A Novel Approach to Motor Imagery Classification via Mini-Epoch Generation

Litao Qiao
l1qiao@ucsd.edu

Gaopo Huang
ghuang@ucsd.edu

Darin Tsui
dtsui@ucsd.edu

Siddhant Jadhav
sijadhav@ucsd.edu

Abstract

Classification of motor imagery tasks based on EEG signals has a wide range of applications in clinical science and Brain-Computer Interfaces (BCIs). A number of machine learning and deep learning algorithms fused together with signal processing techniques have been used in recent years. However, these attempts fail to retain the temporal information in the EEG signals that is critical for motor imagery. In this paper, we propose two algorithms, mini-epoch learning and mini-epoch ensemble learning, that aim to tackle this limitation. We find that implementing our mini-epoch learning algorithm with SVM proved to be the most accurate in identifying different motor imagery tasks, with an accuracy of 63.16%. Implementing mini-epoch ensemble learning with XGB recorded the highest accuracy of 61.04%. The mini-epoch learning algorithm offers a better classification performance compared to other state-of-the-art methods.

1. Introduction

1.1. Background

Brain-Computer Interfaces (BCIs) allow individuals to interact with their surroundings without the usage of peripheral nerves and muscles [11]. BCIs accomplish this by capturing and interpreting brain signals. Subsequently, the extracted information is transformed into commands that can be communicated to an output device, enabling the execution of specific actions [17]. One key modality for BCIs is electroencephalography (EEG), which measures the electrical activity of the brain through the scalp. In this instance, brain signals are collected through noninvasive procedures [10]. Recent studies have explored the possibility of using BCI in a range of clinical applications. In particular, motor imagery-based BCIs, which involve the mental simulation of a movement, offer ways for individuals with neurological motor disorders to interact with the world [15]. For example, studies have shown that rehabilitation involving motor imagery and EEG-based BCIs can help post-stroke patients with the functional recovery of limbs [14].

The application of machine learning in motor imagery classification is an exciting and rapidly developing research area. However, it remains challenging to achieve high accuracy due to the presence of noise and other artifacts in the signals [1]. Moreover, the temporal nature of motor imagery introduces difficulties in processing EEG data to capture adequate signals [7]. In [18], three conventional classification methods (power + SVM, common spatial patterns (CSP) + SVM, and Autoregression (AR) + SVM) are attempted for binary class classification. In this paper, we propose a novel approach for the classification of motor imagery movements. We do this by performing feature extraction over small time windows. From here, these small windows constitute a feature set inputted to our proposed architecture to make the final classification. Our approach aims to address the challenges associated with motor imagery classification in BCI applications.

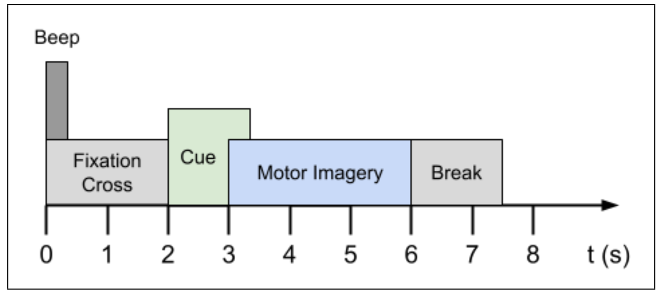

Figure 1. Timing sequence of motor imagery experiments.

1.2. Description of dataset

To test our algorithm implementation, we use the publicly available BCI Competition IV 2a data set [3]. The dataset consists of EEG recordings from 9 subjects, with each subject providing 2 sessions of data. Each session contains 6 runs separated by breaks, with each run containing 48 trials. In total, each session contains 288 trials.

Each session of data includes 22 EEG recordings and 3 electrooculography (EOG) recordings. In these sessions, subjects were asked to perform four different motor imagery tasks: imagining the movement of the left hand, right hand, feet, and tongue. The timing sequence of the experiment is shown in Figure 1.

