Dean&Francis
ISSN 2959-6157

EEG Feature Extraction and Classification of SSVEP

Zihan Wang[1, *], **Wenxuan Zou**[2] **and Yunxin Zhang**[3]

[1]Electronic Information Engineering, Hubei University of Technology, Wuhan, China
[2]Electronic Science and Technology, Nanjing University of Information and Technology, Nanjing, China
[3]Biomedical Engineering, City University of Hong Kong, Hong Kong, China

*Corresponding author: Jichuansai@ldy.edu.rs

Abstract:
As an innovative communication method, brain-computer interface (BCI) can directly convert human brain activity into control signals, which is of great significance in improving the quality of life of people with disabilities. In this paper, the application of steady-state visual evoked potential (SSVEP) in BCI is discussed, and the feature extraction and classification methods of EEG signals are studied. Feature extraction of EEG signals was performed by preprocessing them using the EEGLAB toolbox and classification using support vector machine (SVM) to identify different patterns of EEG activity. The experimental results show that this feature extraction and classification method significantly improves the performance of BCI system. Future research can further optimize the feature extraction algorithm and improve the visual stimulation paradigm to improve the recognition accuracy and practicality of the system. Additionally, integrating advanced machine learning techniques such as deep learning and transfer learning could potentially enhance the system's ability to adapt to individual users and generalize across different tasks, thereby increasing the robustness and versatility of the BCI system in real-world applications.

Keywords: Steady-state visual evoked potentials; Feature extraction; Visual neural networks; Brain-computer interface

1. Introduction

In recent years, research on Brain Computer Interface (BCI) has rapidly developed, providing innovative methods for direct communication between the human brain and external devices. Among the various modes of BCI, electroencephalography (EEG) stands out as a particularly promising approach, owing to its non-invasive nature, high temporal resolution, and relatively low cost [1]. EEG records the electrical activity of the brain, making it a powerful tool for analyzing the brain's response to different stimuli, including visual stimuli.

Visual evoked potential (VEP) is a specific EEG response triggered by visual stimuli and has been widely used in clinical and research fields. VEP provides

ISSN 2959-6157

important insights into visual cortex function and overall neural processing of visual information [2]. By analyzing EEG data of various visual stimuli in the brain, researchers can decode potential neural mechanisms and develop robust classification systems for BCI [3]. The core objective of this study is to explore EEG feature extraction and classification related to visual neural activity. Specifically, we focus on developing effective classification methods to distinguish between different visual stimuli. The achievement of this goal will not only improve the performance of BCI systems, but also provide a new perspective for understanding the neural mechanisms of visual information processing.

2. EEG Feature Extraction and Classification of SSVEP

2.1 Research Methodology

This study utilizes a publicly available EEG dataset as the data source, aiming to conduct an in-depth analysis of brain activity patterns to enhance our understanding of neural activities. Our methodology encompasses several key steps, including data preprocessing, feature extraction, and classification model application. In the data preprocessing phase, we employ the MATLAB EEGLAB toolbox, starting with electrode localization to ensure the spatial accuracy of the signals. Subsequently, a bandpass filter, typically set in the range of 1-40 Hz, is applied to remove low-frequency drift and high-frequency noise, thereby ensuring the quality of the signals. Following this, Independent Component Analysis (ICA) is performed to separate independent components from the mixed signals. ICA operates on the assumption that observed signals are linear mixtures of independent source signals, using statistical methods to identify these sources, effectively removing artifacts such as eye movements and muscle activity, resulting in cleaner EEG signals [4]. During the feature extraction phase, various key features are extracted from the preprocessed EEG signals, primarily including the mean and standard deviation in the time domain. These features reflect the fundamental statistical properties of the signals and are organized into a feature matrix that combines the features of each trial, providing a solid foundation for subsequent classification. Our current approach focuses on time-domain features, offering simplicity and computational efficiency. However, considering the complexity of EEG signals, future work may incorporate frequency domain and time-frequency features to further enhance model performance. In terms of classification models, the study applies Support Vector Machine (SVM) for feature classification [5]. SVM is a powerful supervised learning algorithm capable of effectively handling high-dimensional data and finding the optimal hyperplane to separate different classes. To comprehensively evaluate the classifier's perform, a 10-fold cross-validation method is employed, dividing the dataset into ten folds to ensure that each sample has the opportunity to serve as a test set, thereby improving the model's generalization ability. Furthermore, the study visualizes the classification results through a confusion matrix, providing an intuitive display of the classifier's performance across various categories. The confusion matrix not only presents the predicted results for each category but also reveals the model's confusion between different classes, highlighting which brain activity patterns are more prone to misclassification and providing direction for further improvements. This series of methods collectively provides effective technical support for the analysis of EEG signals, advancing our understanding of brain activity patterns. Specifically, it enables more accurate identification and classification of different brain activity states, laying the foundation for applications in brain-computer interfaces, neurological disease diagnosis, and other related fields. Future research will continue to explore more complex features and models to further enhance classification performance and broaden the scope of applications.

