



# Auto-Weighted Multi-View Discriminative Metric Learning Method With Fisher Discriminative and Global Structure Constraints for Epilepsy EEG Signal Classification

Jing Xue<sup>1</sup>, Xiaoqing Gu<sup>2</sup> and Tongguang Ni<sup>2\*</sup>

<sup>1</sup> Department of Nephrology, The Affiliated Wuxi People's Hospital of Nanjing Medical University, Wuxi, China, <sup>2</sup> School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou, China

## OPEN ACCESS

### Edited by:

Mohammad Khosravi,  
Persian Gulf University, Iran

### Reviewed by:

Yufeng Yao,  
Changshu Institute of  
Technology, China  
Hongru Zhao,  
Soochow University, China

### \*Correspondence:

Tongguang Ni  
ntg@cczu.edu.cn

### Specialty section:

This article was submitted to  
Neuroprosthetics,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 22 July 2020

**Accepted:** 24 August 2020

**Published:** 29 September 2020

### Citation:

Xue J, Gu X and Ni T (2020)  
Auto-Weighted Multi-View  
Discriminative Metric Learning Method  
With Fisher Discriminative and Global  
Structure Constraints for Epilepsy EEG  
Signal Classification.  
*Front. Neurosci.* 14:586149.  
doi: 10.3389/fnins.2020.586149

Metric learning is a class of efficient algorithms for EEG signal classification problem. Usually, metric learning method deals with EEG signals in the single view space. To exploit the diversity and complementariness of different feature representations, a new auto-weighted multi-view discriminative metric learning method with Fisher discriminative and global structure constraints for epilepsy EEG signal classification called AMDML is proposed to promote the performance of EEG signal classification. On the one hand, AMDML exploits the multiple features of different views in the scheme of the multi-view feature representation. On the other hand, considering both the Fisher discriminative constraint and global structure constraint, AMDML learns the discriminative metric space, in which the intraclass EEG signals are compact and the interclass EEG signals are separable as much as possible. For better adjusting the weights of constraints and views, instead of manually adjusting, a closed form solution is proposed, which obtain the best values when achieving the optimal model. Experimental results on Bonn EEG dataset show AMDML achieves the satisfactory results.

**Keywords:** metric learning, multi-view learning, auto-weight, EEG signal classification, epilepsy

## INTRODUCTION

Epilepsy is characterized by an unexpected seizure periodicity, where brain temporary dysfunction is caused by abnormal discharge of neurons (Kabir and Zhang, 2016; Gummadavelli et al., 2018; Li et al., 2019). During the seizure, motor dysfunction, intestinal and bladder dysfunction, loss of consciousness, and other cognitive dysfunction often occur. Since the occurrence of epilepsy is often accompanied by changes in spatial organization and temporal dynamics of brain neural neurons, many brain imaging methods are used to reveal abnormal changes in brain neural neurons caused by epilepsy. EEG signal is an important signal to record the activity of neurons in the brain. It uses electrophysiological indicators to record the changes in the electrical wave of the cerebral cortex generated during brain activity. It is the overall reflection of the activity of brain neurons in the cerebral cortex. Many clinical studies have shown that due to abnormal discharge of brain neurons, epilepsy-specific waveforms, such as spikes and sharp waves, appear during or shortly before the onset of seizures, so identifying EEG signals is an effective detection of epilepsy method. Clinically, the detection of seizures based on EEG signals mainly relies on the personal experience

of doctors. However, modern EEG recorders can generate up to 1,000 data points per second, and the standard recording process can last for several days. This procedure will make manual screening require a lot of physical and mental exhaustion, and after a long period of observation, the doctor's judgment is easily affected by fatigue.

With the gradual development of smart healthcare, more and more machine learning algorithms are applied to the detection of epilepsy of EEG signals (Jiang et al., 2017a; Juan et al., 2017; Usman and Fong, 2017; Richhariya and Tanveer, 2018; Cury et al., 2019). In the view of machine learning, the EEG signal recognition contains two stages: feature extraction and classification method. The commonly used feature extraction methods for EEG signals are time-domain feature extractions and frequency-domain feature extractions (Srinivasan et al., 2005; Tzallas et al., 2009; Iscan et al., 2011). Since the original EEG signal is the time series signal, the time-domain feature extractions are generally based on the original EEG signal; then, the relevant statistics of the time series are calculated, and the epilepsy EEG features are extracted, using the kernel principal component analysis (KPCA) (Smola, 1997). The frequency-domain features are to transform the original EEG signal in the time domain to the frequency domain and then extract the relevant frequency-domain features as EEG features (Griffin and Lim, 1984). Although these feature extraction methods provide good performance in some practical applications, there is no feature extraction method that can be applied to all application scenarios. EEG signals are generated by numerous brain neuron activities. Due to the non-linear and non-static nature of EEG signals, how to extract effective features is still an important challenge. For those reasons, the multiple feature based multi-view learning concept has become a hot topic in EEG signal classification (Yuan et al., 2019; Wen et al., 2020). Different from using the single feature type, the multi-view learning method can comprehensively use a set of data features obtained from multiple ways or multiple levels. Each view of the data features may contain specific information not available in other views. Specifically, these independent and diverse features can be extracted from time-domain, frequency-domain, and multilevel features of signals. Appropriately designed multi-view learning can significantly promote the performance of EEG signal classification. For example, Spyrou et al. (2018) proposed a multiple features-based classifier to use spatial, temporal, or frequency EEG data. This classifier performs dimensionality reduction and rejects components through evaluating the classification performance. Two multi-view Takagi–Sugeno–Kang fuzzy systems for epileptic EEG signals classification are proposed in Zhou et al. (2019) and Jiang et al. (2017b), respectively. The former fuzzy system is developed in a deep view-reduction framework, and the latter fuzzy system is developed in a multi-view collaborative learning mechanism.

Besides the multi-view learning, classification algorithm is very important for EEG signal classification. One of the recent trends is the metric learning method. Metric learning method learns a more suitable distance measurement criterion in the feature space from the training data. Metric learning can be

used for specific tasks, such as classification and clustering, so as to more accurately represent the similarity between samples. Different from traditional Euclidean distance, such as nearest-neighbors classifiers and K-means, metric learning aims to find the appropriate similarity measures between data pairs to maintain the required distance structure (Cai et al., 2015; Wang et al., 2015; Lu et al., 2016). The appropriately distance metrics can provide a good measure of the similarity and dissimilarity between different samples. For example, Liu et al. (2014) developed a similarity metric-learning in the process of EEG P300 wave recognition. Compared with traditional Euclidean metric, the proposed global Mahalanobis distance metric shows the better discriminative representation. Phan et al. (2013) proposed a metric learning method using the global distance metric from labeled examples. This method successfully applied on single-channel EEG data for sleep staging and does not need artifact removal or bootstrapping preprocessing steps. Alwasiti et al. (2020) proposed a deep metric learning model and tested it for motor imagery EEG signals classification. The experimental results show that the proposed deep metric learning model can converge with very small number of training EEG signals.

Inspired by the distance metric and multi-view learning, we present a new auto-weighted multi-view discriminative metric learning method with Fisher discriminative and global structure constraints for EEG signal classification called AMDML. To better exploit the correlation and complementary data features of multiple views, both the Fisher discriminative constraint and global structure constraint are adopted in the construction process of the distance metric matrix. In such common metric space, the intraclass EEG signals are compact, and interclass EEG signals are separable as much as possible. Simultaneously, an auto-weighted learning strategy is developed to automatically adjust constraint and view weights during the model learning process. The contributions of our work are as follows: (1) Both Fisher discriminative and global structure information of multiple view data features are considered in the multi-view metric learning model with the high discriminative performance; (2) in the optimization process, the constraint and view weights can be adjusted auto-weighted by the closed form solution, instead of adjusted manually. Thus, the constraints balance and multiple view collaboration can be optimized; and (3) the experimental results on Bonn EEG dataset justify the applicability of AMDML for EEG signal classification.