2. Related work

2.1. Motor imagery classification in BCI application

Several approaches have been developed for improving the classification accuracy of EEG signals in BCI systems. In this section, we will examine various classification algorithms employed towards the BCI Competition IV 2a data set [3].

Echtioui et al. proposed fusing convolutional neural networks (CNN) for multi-class motor imagery EEG signals classification [6]. Their implementation fused two CNNs with the Long Short-Term Memory (LSTM) layers. The proposed architecture achieved an accuracy of 61.68% on the BCI Competition IV data set 2a, which outperforms other state-of-the-art methods.

In a separate study, Luo et al. proposed an ensemble support vector learning (ESVL) method for motor imagery EEG classification [9]. The proposed method used multiple support vector machines (SVMs) with different kernels and parameters, which were combined using a weighted voting scheme to improve classification accuracy. In their implementation, this EVSL method achieved a kappa value of 0.6.

Other approaches involving signal processing techniques have also seen success. Specifically, Ang et al., the winner in BCI Competition IV data set 2a, proposed using the filter bank common spatial pattern (FBCSP) algorithm to extract discriminative features [2]. The features are then fed into the Naive Bayes Parzen Window classifier. The proposed FBCSP algorithm achieved a kappa value of 0.569.

Overall, these studies demonstrate the effectiveness of using machine learning techniques on motor imagery classification. However, the real-time neural processes underlying motor imagery are still not fully understood. Research has shown that the neurological response to motor imagery is not only influenced by the elapsed time since the stimulus was cued, but also by the task's difficulty level. Specifically, the duration of imagining a movement is directly proportional to the time required to physically execute that movement [12]. As a result, it is possible that taking into account individual, brief time steps could improve classification accuracy.

2.2. Multi-class classifiers

The BCI Competition IV 2a data set poses a multi-class classification task where one out of four possible labels should be outputted given an input time signal. Many multi-class classification algorithms have been developed and applied to similar problems. In our approach, we either use one multi-class classifier or an ensemble of them to predict the label for each trial. We implement them with several classifiers commonly used in other machine learning papers, including SVM, Deep Learning (DL) methods, and XGBoost. SVM classifies instances with a hyperplane parameterized by the support vectors based on the idea of margin maximization [16]. DL methods such as fully connected neural networks and convolutional neural networks can be used to extract abstract features via propagation through multi-layer perceptrons [8]. XGBoost is an end-to-end gradient boosted tree algorithm that minimizes the loss function over the weak learners in the form of decision trees [4].

3. Methods

In this section, we discuss the filtering and signal processing steps used in our approach. Then, we provide a detailed description of two proposed architectures to classify motor imagery tasks based on small time windows.

3.1. Mini-epoch generation

The algorithms of most previous works employed on the motor imagery classification problem directly use the frequency-domain transform of all the EEG signals as the input to the classification model. However, these approaches do not capture the relative significance among all the time steps. We want to approach this problem by simultaneously making use of the discriminating frequency-domain features and preserving the time-varying information.

We pull relevant information from the EEG signals by extracting the data 1 second to 4 seconds after the stimulus was cued. From here, we divide the input signals in the time domain into small sub-intervals, which we will refer to as "mini-epochs". We experimented with two mini-epoch generation strategies: fixed and sliding.

1. Fixed mini-epoch window: The time signal is divided into N non-overlapping intervals with equal length. In this approach, the next interval starts from the point where the first interval ends as shown in Figure 2. Each mini-epoch spans $T = \frac{3}{N}$ seconds.

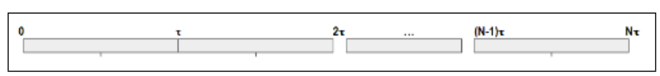

Figure 2. Fixed mini-epoch window for generating time intervals

2. Sliding mini-epoch window: With fixed window length T second of each epoch and the number of mini-epochs N determined, the mini-epoch can be generated by sliding the window by increment of $\Delta t = \frac{3-T}{N-1}$ seconds. Using this approach, we can more flexibly determine the mini-epoch duration and the number of mini-epochs while spanning the full time sequence via varying the parameters T and N. These mini-epochs can be overlapping as shown in Figure 3. Using this approach, we can more clearly observe the effect of

trimming the EEG signal into sub-intervals and look for an optimal mini-epoch length that extracts the significance of different time steps for the EEG signal.