2.2 Experimental Principles

The experimental design consists of four fundamental steps: data preprocessing, feature extraction, classification, and results visualization. We utilize a publicly available EEG dataset from the Brain-Computer Interface Laboratory at Tsinghua University as our primary data source. In the data preprocessing step, we employ the MATLAB EEGLAB toolbox to ensure the quality and spatial accuracy of the EEG signals, which includes electrode localization and the application of a bandpass filter (typically 1-40 Hz) to remove noise and artifacts. The feature extraction phase focuses on deriving key characteristics from the preprocessed EEG signals, primarily including statistical features such as the mean and standard deviation in the time domain. These features are organized into a feature matrix, providing the foundational data for subsequent classification. For the classification step, we apply Support Vector Machine (SVM) to identify different brain activity patterns, leveraging its ability to handle high-dimensional data effectively. To evaluate the model's performance comprehensively, we implement a 10-fold cross-validation method [6]. Finally, in the results visualization step, we utilize a confusion matrix to provide an intuitive display of the classifier's performance across various categories, offering insights into the model's strengths and areas for potential improvement. This systematic approach not only advances our understanding of neural activities but also lays the groundwork for applications in brain-computer interfaces and neurological research. Regarding the data type and basic data, the EEG signal data used in this experiment is sourced from the Brain-Computer Interface Laboratory at Tsinghua University. EEG signals are typically stored in matrix form, with dimensions represented as [number of channels, number of time points, and number of trials]. In this experiment, the data dimensions are [64, 1500, 40, 6], indicating 64 channels, 1500 time points, 40 trials, and 6 conditions. The specific basic data includes 64 channels (i.e., 64 electrodes), 1500 time points (sampling points for each trial), 40 trials (the number of measurements conducted in the experiment), and 6 conditions (the six different conditions involved in the experiment). Through these steps and data, the experiment aims to conduct an in-depth analysis of EEG signals and identify brain activity patterns.

2.3 Experimental Procedures

The experimental design for analyzing EEG signals encompasses a systematic workflow that ensures the integrity and accuracy of the data throughout various stages. The process initiates with data acquisition, where we retrieve EEG signal data and electrode location files from the database and efficiently integrate them into the EEGLAB environment for subsequent analysis. This crucial step lays the foundation for the entire analytical workflow, ensuring the integrity and accessibility of the raw data. Following this, the signal preprocessing phase applies essential techniques such as bandpass filtering and Independent Component Analysis (ICA) to enhance signal quality by removing artifacts and noise. Next, the feature extraction stage focuses on deriving key statistical features from the preprocessed signals, which serve as the foundation for subsequent classification. The label preparation step ensures that each trial is accurately labeled according to the experimental design, facilitating effective model training. The classification phase employs Support Vector Machine (SVM) to identify distinct brain activity patterns, while the final results visualization stage presents the classification outcomes through confusion matrices and scatter plots, providing insights into the model's performance and the underlying EEG data characteristics. Each of these steps is crucial for achieving a comprehensive understanding of brain activity patterns through EEG analysis, as shown in Fig. 1.