## RELATED WORK

### Metric Learning

Here, we introduce the baseline method in the study. Xing et al. (2003) proposed a distance metric considering side-information (DMSI) method. Using the given similar and dissimilar pairs of samples, DMSI learns a good distance metric to identify the “similar” relationship between all pairs of samples so that similar pairs are close and dissimilar pairs are separated. Let  $S$  and  $D$  be two sets of pair as

$$S = \{(\mathbf{x}_i, \mathbf{x}_j) | \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are similar}\}, \quad (1)$$

$$D = \{(\mathbf{x}_i, \mathbf{x}_j) | \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are dissimilar}\}. \quad (2)$$

The optimization problem of DMSI is represented as

$$\begin{aligned} \min_M \quad & \sum_{(x_i, x_j) \in S} \|x_i - x_j\|_M^2 \\ \text{s.t.} \quad & \sum_{(x_i, x_j) \in D} \|x_i - x_j\|_M \geq 1, \\ & M \geq 0. \end{aligned} \tag{3}$$

A key point in DMSI is that all samples that are not clearly identified as similar are dissimilar. In addition, metric learning tries to find an appropriate measurement to preserve the distance structure. The distance metric considers a positive semidefinite matrix  $M$ , and  $\|x_i - x_j\|_M$  is induced as a Mahalanobis distance

$$\|x_i - x_j\|_M = \sqrt{(x_i - x_j)^T M (x_i - x_j)} \tag{4}$$

When the learned  $M$  is a diagonal matrix, **Equation (3)** can be solved by the Newton–Raphson method; when  $M$  is a full matrix, **Equation (3)** can be solved by an iterative optimization algorithm with gradient ascent and iterative projection strategies.

### Bonn EEG Dataset

The EEG signal data in the experiment is from the website of the Bonn University, Germany (Tzallas et al., 2009). The Bonn EEG dataset contains five groups of EEG signal sets called as groups A–E. The example samples in groups A–E are shown in **Figure 1**. Each EEG data group consists of 100 single-channel EEG signal segments of 23.6 s and 173.6 Hz rate. The basic information of five groups is listed in **Table 1**. EEG signal data in groups A and B is sampled from five healthy volunteers, and EEG signal data

in groups C–E is sampled from five patients at different states of epileptic seizure.

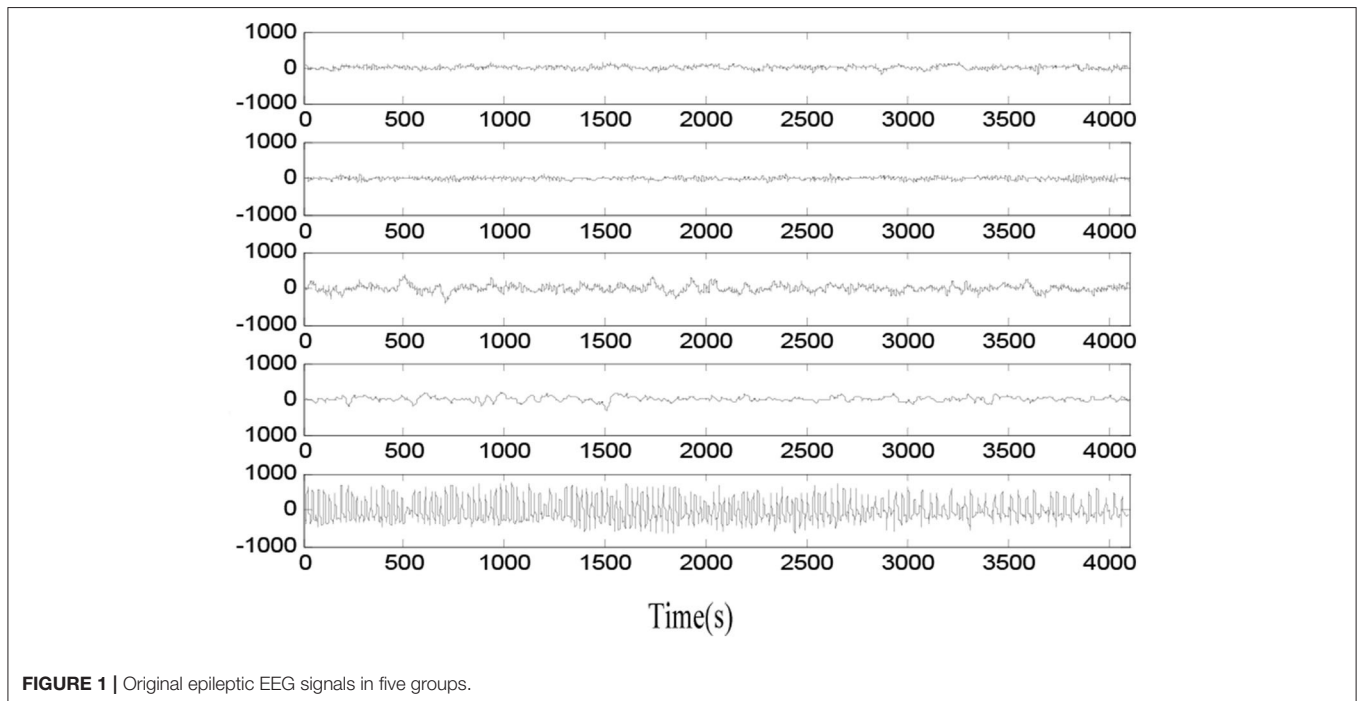
## PROPOSED METHOD

### Objective Function

After collecting a set of EEG signals, we obtain  $N$  samples presented as  $\{x_i, l_i\}_{i=1}^N$ , where  $l_i$  is the class label of sample  $x_i$ . According to the label information, we construct two sets of sample pairs such that  $X_s = \{(x_i, x_j) | l_i = l_j\}$  is the intraclass sample set and  $X_d = \{(x_i, x_j) | l_i \neq l_j\}$  is the interclass sample set. Then, generating multiple view of data samples, we obtain  $\{x_i^m\}_{m=1}^M$  from  $M$  different view features of each sample, where  $x_i^m$  is the  $m$ th view of sample  $x_i$ . For discriminative projecting, we build the  $k$ -nearest neighbor intraclass graph  $\{G^m\}_{m=1}^M$  and interclass graph  $\{P^m\}_{m=1}^M$  in each view, which use the supervised information to describe the local geometrical structure of the

**TABLE 1** | The basic information of EEG data groups of A–E.

Group	Basic information
A	EEG signals of healthy volunteers in awoken state with eyes open
B	EEG signals of healthy volunteers in awoken state with eyes closed
C	EEG signals of patients in hippocampal formation of the opposite hemisphere of brain
D	EEG signals of patients in epileptogenic zone during periodic lulls
E	EEG signals of patients during seizure activity



**FIGURE 1** | Original epileptic EEG signals in five groups.

data. The intraclass graph  $G^m$  can be computed as

$$G_{ij}^m = \begin{cases} 1/|X_i| & \text{if } \mathbf{x}_i^m \in N_{k_1}(\mathbf{x}_j^m) \text{ or } \mathbf{x}_j^m \in N_{k_1}(\mathbf{x}_i^m) \text{ and } (\mathbf{x}_i, \mathbf{x}_j) \in X_s, \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where  $N_{k_1}(\mathbf{x}_j^m)$  denotes the intraclass sample set containing the  $k_1$  nearest neighbors of  $\mathbf{x}_j^m$ .