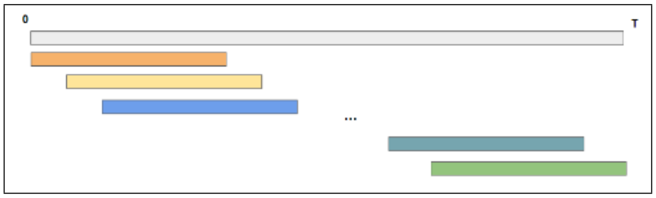

Figure 3. Sliding mini-epoch window for generating time intervals

3.2. Signal processing

A suitable signal processing procedure that performs desired feature extraction is crucial for motor imagery classification performance for EEG-based BCI systems, since the low-voltage electric signals received by EEG through scalp can be highly susceptible to noises. Furthermore, not all channels measured in the data set are responsible for the stimulus cued by the motor imagery tasks. We attempt to extract the relevant, filtered features that minimize the effect of the noise. Another goal of signal processing is to reduce the feature dimension so that only the discriminating features are used in classification. [6] provides a signal processing pipeline that they employ in their state-of-the-art method for this competition. Figure 4 shows the complete signal processing pipeline. In our approach, after adding our mini-epoch window implementation, we follow their pipeline which is described below.

1. Filter the signal through a bandpass filter from 7Hz to 30Hz, which is the only frequency range containing relevant information about the motor imagery tasks.

2. Remove signals from EOG channels and consider only the 22 EEG channels for motor imagery tasks.

3. Apply wavelet packet decomposition (WPD) technique to the resulting signal to extract features in the frequency domain. WPD is an extension of Wavelet Decomposition that comprises several bases with more filtering operations applied to the wavelets. [19].

4. Apply Common Spatial Pattern (CSP) technique for extracting spatial features from the signal. It has been widely used for feature extraction in EEG-based BCI systems for motor imagery [6].

3.3. Proposed algorithm

To validate our mini-epoch window implementation, we construct two algorithms. The first algorithm is based on mini-epoch learning. The second algorithm is based on mini-epoch ensemble learning.

3.3.1 Mini-epoch learning

Figure 5 shows our proposed architecture for mini-epoch learning.

After signal processing is applied to our data, we create a feature set out of the resulting signals. From here, this feature set is fed into a classifier to output the most probable class out of the entire time series. This classifier can be any classification algorithm, such as multi-layer perception (MLP), XGBoost, and support vector machine (SVM).

3.3.2 Mini-epoch ensemble learning

Figure 6 shows our proposed architecture for mini-epoch ensemble learning.

In this implementation, after signal processing is applied to our data, we feed the resulting signals to a selected base multi-class classifier. However, instead of directly yielding the predicted label, these base classifiers are employed to generate the normalized scores for each label. These normalized scores indicate the probability distribution of each label predicted within each mini-epoch. Like the classifier in the mini-epoch learning algorithm, the base classifiers can be implemented with any classification algorithm.

These N base classifiers form the ensemble to make the final prediction for the last step. The normalized scores generated from each classifier are fed into the ensemble learner to weigh the importance of each mini-epoch. Like the base classifier, the ensemble learner can be implemented with any algorithm. For example, a simple logistic regression model would weigh the mini-epoch based on each corresponding normalized score as a 4d vector. It can also be implemented with a single-layer neural network such that the probability score of each label for each base classifier will be evaluated individually to make the final prediction. Therefore, the ensemble learner can integrate the information from all mini-epochs, thus capturing the time-variant features via the normalized scores from their base classifiers.

4. Experiments

In this section, we describe the particular model and parameter selections for the experiment and discuss the performance of our proposed architecture under different setups.