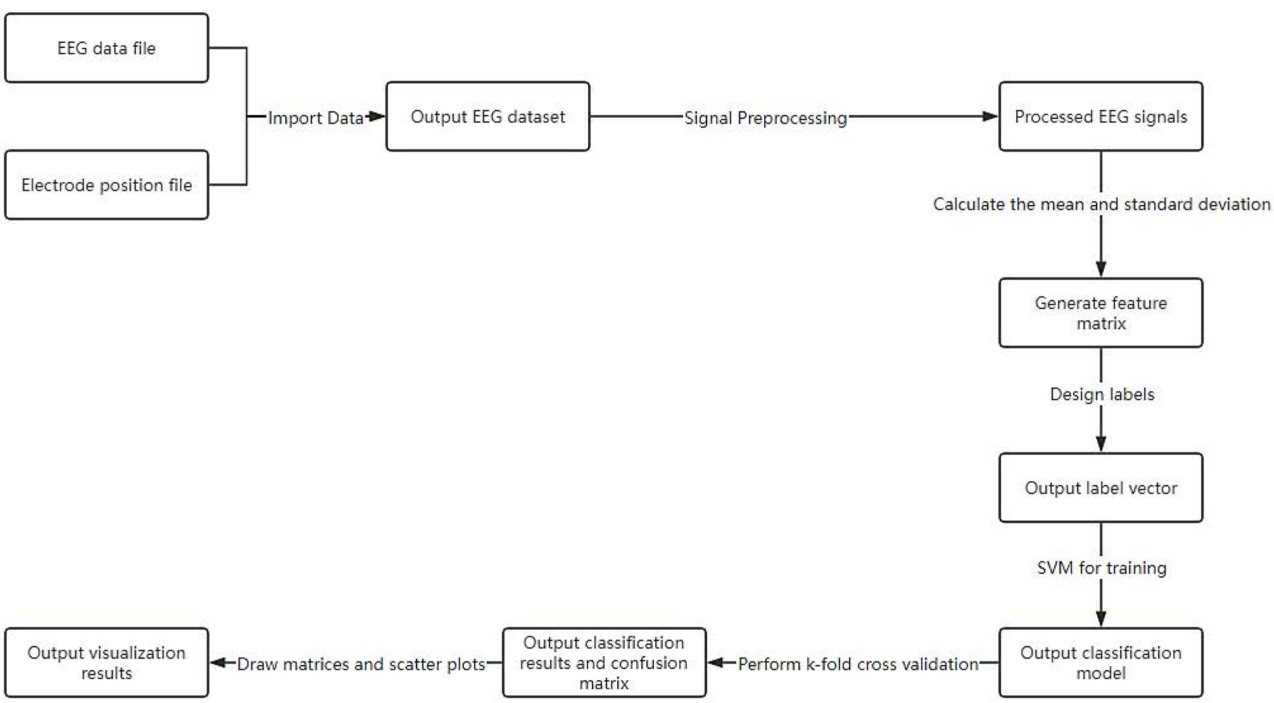

Fig. 1 Experimental Procedures (Photo/Picture credit: Original).

2.3.1 Data Loading:

The experiment begins by setting the file path for the data, ensuring proper access to the EEG signal data and electrode location files stored on the local computer. The EEG data file (e.g., S1.mat) containing the recorded signals are loaded using MATLAB's load function. This step

ISSN 2959-6157

imports the experimental EEG data into the workspace, making it readily available for subsequent preprocessing and analysis within the EEGLAB environment. Subsequently, the 'pop_importdata' function from the EEGLAB toolbox is employed to import the data into the EEGLAB environment, specifying the data format as MATLAB, setting the sampling rate to 256 Hz, the number of channels to 64, and the number of time points to 1500, thereby ensuring that the data structure meets the requirements of EEGLAB. Additionally, the electrode location file (e.g., '64-channels. loc') is loaded to ensure the spatial accuracy of the EEG signals. Finally, the 'eeg_checkset' function is used to verify the integrity and consistency of the dataset, ensuring that the foundational data for subsequent processing is reliable, as shown in Fig. 2.