The graph  $P^m$  can be computed as

$$P_{ij}^m = \begin{cases} 1/|X_d| & \text{if } \mathbf{x}_i^m \in N_{k_2}(\mathbf{x}_j^m) \text{ or } \mathbf{x}_j^m \in N_{k_2}(\mathbf{x}_i^m) \text{ and } (\mathbf{x}_i, \mathbf{x}_j) \in X_d, \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where  $N_{k_2}(\mathbf{x}_j^m)$  denotes the interclass sample set containing the  $k_2$  nearest neighbors of  $\mathbf{x}_j^m$ .

Then, the intraclass correlation constraint  $F_G^m$  from the  $m$ th view can be written as

$$\begin{aligned} F_G^m &= \frac{1}{2} \sum_{i,j} G_{ij}^m \left\| \mathbf{H}^T \mathbf{x}_i^m - \mathbf{H}^T \mathbf{x}_j^m \right\|_2^2 \\ &= \frac{1}{2} \sum_{i,j} P_{ij}^m (\mathbf{H}^T (\mathbf{x}_i^m - \mathbf{x}_j^m) (\mathbf{x}_i^m - \mathbf{x}_j^m)^T \mathbf{H}) \\ &= \text{Tr}(\mathbf{H}^T \mathbf{X}^m \mathbf{L}_G^m \mathbf{X}^{mT} \mathbf{H}) \end{aligned} \quad (7)$$

where  $\mathbf{L}_G^m$  is the Laplacian matrix on  $G^m$ , and  $\mathbf{L}_G^m$  is computed as  $\mathbf{L}_G^m = \mathbf{D}_G^m - \mathbf{G}^m$ .  $\mathbf{D}_G^m$  is a diagonal matrix, and the element in  $\mathbf{D}_G^m$  is  $D_{G,i,i}^m = \sum_j G_{ij}^m$ .  $\text{Tr}()$  is for the trace operator.

The interclass correlation constraint  $F_P^m$  from the  $m$ th view can be written as

$$\begin{aligned} F_P^m &= \frac{1}{2} \sum_{i,j} P_{ij}^m \left\| \mathbf{H}^T \mathbf{x}_i^m - \mathbf{H}^T \mathbf{x}_j^m \right\|_2^2 \\ &= \frac{1}{2} \sum_{i,j} P_{ij}^m (\mathbf{H}^T (\mathbf{x}_i^m - \mathbf{x}_j^m) (\mathbf{x}_i^m - \mathbf{x}_j^m)^T \mathbf{H}) \\ &= \text{Tr}(\mathbf{H}^T \mathbf{X}^m \mathbf{L}_P^m \mathbf{X}^{mT} \mathbf{H}) \end{aligned} \quad (8)$$

where  $\mathbf{L}_P^m$  is the Laplacian matrix on  $P^m$ , and  $\mathbf{L}_P^m$  is computed as  $\mathbf{L}_P^m = \mathbf{D}_P^m - \mathbf{P}^m$ .  $\mathbf{D}_P^m$  is a diagonal matrix, and the element in  $\mathbf{D}_P^m$  is  $D_{P,i,i}^m = \sum_j P_{ij}^m$ .

For global structure knowledge of multiple-view data preservation, the following global structure consistency  $Q^m$  is employed

$$\begin{aligned} Q^m &= \frac{1}{2} \sum_{i,j} W_{ij} \left\| \mathbf{H}^T \mathbf{x}_i^m - \mathbf{H}^T \mathbf{x}_j^m \right\|_2^2 \\ &= \frac{1}{2} \sum_{i,j} W_{ij} (\mathbf{H}^T (\mathbf{x}_i^m - \mathbf{x}_j^m) (\mathbf{x}_i^m - \mathbf{x}_j^m)^T \mathbf{H}) \\ &= \text{Tr}(\mathbf{H}^T \mathbf{X}^m \mathbf{L}_W \mathbf{X}^{mT} \mathbf{H}) \end{aligned} \quad (9)$$

where  $W$  is an adjacent matrix whose element is  $W_{ij} = 1/N^2$ .  $L_W$  is the Laplacian matrix on  $W$ , and  $L_W$  is computed as  $L_W = D_W - W$ ,  $D_W$  is a diagonal matrix, and the element in  $D_W$  is  $D_{W,i,i}^m = 1/N$ .

The basic principle of  $Q^m$  is to use the global structural information through cross-view data covariance. The term  $X^m L_W X^{mT}$  is equivalent to the centering matrix of the  $m$ th

view data, i.e.,  $X^m L_W X^{mT} = E[(\mathbf{x} - \frac{1}{N} \sum_i \mathbf{x}_i)(\mathbf{x} - \frac{1}{N} \sum_i \mathbf{x}_i)^T]$ .

It represents the average squared distance between all samples of the  $m$ th view in the metric space. Therefore,  $Q^m$  can be considered as a principal component analysis (PCA) (Smola, 1997)—like the regularization term in the  $m$ th view.

The goal of AMDML is to find an optimal discriminative distance metric in a multi-view learning model, and in such metric space, it can exploit the complementary information of different view data features and further enforce the proposed method to be more discriminative. To achieve this goal, we learn a metric that maximizes the Fisher analysis constraint (interclass/intraclass correlation ratio), simultaneously maximizing the preservation of the global structure consistency constraint. The objective function of AMDML is designed as

$$\begin{aligned} \min_{\mathbf{H}, \Theta} & \sum_{m=1}^M \Theta_m^r \frac{F_G^m - F_P^m}{Q^m}, \\ \text{s.t. } & \mathbf{H}^T \mathbf{H} = \mathbf{I}, \\ & \sum_{m=1}^M \Theta_m = 1, \Theta_m \geq 0, \end{aligned} \quad (10)$$

The projection matrix  $H$  helps to build a discriminative metric space among multiple views, such that the feature correlation and complementary structural information among multiple views can be exploited. The vector  $\Theta = [\Theta_1, \Theta_2, \dots, \Theta_M]$  is the view weight vector, and its element  $\Theta_m$  indicates the role importance of the  $m$ th view. When  $\Theta_m$  tends to 0, it means the data features of the  $m$ th view are useless for discrimination task. The  $\Theta_m=1$  means that only one type of data features from one view is used in AMDML, and in this case, Equation (10) is a single view learning problem. To better utilize the complementary information of multiple features rather than the best feature, we use index parameter  $r$  ( $r > 1$ ) on  $\Theta_m$ .

