Int. Conf. Artif. Intell. Circuits Syst. AICAS 2019*, pp. 142–145, Mar. 2019, doi: 10.1109/AICAS.2019.8771581.

[16] K. Guo, H. Yu, R. Chai, H. Nguyen, and S. W. Su, "A Hybrid Physiological Approach of Emotional Reaction Detection Using Combined FCM and SVM Classifier," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2019, pp. 7088–7091, Jul. 2019, doi: 10.1109/EMBC.2019.8857698.

[17] P. Santhiya and S. Chitrakala, "A Survey on Emotion Recognition from EEG Signals: Approaches, Techniques Challenges," *Proc. - Int. Conf. Vis. Towar. Emerg. Trends Commun. Networking, ViTECoN 2019*, Mar. 2019, doi: 10.1109/VITECON.2019.8899671.

[18] H. Ullah, M. Uzair, A. Mahmood, M. Ullah, S. D. Khan, and F. A. Cheikh, "Internal Emotion Classification Using EEG Signal with Sparse Discriminative Ensemble," *IEEE Access*, vol. 7, pp. 40144–40153, 2019, doi: 10.1109/ACCESS.2019.2904400.

[19] R. Yan, N. Lu, Y. Yan, X. Niu, and J. Wu, "A Fine-grained Hemispheric Asymmetry Network for accurate and interpretable EEG-based emotion classification," *Neural Networks*, vol. 184, p. 107127, Apr. 2025, doi: 10.1016/J.NEUNET.2025.107127.

[20] C. G. Coogan and B. He, "Brain-Computer Interface Control in a Virtual Reality Environment and Applications for the Internet of Things," *IEEE Access*, vol. 6, pp. 10840–10849, Feb. 2018, doi: 10.1109/ACCESS.2018.2809453.

[21] L. Song and J. Epps, "Classifying EEG for Brain-Computer Interface: Learning Optimal Filters for Dynamical System Features," *Comput. Intell. Neurosci.*, vol. 2007, no. 1, p. 057180, Jan. 2007, doi: 10.1155/2007/57180.

[22] A. Mert and A. Akan, "Emotion recognition from EEG signals by using multivariate empirical mode decomposition," *Pattern Anal. Appl.*, vol. 21, no. 1, pp. 81–89, Feb. 2018, doi: 10.1007/S10044-016-0567-6/METRICS.

[23] C. N. Anagnostopoulos, T. Iliou, and I. Giannoukos, "Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011," *Artif. Intell. Rev.*, vol. 43, no. 2, pp. 155–177, Feb. 2015, doi: 10.1007/S10462-012-9368-5/METRICS.

[24] M. T. Sadiq *et al.*, "Motor imagery EEG signals classification based on mode amplitude and frequency components using empirical wavelet transform," *IEEE Access*, vol. 7, pp. 127678–127692, 2019, doi: 10.1109/ACCESS.2019.2939623.

[25] J. Kevric and A. Subasi, "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system," *Biomed. Signal Process. Control*, vol. 31, pp. 398–406, Jan. 2017, doi: 10.1016/J.BSPC.2016.09.007.

[26] M. A. Asghar, Fawad, M. J. Khan, Y. Amin, and A. Akram, "EEG-based Emotion Recognition for Multi Channel Fast Empirical Mode Decomposition using VGG-16," *2020 Int. Conf. Eng. Emerg. Technol. ICEET 2020*, Feb. 2020, doi: 10.1109/ICEET48479.2020.9048217.

[27] W. L. Zheng and B. L. Lu, "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks," *IEEE Trans. Auton. Ment. Dev.*, vol. 7, no. 3, pp. 162–175, Sep. 2015, doi: 10.1109/TAMD.2015.2431497.

[28] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial-Temporal Recurrent Neural Network for Emotion Recognition," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 939–947, Mar. 2019, doi: 10.1109/TCYB.2017.2788081.

[29] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring EEG features in cross-subject emotion recognition," *Front. Neurosci.*, vol. 12, no. MAR, p. 294333, Mar. 2018, doi: 10.3389/FNINS.2018.00162/BIBTEX.

[30] R. W. Homan, J. Herman, and P. Purdy, "Cerebral location of international 10–20

system electrode placement," *Electroencephalogr. Clin. Neurophysiol.*, vol. 66, no. 4, pp. 376–382, Apr. 1987, doi: 10.1016/0013-4694(87)90206-9.

[31] A. K. Vuong, J. Zhang, R. L. Gibson, and W. W. Sager, "Application of the two-dimensional continuous wavelet transforms to imaging of the Shatsky Rise plateau using marine seismic data," *Spec. Pap. Geol. Soc. Am.*, vol. 511, pp. 127–146, 2015, doi: 10.1130/2015.2511(07).

[32] V. Gupta, M. D. Chopda, and R. B. Pachori, "Cross-Subject Emotion Recognition Using Flexible Analytic Wavelet Transform from EEG Signals," *IEEE Sens. J.*, vol. 19, no. 6, pp. 2266–2274, Mar. 2019, doi: 10.1109/JSEN.2018.2883497.

[33] B. Nakisa, M. N. Rastgoo, D. Tjondronegoro, and V. Chandran, "Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors," *Expert Syst. Appl.*, vol. 93, pp. 143–155, Mar. 2018, doi: 10.1016/J.ESWA.2017.09.062.

[34] C. Szegedy *et al.*, "Going deeper with convolutions," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June-2015, pp. 1–9, Oct. 2015, doi: 10.1109/CVPR.2015.7298594.

[35] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 2818–2826, Dec. 2016, doi: 10.1109/CVPR.2016.308.

[36] Y. Qu, R. Li, A. Deng, C. Shang, and Q. Shen, "Non-unique decision differential entropy-based feature selection," *Neurocomputing*, vol. 393, pp. 187–193, Jun. 2020, doi: 10.1016/J.NEUCOM.2018.10.112.

[37] S. O'Hara and B. A. Draper, "Introduction to the Bag of Features Paradigm for Image Classification and Retrieval," Jan. 2011, Accessed: Oct. 06, 2025. [Online]. Available: https://arxiv.org/pdf/1101.3354

[38] L. Elazary and L. Itti, "Interesting objects are visually salient," *J. Vis.*, vol. 8, no. 3, pp. 3–3, Mar. 2008, doi: 10.1167/8.3.3.

[39] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," 1996, Accessed: Oct. 06, 2025. [Online]. Available: www.aaai.org

Copyright © by the authors and 50Sea. This work is licensed under the Creative Commons Attribution 4.0 International License.