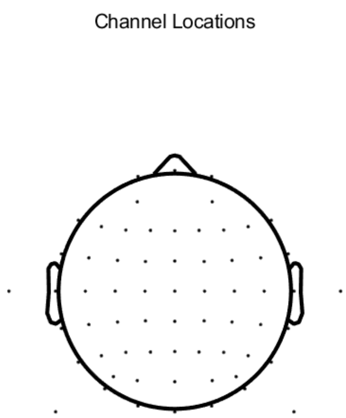

Fig. 2 Electrode position (Photo/Picture credit: Original).

2.3.2 Signal Preprocessing:

In the signal preprocessing phase, a bandpass filter is first applied, typically set in the range of 1-40 Hz. This filter removes low-frequency drift and high-frequency noise, ensuring the quality of the EEG signals. The low-frequency components often include artifacts caused by heartbeat and respiration, while high-frequency noise may result from electrical interference and other environmental factors. By eliminating these unwanted elements, we enhance the signal-to-noise ratio and focus on the brain activity patterns of interest for our analysis. Thereby ensuring the quality of the EEG signals. Next, Independent Component Analysis (ICA) is performed to separate independent components from the mixed signals, effectively removing artifacts and noise (such as eye movements and muscle activity), resulting in cleaner EEG signals. The ICA is executed using the 'pop_runica' function, and the ICA components are examined with the 'pop_selectcomps' function, allowing the researcher to manually identify and remove components associated with eye and muscle activity. This process typically requires the researcher to assess the waveforms and spectra of the components to ensure that the removed components are indeed artifacts. Finally, the processed dataset is saved for further analysis using the 'pop_saveset' function, storing the processed EEG data in '.set' file format.

2.3.3 Feature Extraction:

During the feature extraction phase, key features are extracted from each trial, primarily including the mean and standard deviation in the time domain. These statistical measures provide fundamental insights into the signal characteristics. However, EEG analysis often benefits from a multi-domain approach. Frequency domain features, such as power spectral density (PSD), offer valuable information about the signal's energy distribution across different frequency bands. Additionally, time-frequency features like wavelet coefficients capture both temporal and spectral properties of the EEG signals, providing a more comprehensive representation of brain activity patterns. While our current study focuses on time domain features for their computational efficiency, future work may incorporate these additional feature types to enhance the model's discriminative power and capture more nuanced aspects of neural dynamics.

These features reflect the fundamental statistical properties of the signals and are organized into a feature matrix that combines the features of each trial. Initially, a feature matrix is initialized, with the number of rows corresponding to the number of trials and the number of columns corresponding to the number of extracted features (mean and standard deviation for each channel). By iterating through each channel and trial, the mean and standard deviation for each trial are calculated and stored in the appropriate positions within the feature matrix. For each trial, the mean and standard deviation for each channel are computed and stored. This process results in a comprehensive feature matrix, providing a solid foundation for subsequent classification. The matrix encapsulates essential statistical properties of the EEG signals across all channels and trials, effectively reducing the high-dimensional raw data while retaining critical information about brain activity patterns [7].

2.3.4 Label Preparation:

Labels are generated based on the experimental design, reflecting the specific conditions or tasks associated with each trial. For example, in a study with 6 distinct experimental conditions and 40 trials per condition, labels 1 to 6 would be repeated 40 times each. This process creates a label vector that precisely corresponds to the number of rows in the feature matrix, ensuring a one-to-one mapping between features and their respective classifications. This careful alignment of features and labels is crucial for training the classification model accurately and evaluating its performance in distinguishing between different brain activity patterns. Assuming there are 6 conditions in the experiment, each corresponding to 40 trials, the appropriate labels are generated and converted into a column vector. The 'repmat' function is used to repeat the labels, ensuring that each trial has a corresponding label, thereby providing the necessary supervisory information for the classification model. Specifically, the label generation process involves repeating each condition's label (e.g., 1 to 6) 40 times to form a complete label vector, ensuring consistency with the number of rows in the feature matrix.

2.3.5 Classification:

Support Vector Machine (SVM) is employed for feature classification, training the model and making predictions. SVM is chosen for its effectiveness in high-dimensional spaces and its ability to handle non-linear decision boundaries through kernel tricks, making it particularly suitable for EEG classification tasks. This algorithm excels at finding the optimal hyperplane that maximizes the margin between different classes of brain activity patterns [8].