Equation (10) can be represented as

$$\begin{aligned} \min_{\mathbf{H}, \Theta} & \sum_{m=1}^M \Theta_m^r \frac{\text{Tr}(\mathbf{H}^T \mathbf{X}^m (\mathbf{L}_G^m - \mathbf{L}_P^m) \mathbf{X}^{mT} \mathbf{H})}{\text{Tr}(\mathbf{H}^T \mathbf{X}^m \mathbf{L}_W \mathbf{X}^{mT} \mathbf{H})}, \\ \text{s.t. } & \mathbf{H}^T \mathbf{H} = \mathbf{I}, \\ & \sum_{m=1}^M \Theta_m = 1, \Theta_m \geq 0. \end{aligned} \quad (11)$$

However, the optimization of Equation (11) involves a complex operation of inverse. Using a constraint weight parameter  $\gamma$ , we reconstruct Equation (11) into the following weighted optimization model

$$\begin{aligned} \min_{\mathbf{H}, \Theta, \gamma} & \sum_{m=1}^M \Theta_m^t (\gamma^2 \mathbf{H}^T \mathbf{X}^m (\mathbf{L}_G^m - \mathbf{L}_P^m) \mathbf{X}^{mT} \mathbf{H} - \gamma \text{Tr}(\mathbf{H}^T \mathbf{X}^m \mathbf{L}_W \mathbf{X}^{mT} \mathbf{H})), \\ \text{s.t. } & \mathbf{H}^T \mathbf{H} = \mathbf{I}, \\ & \sum_{m=1}^M \Theta_m = 1, \Theta_m \geq 0. \end{aligned} \quad (12)$$

where  $\gamma$  represents a constraint weight tradeoff Fisher discriminative constraint and global structure constraint. It is noted that the constraint weight  $\gamma$  and view weight  $\Theta$  are not a manually adjusted parameters. In this study, we adaptively adjust  $\gamma$  and  $\Theta$  in two closed form solutions, respectively.

### Optimization

Because the optimization problem of Equation (12) is a non-linear constrained non-convex problem, in this study, we solve the optimization problem in Equation (12) using the iteratively optimization strategy to obtain the AMDML parameters of  $\mathbf{H}$ ,  $\Theta$ , and  $\gamma$ . First, we tune parameter  $\mathbf{H}$  while fixing parameters  $\Theta$  and  $\gamma$ . The optimization problem in Equation (12) can be reformulated as follows:

$$\min_{\mathbf{H}} \sum_{m=1}^M \Theta_m^t (\gamma^2 \text{Tr}(\mathbf{H}^T \mathbf{X}^m (\mathbf{L}_G^m - \mathbf{L}_P^m) \mathbf{X}^{mT} \mathbf{H}) - \gamma \text{Tr}(\mathbf{H}^T \mathbf{X}^m \mathbf{L}_W \mathbf{X}^{mT} \mathbf{H})), \quad (13)$$

s.t.  $\mathbf{H}^T \mathbf{H} = \mathbf{I}$ ,

Thus,  $\mathbf{H}$  can be easily calculated by solving the eigenvalue decomposition problem as follows:

$$\left( \sum_{m=1}^M \Theta_m^t (\gamma^2 \mathbf{X}^m (\mathbf{L}_G^m - \mathbf{L}_P^m) \mathbf{X}^{mT} - \gamma (\mathbf{X}^m \mathbf{L}_W \mathbf{X}^{mT})) \right) \mathbf{H} = \alpha \mathbf{H} \quad (14)$$

In terms of the Lagrange optimization, the minimization of Equation (14) can be converted with multiplier as follows:

$$J(\Theta, \gamma, \alpha) = \sum_{m=1}^M \Theta_m^t \left( \gamma^2 \text{Tr}(\mathbf{H}^T \mathbf{X}^m (\mathbf{L}_G^m - \mathbf{L}_P^m) \mathbf{X}^{mT} \mathbf{H}) - \gamma \text{Tr}(\mathbf{H}^T \mathbf{X}^m \mathbf{L}_W \mathbf{X}^{mT} \mathbf{H}) \right) - \alpha \left( \sum_{m=1}^M \Theta_m^t - 1 \right). \quad (15)$$

Next, we tune parameter  $\Theta$  while fixing parameters  $\mathbf{H}$  and  $\gamma$ .

Let  $\frac{\partial J(\Theta, \gamma, \alpha)}{\partial \Theta_m} = 0$  and  $\frac{\partial J(\Theta, \gamma, \alpha)}{\partial \alpha} = 0$ , we have

$$\begin{cases} t \Theta_m^{t-1} \gamma^2 (\text{Tr}(\mathbf{H}^T \mathbf{X}^m (\mathbf{L}_G^m - \mathbf{L}_P^m) \mathbf{X}^{mT} \mathbf{H}) - \gamma \text{Tr}(\mathbf{H}^T \mathbf{X}^m \mathbf{L}_W \mathbf{X}^{mT} \mathbf{H})) - \alpha = 0, \\ \sum_{m=1}^M \Theta_m - 1 = 0. \end{cases} \quad (16)$$

We can obtain  $\Theta_m$  as follows:

$$\Theta_m = \frac{(1/(\gamma^2 \text{Tr}(\mathbf{H}^T \mathbf{X}^m (\mathbf{L}_G^m - \mathbf{L}_P^m) \mathbf{X}^{mT} \mathbf{H}) - \gamma \text{Tr}(\mathbf{H}^T \mathbf{X}^m \mathbf{L}_W \mathbf{X}^{mT} \mathbf{H})))^{1/(t-1)}}{\sum_{m=1}^M (1/(\gamma^2 \text{Tr}(\mathbf{H}^T \mathbf{X}^m (\mathbf{L}_G^m - \mathbf{L}_P^m) \mathbf{X}^{mT} \mathbf{H}) - \gamma \text{Tr}(\mathbf{H}^T \mathbf{X}^m \mathbf{L}_W \mathbf{X}^{mT} \mathbf{H})))^{1/(t-1)}}. \quad (17)$$

Finally, we tune parameter  $\gamma$  while fixing parameters  $\mathbf{H}$  and  $\Theta$ . In terms of the Lagrange optimization, the solution of  $\gamma$  is  $\frac{\partial J(\Theta, \gamma, \alpha)}{\partial \gamma} = 0$ ; we can obtain  $\gamma$  as follows:

$$\gamma = \frac{\Theta_m^t \sum_{m=1}^M \text{Tr}(\mathbf{H}^T \mathbf{X}^m \mathbf{L}_W \mathbf{X}^{mT} \mathbf{H})}{2 \Theta_m^t \sum_{m=1}^M \text{Tr}(\mathbf{H}^T \mathbf{X}^m (\mathbf{L}_G^m - \mathbf{L}_P^m) \mathbf{X}^{mT} \mathbf{H})} \quad (18)$$

Based on the above analysis, the proposed AMDML method is presented in **Algorithm 1**.

**Algorithm 1** | The proposed AMDML method.

**Input:** M views of  $m$  pairs of EEG signals;

**Output:** the best metric  $H = H_t$ .

Set  $\Theta = [1/M, 1/M, \dots, 1/M]$  and compute  $\mathbf{H}$  using **Equation (13)**.