4.1. Models

As discussed earlier, any classification model can be used as the multi-class classifier in the mini-epoch learning architecture. Similarly, any classification model can be used as the base classifier in the mini-epoch ensembling architecture. In this set of experiments, we select three classifiers that are commonly used in similar multi-class classification problems to demonstrate their improvements in ac-

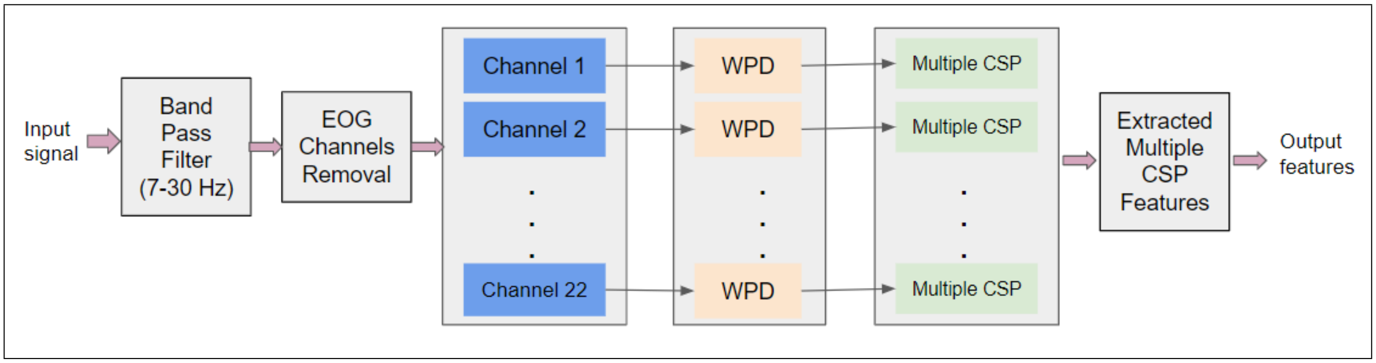

Figure 4. Signal processing pipeline.