The 'fitcecoc' function is used to train a multi-class SVM model to identify different brain activity patterns. To evaluate the model's performance, k-fold cross-validation (typically k=10) is adopted, calculating classification accuracy and classification loss in each fold. This cross-validation approach ensures that each sample has the opportunity to serve as a test set, thereby enhancing the model's generalization ability, as shown in Fig. 3.

After training is complete, the trained model is used to predict the features, and a confusion matrix is generated to visualize the model's performance across different categories. The confusion matrix not only provides the predicted results for each category but also reveals the model's confusion between different classes, aiding in the analysis of the model's classification capability, as shown in Fig. 4.

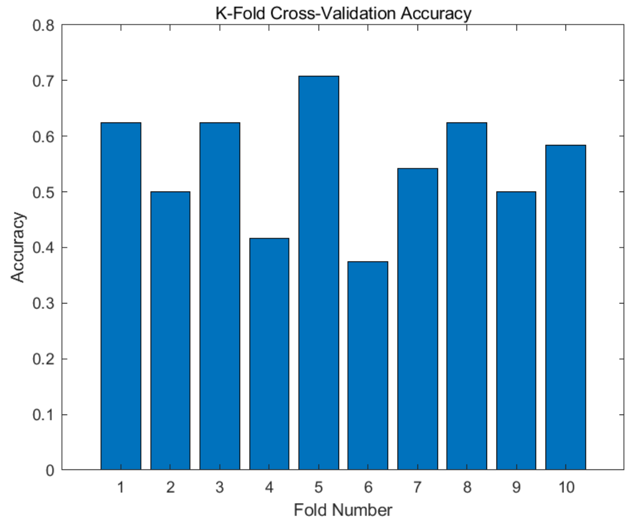

Fig. 3 K-Fold cross-validation accuracy (Photo/Picture credit: Original).

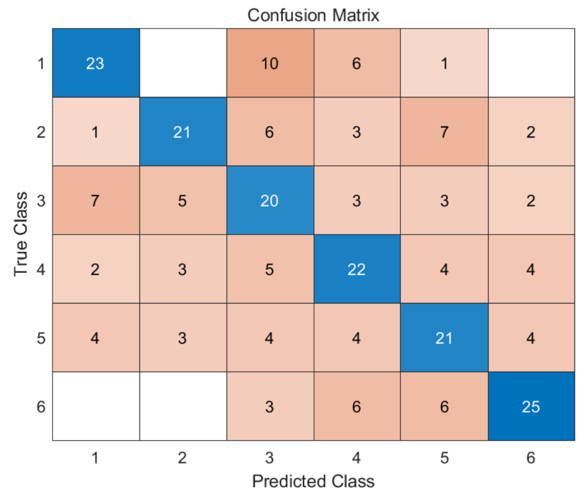

Fig. 4 Confusion matrix (Photo/Picture credit: Original).

2.3.6 Results Visualization:

In the results visualization phase, a confusion matrix is plotted to intuitively display the classification results, assisting in the analysis of the model's performance across various categories. Additionally, a feature scatter plot is created to illustrate the distribution of different category samples in the feature space, helping to understand the model's classification ability and the discriminative power of the features [9]. The 'gscatter' function is utilized to color the scatter plot based on the prediction results, providing a visual representation of the distribution of different category samples. Through these visualization techniques, researchers can better assess the effectiveness of the model and provide insights for future improvements, as shown in Fig. 5.