Repeat

$t = t+1$

Fix  $\mathbf{H}(t)$ , and compute  $\Theta(t)$  using **Equation (17)**;

Fix  $\Theta(t)$ , and compute  $\mathbf{H}(t)$  using **Equation (18)**

Compute  $L(t)$  using **Equation (15)**;

Until  $\|J(t)-J(t-1)\| \leq \delta$  or  $t \geq t_{\max}$

**TABLE 2** | Ten EEG classification tasks.

Tasks	EEG signal groups
Task 1	A and C
Task 2	A and D
Task 3	A and E
Task 4	B and C
Task 5	B and D
Task 6	B and E
Task 7	{A, B} and {C, D}
Task 8	{A, B} and E
Task 9	{A, B} and {D, E}
Task 10	{A, B} and {C, E}

## EXPERIMENT

### Experimental Settings

In the experiment, we extract three types of data feature including KPCA, wavelet packet decomposition (WPD) (Wu et al., 2008), and short-time Fourier transform (STFT) (Griffin and Lim, 1984). We design 10 classification tasks, and the basic information of tasks is as shown in **Table 2**. In order to show the performance of our method, we compare AMDML with four single-view classification methods [including DMSI (Xing et al., 2003), large margin nearest neighbor (LMNN) (Weinberger and Saul, 2009), neighborhood preserving embedding (NPE) (Wen

et al., 2010), and RDML-CCPVL (Ni et al., 2018)] and three multi-view methods [including MvCVM (Huang et al., 2015), VMRML-LS (Quang et al., 2013), and DMML (Zhang et al., 2019)]. In the LMNN method, the number of target neighbors  $k$  was set to  $k = 3$ , and the weighting parameter  $\mu$  is selected within the grid  $\{0, 0.2, \dots, 1\}$ . In the RDML-CCPVL method, the regularization parameter is selected within the grid  $[0.01, 0.1, 0.5, 1, 5, 10, 20]$  and the number of clusters is selected within the grid  $[2, 3, \dots, 20]$ . In the MvCVM method, the regularization parameter is selected within the grid  $[0.01, 0.1, 0.5, 1, 5, 10, 20]$ . In VMRML-LS, the regularization parameters  $\gamma_A = 10^{-5}$ ,  $\gamma_B = 10^{-6}$ , and  $\gamma_W = 10^{-6}$ , and the element in weight vector is selected in  $[1,$

**TABLE 3** | The classification performances of AMDML on 10 classification tasks.

		WPD	STFT	KPCA	Full views
Task 1	LMNN	95.00	96.14	96.19	96.18
	NPE-KNN	95.05	96.29	96.34	96.38
	NPE-SVM	95.08	96.17	96.29	96.49
	RDML-CCPVL	96.07	96.64	96.58	96.97
	DMSI	96.11	96.53	96.52	96.88
	AMDML-KNN	97.75	97.62	97.88	<b>99.19</b>
	AMDML-SVM	97.86	97.55	97.86	99.18
Task 2	LMNN	93.92	94.73	95.00	95.88
	NPE-KNN	94.01	95.03	95.21	95.71
	NPE-SVM	94.06	95.00	94.54	96.07
	RDML-CCPVL	94.60	95.06	95.88	96.17
	DMSI	94.45	95.08	94.64	96.22
	AMDML-KNN	97.29	97.86	98.38	99.77
	AMDML-SVM	97.22	98.03	98.30	<b>99.78</b>
Task 3	LMNN	94.75	94.92	95.11	95.40
	NPE-KNN	94.01	94.01	95.42	95.72
	NPE-SVM	94.98	94.08	95.41	95.64
	RDML-CCPVL	94.56	95.59	95.88	95.70
	DMSI	94.72	95.34	95.26	96.24
	AMDML-KNN	96.72	98.82	98.35	99.35
	AMDML-SVM	96.83	98.90	98.48	<b>99.41</b>
Task 4	LMNN	92.76	92.90	93.07	94.45
	NPE-KNN	94.00	93.02	94.48	94.70
	NPE-SVM	93.00	93.04	93.42	94.65
	RDML-CCPVL	93.52	93.56	93.90	94.92
	DMSI	94.21	94.01	94.45	95.03
	AMDML-KNN	96.71	97.88	97.44	<b>99.42</b>
	AMDML-SVM	96.88	97.86	97.42	99.35
Task 5	LMNN	91.81	93.89	94.10	95.46
	NPE-KNN	92.99	93.08	93.43	95.71
	NPE-SVM	92.98	93.13	94.46	95.66
	RDML-CCPVL	94.54	94.66	94.91	95.99
	DMSI	94.50	95.04	95.11	95.37
	AMDML-KNN	98.76	98.87	98.40	99.39
	AMDML-SVM	99.83	98.89	98.39	<b>99.41</b>
Task 6	LMNN	96.72	96.92	96.16	97.36
	NPE-KNN	96.05	96.93	97.09	97.22
	NPE-SVM	96.91	97.01	97.22	97.42
	RDML-CCPVL	96.57	97.62	96.90	97.37
	DMSI	96.49	97.49	96.81	97.02
	AMDML-KNN	97.73	98.89	97.39	<b>99.45</b>
	AMDML-SVM	97.89	98.94	97.35	99.38
Task 7	LMNN	93.76	93.87	93.05	95.37
	NPE-KNN	93.93	94.06	93.41	95.73
	NPE-SVM	96.03	95.97	95.43	95.68
	RDML-CCPVL	95.54	95.59	94.94	95.94
	DMSI	95.29	95.36	94.73	95.08

*(Continued)*

TABLE 3 | Continued

		WPD	STFT	KPCA	Full views
Task 8	AMDML-KNN	97.74	97.89	97.31	99.39
	AMDML-SVM	97.92	97.81	97.36	<b>99.44</b>
	LMNN	94.79	95.86	95.12	96.40
	NPE-KNN	94.95	96.00	95.44	96.41
	NPE-SVM	94.94	95.99	95.45	96.09
	RDML-CCPVL	95.57	96.09	95.89	96.14
	DMSI	95.63	95.94	96.15	96.18
Task 9	AMDML-KNN	97.72	98.82	98.41	99.45
	AMDML-SVM	97.90	98.92	98.45	<b>99.46</b>
	LMNN	92.78	94.94	94.06	95.37
	NPE-KNN	92.93	95.05	94.44	95.68
	NPE-SVM	93.01	95.04	95.43	95.62
	RDML-CCPVL	94.52	95.66	95.94	95.98
	DMSI	94.67	92.51	96.04	96.40
Task 10	AMDML-KNN	96.69	97.83	97.37	<b>99.42</b>
	AMDML-SVM	96.99	97.90	97.43	99.38
	LMNN	93.87	93.86	93.13	94.31
	NPE-KNN	94.02	92.97	93.44	94.49
	NPE-SVM	93.04	94.09	94.48	94.66
	RDML-CCPVL	95.57	95.56	94.98	96.01
	DMSI	95.17	94.22	94.10	94.56
	AMDML-KNN	97.78	97.83	96.41	<b>98.40</b>
	AMDML-SVM	97.89	97.97	96.38	98.37

The bold values represents mean the best classification performance in the tasks.

5, 10]. In the DMML method, the number of interclass marginal samples is selected within the grid [1, 2,..., 10]. In the proposed AMDML method, the parameters  $k_1$  and  $k_2$  in Equations (5) and (6) are selected in [2, 3,..., 10]. The widely used  $K$ -nearest neighbor (KNN) and support vector machine (SVM) are used as the classifiers for the proposed AMDML, and we name them as AMDML-KNN and AMDML-SVM, respectively. We empirically set the nearest neighborhood number of KNN classifier as [1, 3,..., 9] and train SVM model using LIBSVM (Chang and Lin, 2011). All methods are implemented in MATLAB using a computer with 2.6 GHz dual-core CPU and 8 GB RAM.