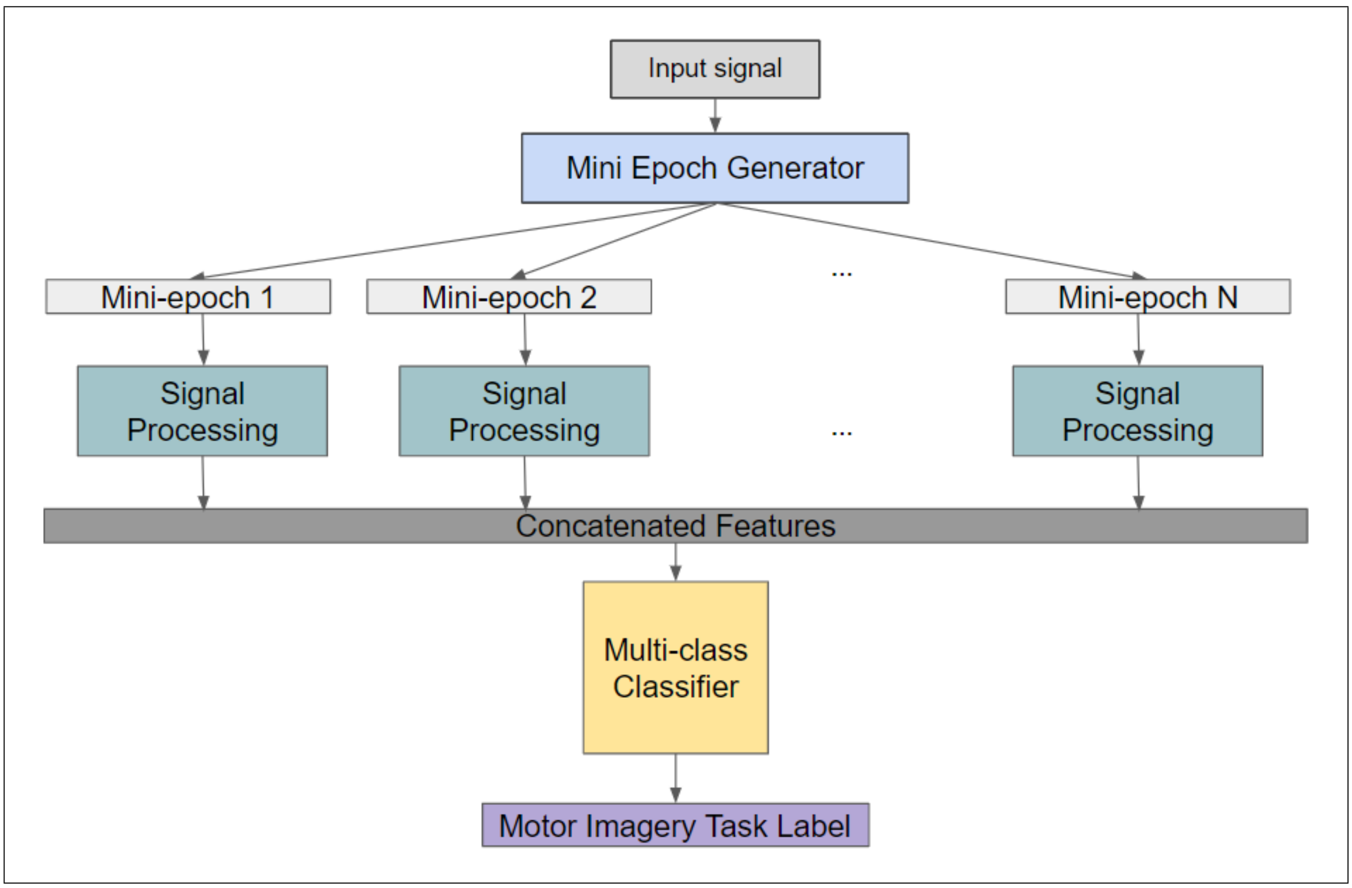

Figure 5. Proposed mini-epoch learning architecture.

curacy and their performance in comparison to the current state-of-the-art method. The three base classifiers are:

- Multi-layer perception (MLP): MLP is the one of most popular machine learning solutions nowadays to solve many regression and classification problems. It can fit as a base multi-class classifier with a final soft-max layer, and also function as an ensemble learner with a single layer of cross-entropy loss. Therefore, for our ensemble based architecture, it is possible to implement MLP in both places and realize an end-to-end training.
- Support vector machine (SVM): SVM is a classical machine learning method that is highly robust to dimensionality. With the mini-epoch approach, the number of features for each training instance will be scaled up by N since we will obtain frequency-domain trans-