ISSN 2959-6157

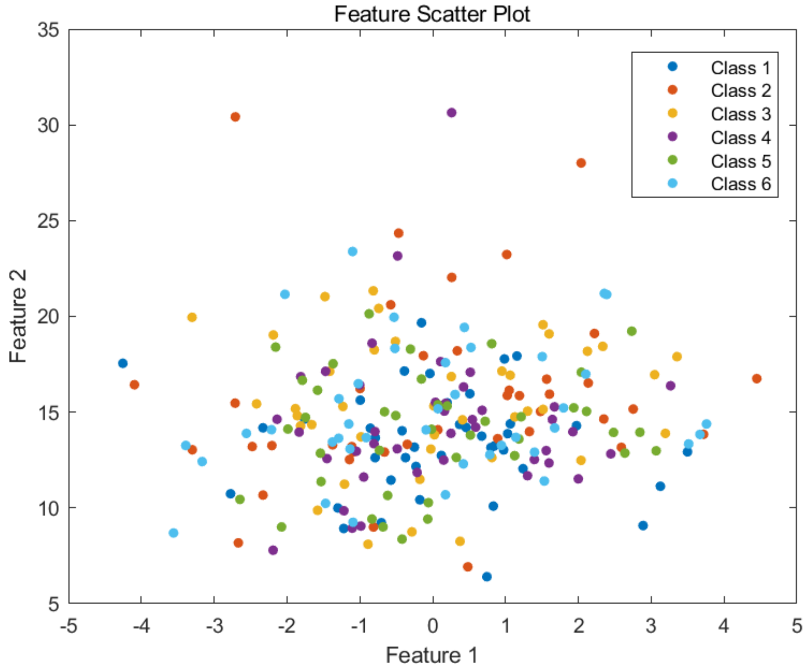

Fig. 5 Feature scatter plot (Photo/Picture credit: Original).

3. Conclusion

By loading and preprocessing EEG data, independent component analysis (ICA) was applied to remove artifacts, and features in time and frequency domains were extracted to construct an effective feature matrix. Then, support vector machine (SVM) was used to classify the extracted features, and the performance of the model was evaluated by cross-validation, and the classification accuracy was calculated. The confusion matrix visualizes the prediction effect of the model on various categories, while the feature scatter plot shows the distribution of different categories of samples in the feature space, which improves the understanding and analysis ability of the electrical activity patterns of the brain as a whole.

Future work could focus on increasing dataset diversity through techniques like data augmentation or cross-subject learning, which could significantly improve the model's generalization ability across different individuals and experimental conditions. In addition, we also consider introducing more multi-frequency and nonlinear features to improve the model performance. At the same time, advanced algorithms such as deep learning are explored to better capture complex patterns in EEG signals. Considering the computational efficiency and sensitivity to noise of support vector machines (SVM) when processing large-scale data, it is possible to combine real-time data processing and online classification techniques in the future to promote the practical application of EEG signal analysis in clinical and brain-computer interface applications.

4. Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Liu J. Research on brain-computer interface system based on steady-state visual evoked potential [D]. Hangzhou: Hangzhou Dianzi University, 2016.

[2] Lin C. Research and application of SSVEP feature extraction and classification algorithm [D]. Guangzhou: South China University of Technology, 2021.

[3] Tang J, Zhou Z. A shared-control based BCI system: For a robotic arm control. Proceedings of 2017 First International Conference on Electronics Instrumentation & Information Systems (EIIS), Harbin, China, 2017: 1-5.

[4] Karasawa N, Mitsutake A, Takano H, et al. Identification of slow relaxation modes in a protein trimer via positive definite relaxation mode analysis. The Journal of Chemical Physics, 2019, (8): 084-113.

[5] Korkmaz S, Zararsiz G, Goksuluk D, et al. Drug/nondrug classification using support vector machines with various feature selection strategies. Computer Methods and Programs in Biomedicine, 2014, (2): 51-60.

[6] Wang Y, Ni XS, et al. A Xgboost risk model via feature selection and Bayesian hyper-parameter optimization. International Journal of Database Management Systems, 2019, (01): 01-17.
[7] Ardestani A, Shen W, Darvas F, Toga AW, Fuster JM, et al. Modulation of frontoparietal neurovascular dynamics in working memory. Journal of Cognitive Neuroscience, 2016, (3): 379-401.
[8] Peng X, Xu D, et al. Twin support vector hypersphere (TSVH) classifier for pattern recognition. Neural Computing and Applications, 2020, (5): 1207-1220.
[9] Li X, Wei Y, Zhou Y, Hong B, et al. Subcortical brain segmentation based on a novel discriminative dictionary learning method and sparse coding. IEEE Access, 2019: 149785-149796.