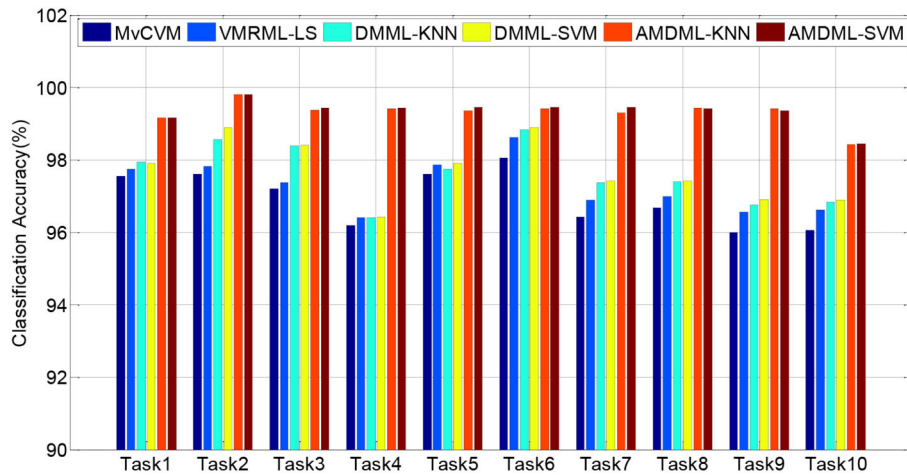
## Performance Comparisons With Single-View Methods

We first compare the performance of AMDML with several single-view classification methods. NPE using two classifiers KNN and SVM are named as NPE-KNN and NPE-SVM, respectively. Table 3 shows the classification performance of these methods on Bonn EEG dataset using three signal views (WPD, STFT, and KPCA) and full views. When AMDML uses single-view feature data, the parameter  $\Theta_m$  is fixed with  $\Theta_m=1$  in its objective function. For a fair comparison, three signal views features are combined for four single-view classification methods in full views. We can see that, on the one hand, both

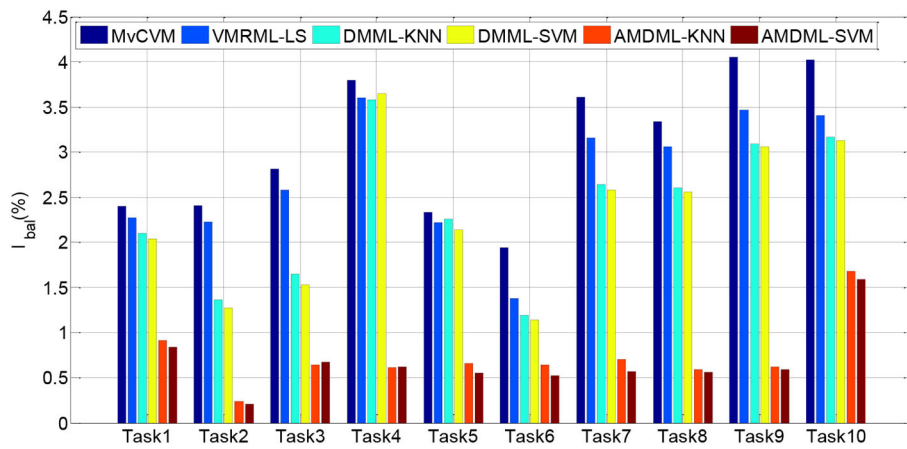
AMDML-KNN and AMDML-SVM using full-view features are better than them using only single-view features. For example, the performances of AMDML-KNN with full-view feature are 1.44, 1.57, and 1.31% higher than its performance in WPD, STFT, and KPCA on Task 1, respectively. On the other hand, the classification accuracy of methods AMDML-KNN and AMDML-SVM are better than single-view methods on 10 tasks. These results show that (1) the simple combination of features is limited to improve classification performance for single-view methods, and (2) due to the inherent diversity and complex of EEG signals, it is suitable to exploit multiple view features to better make use of the correlation and complementary EEG data. Thus, the multi-view learning framework can promote the EEG signal classification performance.

## Performance Comparisons With Multi-View Methods

In this subsection, we compare AMDML with several multi-view classification methods. The multi-view metric learning method DMML uses KNN and SVM as testing classifiers, and two classifiers are named as DMML-KNN and DMML-SVM, respectively. Figure 2 shows the classification accuracies of all methods on all EEG classification tasks. In addition, we use balanced loss  $l_{bal}$  (Wang et al., 2014; Gu et al., 2020) to



**FIGURE 2** | Classification accuracies of six methods on 10 EEG classification tasks.



**FIGURE 3** |  $l_{bal}$  performance of six methods on all EEG classification tasks.

evaluate the classification accuracy on positive class  $ACC_{positive}$  and negative class  $ACC_{negative}$ :