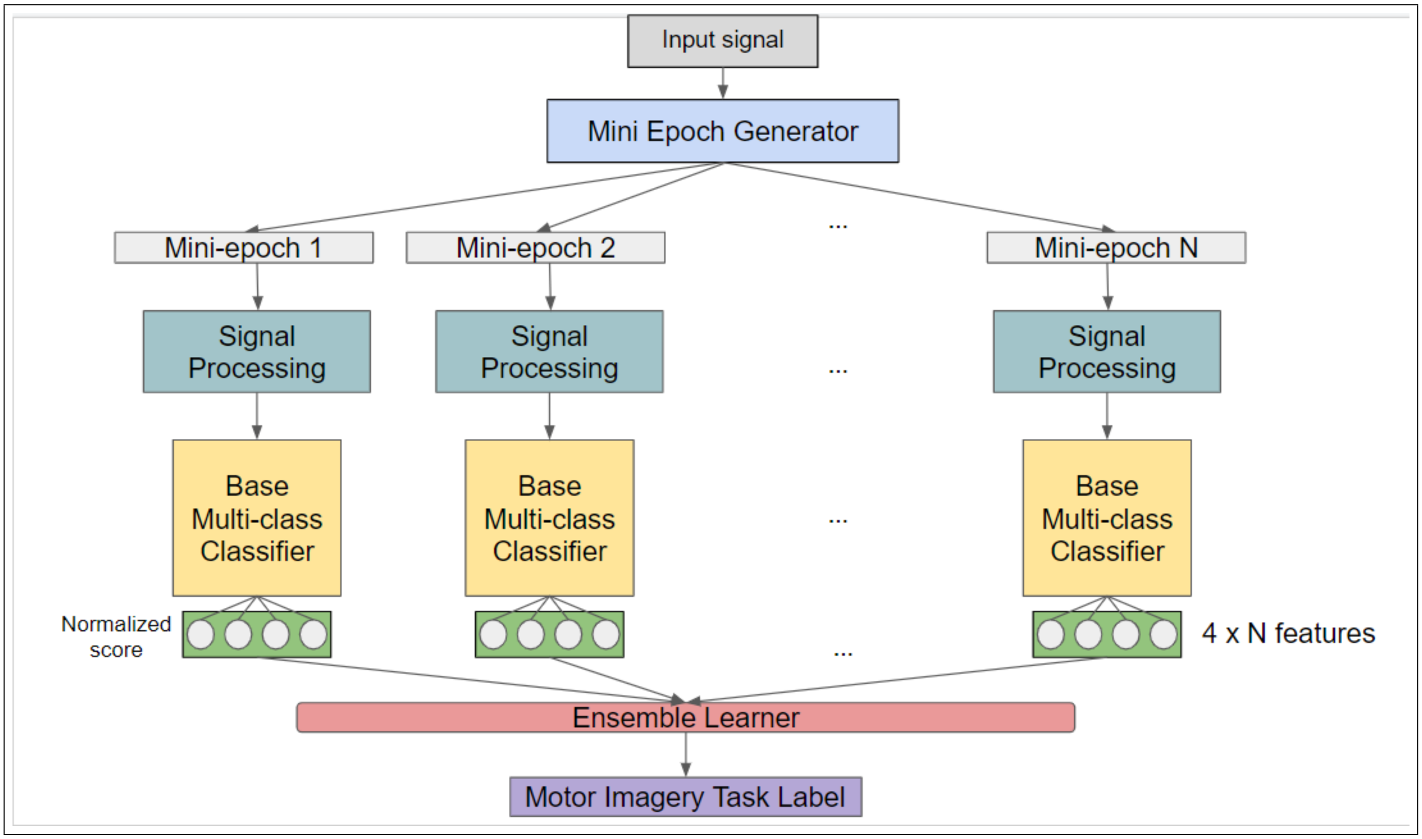

Figure 6. Proposed mini-epoch ensemble learning architecture.

formed features for each mini-epoch. SVM is thus particularly fitting for this architecture as it is reliable to perform in this scaled feature space.

- Gradient boosted trees: gradient boosting is known for being the best-performing model for tabular datasets. It trains an ensemble of weak learners in the form of decision trees to build the final classifier. In particular, XGBoost (XGB) is currently the most efficient implementation which we employ [4].

We also experimented with convolutional neural networks (CNN) for the base classifier which was used to achieve the current state-of-the-art accuracy. However, CNN has a worse performance when deployed into the ensemble-based approach. Due to the limitation of WPD and CSP, mini-epochs with shorter time intervals have much fewer features after passing through these signal processing techniques. The lower feature dimension renders these mini-epochs no longer suitable as input for CNN, since CNN only excels at extracting spatial features from a high-dimensional data example such as images. Therefore, we discarded experiment results with CNN from this paper.

We used the Scikit-learn [13] implementation for MLP and SVM of base learners. To ensure the consistency of our experiments, we manually tuned and fixed the hyper-parameters used for each type of base learner across all experiments. In particular, each MLP base learner consists of 3 fully connected layers of 50 neurons and was trained with 0.01 L_2 regularization. For XGB, we set the learning rate to be 0.3, the maximum depth of the tree to be 4, and strength L_2 regularization to be 100. All other parameters of models are set as default.

4.2. Experimental setup

As mentioned earlier, the dataset consists of 2 sessions for 9 subjects, and each of the sessions records 288 trials. Thus, there are 5184 instances provided in the dataset. We follow the method in [6] to randomly chose 80% of the dataset as the training set and the rest 20% is used as the testing set for performance evaluation purposes.