$$l_{bal} = 1 - (ACC_{positive} + ACC_{negative})/2$$

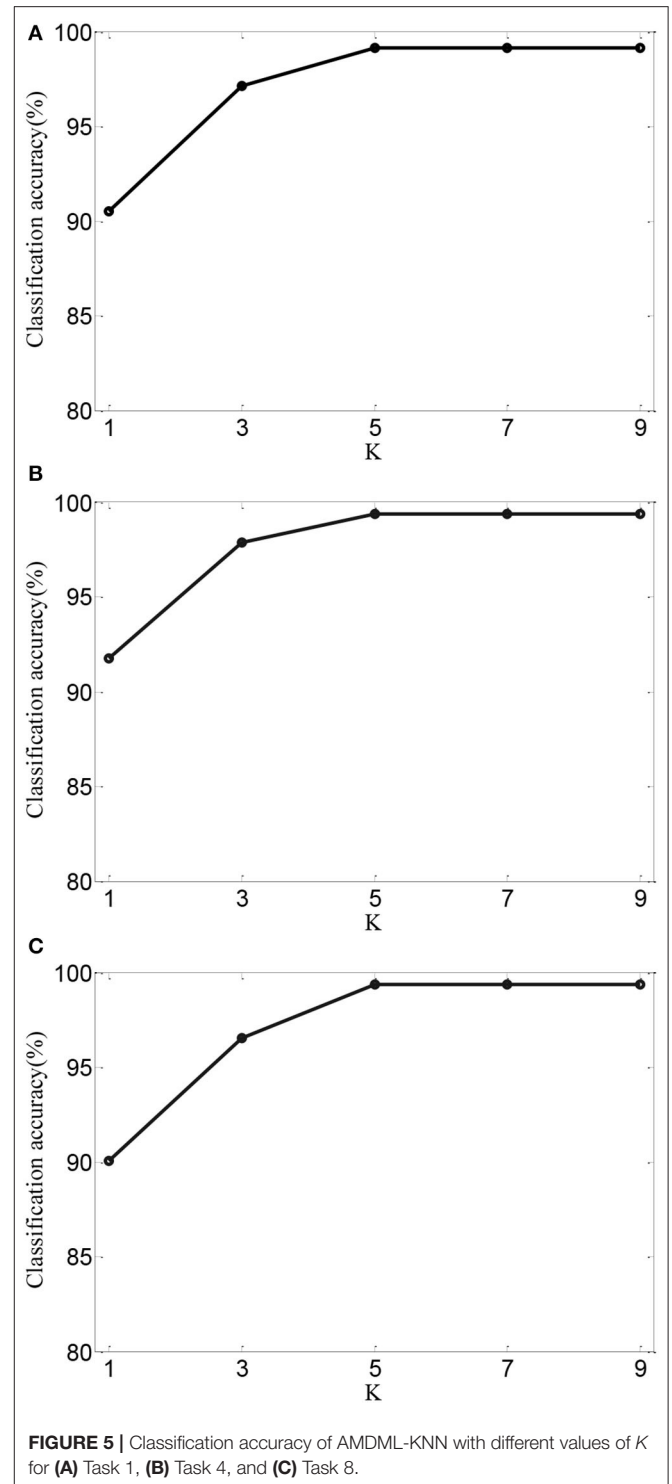
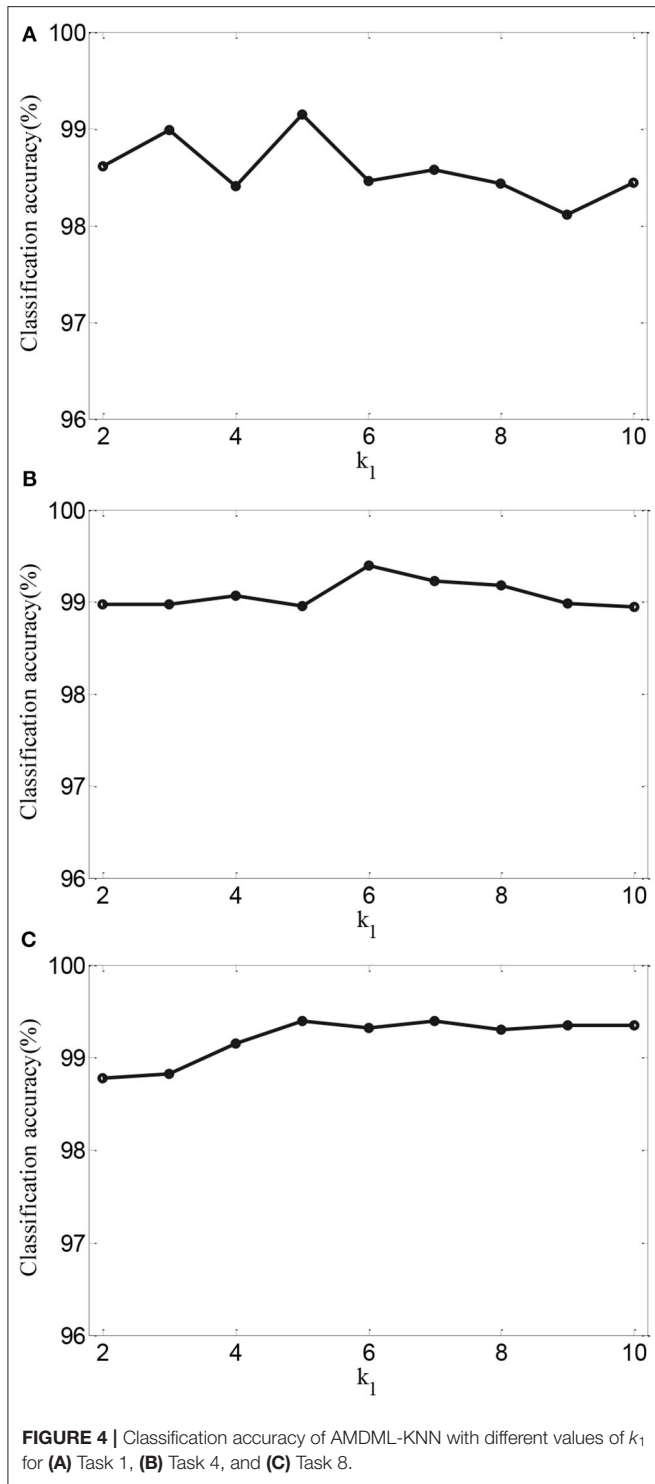
**Figure 3** shows  $l_{bal}$  performance of all methods on EEG classification tasks. Experiment results show that compared with all multi-view classification methods, both AMDML-KNN and AMDML-SVM have a positive effect on improving classification performance. AMDML-KNN and AMDML-SVM achieve the satisfactory classification performance in almost all of the EEG signal categories. In the framework of multi-view learning and to discriminate each emotion category best from all other categories, AMDML learns discriminative metric space to utilize the global and local information by adopting Fisher discriminative constraint and global structure constraint. Thus, the intraclass compactness and interclass EEG signals separability can perform better in the learned metric space. In addition, the

auto-weighted learning strategy used in the proposed method adjusts constraint and view weights. The optimal weights can be obtained adaptively, and multiple feature representation in each view can be collaborative learned. Similar to the results shown in **Table 3**, the classification accuracies of AMDML-KNN and AMDML-SVM are comparative. To summarize, the results in **Figures 2, 3** confirm that the AMDML method is effective in EEG signal classification.

### Model Analysis

To further validate the effects of performance of AMDML, we discuss the effect of the  $k_1$  in **Equation (5)** and  $k_2$  in **Equation (6)** in AMDML. The parameters  $k_1$  and  $k_2$  build the  $k$ -nearest neighbor inter- and intraclass graphs, respectively. For convenience, we set  $k_1 = k_2$  in the range  $\{2, \dots, 10\}$ . **Figure 4** shows the classification accuracy of AMDML-KNN with different values of  $k_1$  for Tasks 1, 4, and 8; meanwhile, the  $k$ -nearest neighbor





in KNN are fixed with 7. We can see that the performance of AMDML-KNN is not high sensitive to the variation  $k_1$  and  $k_2$ .

Next, for AMDML-KNN, we discuss the effect of the  $K$  in KNN classifier. In KNN classifier, the class label of the testing sample is determined by the distance from the  $K$  nearest

training sample. **Figure 5** shows the classification performance of different values of  $K$  for all tasks; meanwhile, fixing  $k_1 = k_2 = 5$ . We can see that the classification accuracy of AMDML-KNN is relatively stable with respect to the variation  $K$ . Therefore, we can set  $K$  empirically to 7 for Tasks 1, 4, and 8.

## CONCLUSION

In this paper, we propose a new multi-views metric learning to achieve the robust distance metric for EEG signal classification. In the scheme of the multi-view data representation, the diversity, and complementariness of features of all views can be exploited; meanwhile, both the Fisher discriminative constraint and global structure constraint are considered, and the learned classifier will obtain high generalization ability. Through learning a discriminative metric space, AMDML shows the higher classification performance. There are several directions of future study. In this paper, we use the  $k$ -nearest neighbor intra- and interclass graphs to exploit local discriminative information; we will consider other discriminative terms in the multi-view metric learning framework. Second, the gradient descent method used in this study is a simple and common solution method. We may develop a more effective method to speed up the solution of our method. Third, we plan to apply the proposed method for more EEG signal classification applications.

## REFERENCES

- Alwasiti, H., Yusoff, M. Z., and Raza, K. (2020). Motor imagery classification for brain computer interface using deep metric learning. *IEEE Access* 8, 109949–109963. doi: 10.1109/ACCESS.2020.3002459
- Cai, D., Liu, K., and Su, F. (2015). "Local metric learning for EEG-based personal identification," in *Proceedings of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (Brisbane, QLD), 842–846. doi: 10.1109/ICASSP.2015.7178088
- Chang, C. C., and Lin, C. J. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intelligent Intellig. Syst. Technol.* 2, 1–27. doi: 10.1145/1961189.1961199
- Cury, C., Maurel, P., Gribonval, R., and Barillot, C. (2019). A sparse EEG-informed fMRI model for hybrid EEG-fMRI neurofeedback prediction. *Front. Neurosci.* 13:1451. doi: 10.3389/fnins.2019.01451
- Griffin, D., and Lim, J. S. (1984). Signal estimation from modified short-time fourier transform. *IEEE Trans. Acoustics Speech Signal Process.* 32, 236–243. doi: 10.1109/TASSP.1984.1164317
- Gu, X., Chung, K., and Wang, S. (2020). Extreme vector machine for fast training on large data. *Int. J. Mach. Learn. Cybernet.* 11, 33–53. doi: 10.1007/s13042-019-00936-3
- Gummadavelli, A., Zaveri, H. P., Spencer, D. D., and Gerrard, J. L. (2018). Expanding brain-computer interfaces for controlling epilepsy networks: novel thalamic responsive neurostimulation in refractory epilepsy. *Front. Neurosci.* 12:474. doi: 10.3389/fnins.2018.00474
- Huang, C., Chung, F. L., and Wang, S. (2015). Multi-view L2-svm and its multi-view core vector machine. *Neural Netw.* 75, 110–125. doi: 10.1016/j.neunet.2015.12.004
- Iscan, Z., Dokur, Z., and Demiralp, T. (2011). Classification of electroencephalogram signals with combined time and frequency features. *Expert Syst. Appl.* 38, 10499–10505. doi: 10.1016/j.eswa.2011.02.110
- Jiang, Y., Deng, Z., Chung, F., Wang, G., Qian, P., Choi, K. S., et al. (2017b). Recognition of epileptic EEG signals using a novel multiview TSK fuzzy system. *IEEE Trans. Fuzzy Syst.* 25, 3–20. doi: 10.1109/TFUZZ.2016.2637405
- Jiang, Y., Wu, D., Deng, Z., Qian, P., Wang, J., Wang, G., et al. (2017a). Seizure classification from EEG signals using transfer learning, semi-supervised learning and TSK fuzzy system. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 2270–2284. doi: 10.1109/TNSRE.2017.2748388
- Juan, D. M. V., Gregor, S., Kristl, V., Pieter, M., German, C., and Dominguez. (2017). Improved localization of seizure onset zones using spatiotemporal constraints and time-varying source connectivity. *Front. Neurosci.* 11:156. doi: 10.3389/fnins.2017.00156
- Kabir, E., and Zhang, Y. (2016). Epileptic seizure detection from EEG signals using logistic model trees. *Brain Inform.* 3, 93–100. doi: 10.1007/s40708-015-0030-2
- Li, X., Yang, H., Yan, J., Wang, X., Li, X., and Yuan, Y. (2019). Low-intensity pulsed ultrasound stimulation modulates the nonlinear dynamics of local field potentials in temporal lobe epilepsy. *Front. Neurosci.* 13:287. doi: 10.3389/fnins.2019.00287
- Liu, Q., Zhao, X., and Hou, Z. (2014). "Metric learning for event-related potential component classification in EEG signals," in *Proceedings of 2014 22nd European Signal Processing Conference (EUSIPCO)* (Lisbon), 2005–2009.
- Lu, X., Wang, Y., Zhou, X., and Ling, Z. (2016). A method for metric learning with multiple-kernel embedding. *Neural Process. Lett.* 43, 905–921. doi: 10.1007/s11063-015-9444-3
- Ni, T., Ding, Z., Chen, F., and Wang, H. (2018). Relative distance metric learning based on clustering centralization and projection vectors learning for person Re-identification. *IEEE Access* 6, 11405–11411. doi: 10.1109/ACCESS.2018.2795020
- Phan, H., Do, Q., Do, T. L., and Vu, D. L. (2013). "Metric learning for automatic sleep stage classification," in *Proceedings of 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Osaka), 5025–5028. doi: 10.1109/EMBC.2013.6610677
- Quang, M. H., Bazzani, L., and Murino, V. (2013). "A unifying framework for vector-valued manifold regularization and multi-view learning," in *Proceedings of the 30th International Conference on International Conference on Machine Learning* (Atlanta, GA), 100–108.
- Richhariya, B., and Tanveer, M. (2018). EEG signal classification using universum support vector machine. *Expert Syst. Appl.* 106, 169–182. doi: 10.1016/j.eswa.2018.03.053
- Smola, A. J. (1997). "Kernel principal component analysis," in *Proceedings of International Conference on Artificial Neural Networks* (Lausanne), 583–588. doi: 10.1007/BFb0020217
- Spyrou, L., Kouchaki, S., and Sanei, S. (2018). Multiview classification and dimensionality reduction of scalp and intracranial EEG data through tensor factorisation. *J. Signal Process. Syst.* 90, 273–284. doi: 10.1007/s11265-016-1164-z
- Srinivasan, V., Eswaran, C., and Sridraam, N. (2005). Artificial neural network based epileptic detection using time-domain and frequency-domain features. *J. Med. Syst.* 29, 647–660. doi: 10.1007/s10916-005-6133-1

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://epileptologie-bonn.de/cms/upload/workgroup/lehnertz/eegdata.html>.

## AUTHOR CONTRIBUTIONS

JX and TN conceived and developed the theoretical framework of the manuscript. XG performed the data process, data evaluation, analysis, interpretation, and designed the figures. All authors drafted the manuscript.

## FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61806026, by the Natural Science Foundation of Jiangsu Province under Grant No. BK20180956.

- Tzallas, A. T., Tsipouras, M. G., and Fotiadis, I. D. (2009). Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE Trans. Inform. Technol. Biomed.* 13, 703–710. doi: 10.1109/TITB.2009.2017939
- Usman, S. M., and Fong, S. (2017). Epileptic seizures prediction using machine learning methods. *Computat. Math. Methods Med.* 2017:9074759. doi: 10.1155/2017/9074759
- Wang, F., Zuo, W., Zhang, L., Meng, D., and Zhang, D. (2015). A kernel classification framework for metric learning, *IEEE Trans. Neural Netw. Learn. Syst.* 26, 1950–1962. doi: 10.1109/TNNLS.2014.2361142
- Wang, S., Wang, J., and Chung, K. (2014). Kernel density estimation, kernel methods, and fast learning in large data sets. *IEEE Trans. Cybernet.* 44, 1–20. doi: 10.1109/TSMCB.2012.2236828
- Weinberger, K. Q., and Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. *Mach. Learn. Res.* 10, 207–244.
- Wen, D., Li, P., Zhou, Y., Sun, Y., Xu, J., Liu, Y., et al. (2020). Feature classification method of resting-state EEG signals from amnesic mild cognitive impairment with type 2 diabetes mellitus based on multi-view convolutional neural network. *IEEE Trans. Neural Syst. Rehabil. Eng.* 28, 1702–1709. doi: 10.1109/TNSRE.2020.3004462
- Wen, J., Tian, Z., She, H., and Yan, W. (2010). “Feature extraction of hyperspectral images based on preserving neighborhood discriminant embedding,” in *Proceedings of IEEE Conference: Image Analysis and Signal Processing (IASP)* (Xiamen), 257–262.
- Wu, T., Yan, G. Z., Yang, B. H., and Hong, S. (2008). EEG feature extraction based on wavelet packet decomposition for brain computer interface. *Measurement* 41, 618–625. doi: 10.1016/j.measurement.2007.07.007
- Xing, E. P., Jordan, M. I., Russell, S. J., and Ng, A. Y. (2003). “Distance metric learning with application to clustering with side-information,” in *Advances in Neural Information Processing Systems*, (Cambridge, MA, USA: MIT Press), 521–528.
- Yuan, Y., Xun, G., Jia, K., and Zhang, A. (2019). A multi-view deep learning framework for EEG seizure detection. *IEEE J. Biomed. Health Inform.* 23, 83–94. doi: 10.1109/JBHI.2018.2871678
- Zhang, L., Shum, H. P. H., Liu, L., Guo, G., and Shao, L. (2019). Multiview discriminative marginal metric learning for makeup face verification. *Neurocomputing* 333, 339–350. doi: 10.1016/j.neucom.2018.12.003
- Zhou, Z., Zhang, Y., and Jiang, Y. (2019). “Deep view-reduction TSK fuzzy system: a case study on epileptic EEG signals detection,” in *Proceedings of 2019 IEEE Symposium Series on Computational Intelligence (SSCI)* (Xiamen), 387–392. doi: 10.1109/SSCI44817.2019.9002722

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Xue, Gu and Ni. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.