In these experiments, we prepare two different datasets. The first dataset comprises the unaltered, full-time sequence of the input signal. The second dataset used is created by generating N mini-epochs, either with the fixed or sliding mini-epoch window generation method. We refer to these datasets as the original and mini-epoch datasets, respectively. Both of these datasets were processed as described in Section 3.

In the mini-epoch datasets, we varied the number of mini-epochs (N) for the fixed mini-epoch window and the length of the interval (T) for the sliding mini-epoch window. For simplicity, we fixed the number of epochs for the sliding mini-epoch window to be 6.

4.3. Evaluation of mini-epoch learning

Our first experiment evaluates the performance of our proposed mini-epoch learning method by comparing the results of the classifier on the original and mini-epoch datasets. Table 1 shows the test accuracy of MLP, SVM, and XGB on the original and mini-epoch dataset with different parameters for N and T.

Table 1. Test accuracies of base learners on original ("original") and mini-epoch dataset ("fixed" and "sliding").

	MLP	**SVM**	**XGB**
Original	43.00%	45.90%	46.19%
Fixed N=3	53.62%	58.15%	56.32%
Fixed N=6	53.71%	61.91%	56.03%
Fixed N=12	58.73%	63.93%	55.83%
Sliding T=0.75s	56.12%	60.37%	56.99%
Sliding T=1.00s	57.47%	63.16%	59.59%
Sliding T=1.25s	57.57%	61.81%	60.46%

Compared to the results of the classifiers on the original dataset, the performances of the classifiers on the mini-epoch datasets are greatly improved by more than 10% on average. In particular, SVM benefits the most from using the mini-epoch dataset, where it increases by around 15% on fixed mini-epoch window with N = 12 and sliding mini-epoch window with T = 1.00s. Note that the best test accuracy (63.93%) attained using SVM on the fixed mini-epoch dataset is better than the best test accuracy (61.68%) reported in the state-of-the-art paper [6]. We also see a trend of test accuracy increases for MLP and SVM on the fixed mini-epoch window dataset and for XGB on the sliding mini-epoch window dataset, which suggests that the increase in the number of input features is generally helpful in boosting the generalization of the classifiers.

4.4. Evaluation of mini-epoch ensemble learning

In this set of experiments, we evaluate the performance of our mini-epoch ensemble learning classifier. We have chosen logistic regression implemented in Scikit-learn as the ensemble learner to combine results from different base learners. Similar to what we have done previously, we use the mini-epoch dataset with different parameters N for the fixed window case and T for the sliding window case. Table 2 shows the test accuracy of MLP, SVM, and XGB as the base learner for the ensemble learning classifier on the mini-epoch datasets with different parameters of N and T.

Table 2. Test accuracies of ensemble learners on the mini-epoch dataset ("fixed" and "sliding").

	MLP	**SVM**	**XGB**
Fixed N=3	43.97%	50.53%	52.94%
Fixed N=6	45.23%	54.58%	55.93%
Fixed N=12	51.11%	53.23%	60.95%
Sliding T=0.75s	50.63%	52.65%	55.74%
Sliding T=1.00s	51.11%	51.69%	57.76%
Sliding T=1.25s	58.24%	54.39%	61.04%

Although the results of the mini-epoch ensemble learning algorithms are generally worse than those of mini-epoch learning algorithms, we can see a clear trend of the performance increases of all base learners as N and T increase. In particular, XGB achieved a comparable result (61.04%) with the state-of-the-art sliding window dataset with T = 1.25s.

4.5. Future work

We have currently used 22 EEG channels in our model. However, according to [5], only the prefrontal cortex part of the brain is responsible for motor imagery. Hence, we can discard all the channels from other parts of the brain and only keep those that are in the prefrontal cortex. This will reduce the number of features while retaining the most important ones. Therefore, this can prove vital to the performance of the base classifiers and consequently, the complete model.

In addition, we would like to test our model on diverse datasets to validate the hypothesis and results. Besides this, more machine learning models can be explored for our multi-class classifiers along with hyperparameter tuning.

5. Conclusion

In this paper, we have shown that integrating mini-epoch generation with multi-class classifiers has promising potential in motor imagery classification. We explore two different algorithms that aim to retain the temporal information in EEG signals: mini-epoch learning and mini-epoch ensemble learning. Mini-epoch learning with SVM and mini-epoch ensemble learning with XGB obtained accuracies of 63.16% and 61.04%, respectively. We find that the mini-epoch learning algorithm beats other state-of-the-art methods when compared over the same dataset.

References

[1] Hamdi Altaheri, Ghulam Muhammad, Mansour Alsulaiman, Syed Amin, Ghadir Altuwaijri, Wadood Abdul, Mohamed Bencherif, and Mohammed Faisal. Deep learning techniques for classification of electroencephalogram (eeg) motor imagery (mi) signals:

A review. *Neural Computing and Applications*, 08 2021.

[2] Kai Keng Ang, Zheng Yang Chin, Chuanchu Wang, Cuntai Guan, and Haihong Zhang. Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b. *Frontiers in Neuroscience*, 6, 2012.

[3] Clemens Brunner, Robert Leeb, Gernot Müller-Putz, Alois Schlögl, and Gert Pfurtscheller. Bci competition 2008–graz data set a. *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, 16:1–6, 2008.

[4] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016. ACM.

[5] Jean Decety. The neurophysiological basis of motor imagery. *Behavioural Brain Research*, 77(1):45–52, 1996.

[6] Amira Echtioui, Wassim Zouch, Mohamed Ghorbel, Chokri Mhiri, and Habib Hamam. Fusion convolutional neural network for multi-class motor imagery of eeg signals classification. In *2021 International Wireless Communications and Mobile Computing (IWCMC)*, pages 1642–1647, 2021.

[7] Aymeric Guillot, Nady Hoyek, Magali Louis, and Christian Collet. Understanding the timing of motor imagery: recent findings and future directions. *International Review of Sport and Exercise Psychology*, 5(1):3–22, 2012.

[8] Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

[9] Jing Luo, Xing Gao, Xiaobei Zhu, Bin Wang, Na Lu, and Jie Wang. Motor imagery eeg classification based on ensemble support vector learning. *Computer Methods and Programs in Biomedicine*, 193:105464, 2020.

[10] D.J. McFarland and J.R. Wolpaw. Eeg-based brain–computer interfaces. *Current Opinion in Biomedical Engineering*, 4:194–200, 2017. Synthetic Biology and Biomedical Engineering / Neural Engineering.

[11] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil. Brain computer interfaces, a review. *sensors*, 12(2):1211–1279, 2012.

[12] Lawrence M Parsons. Imagined spatial transformations of one's hands and feet. *Cognitive Psychology*, 19(2):178–241, 1987.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[14] Girijesh Prasad, Pawel Herman, Damien Coyle, Suzanne Mcdonough, and Jacqueline Crosbie. Applying a brain-computer interface to support motor imagery practice in people with stroke for upper limb recovery: A feasibility study. *Journal of neuroengineering and rehabilitation*, 7:60, 12 2010.

[15] Franck Rienzo, Christian Collet, Nady Hoyek, and Aymeric Guillot. Impact of neurologic deficits on motor imagery: A systematic review of clinical evaluations. *Neuropsychology review*, 24, 04 2014.

[16] C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: a local svm approach. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 3, pages 32–36 Vol.3, 2004.

[17] Jerry J. Shih, Dean J. Krusienski, and Jonathan R. Wolpaw. Brain-computer interfaces in medicine. *Mayo Clinic Proceedings*, 87(3):268–279, 2012.

[18] Zhichuan Tang, Chao Li, and Shouqian Sun. Single-trial eeg classification of motor imagery using deep convolutional neural networks. *Optik - International Journal for Light and Electron Optics*, 130, 11 2016.

[19] Jian-Zhong Xue, Hui Zhang, Chong-Xun Zheng, and Xiang-Guo Yan. Wavelet packet transform for feature extraction of eeg during mental tasks. In *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693)*, volume 1, pages 360–363 Vol.1, 2003.