IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 14, 2020, accepted July 21, 2020, date of publication July 27, 2020, date of current version August 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3011969

MLP With Riemannian Covariance for Motor Imagery Based EEG Analysis

PENGPENG YANG[1,2], JING WANG[1,2], HONGLING ZHAO[3], AND RUNZHI LI[2], (Member, IEEE)

[1]School of Information Engineering, Zhengzhou University, Zhengzhou 450000, China
[2]Cooperative Innovation Center of Internet Healthcare, Zhengzhou University, Zhengzhou 450000, China
[3]School of Distance Learning, Zhengzhou University, Zhengzhou 450000, China

Corresponding author: Runzhi Li (rzli@ha.edu.cn)

This work was supported in part by the China Education and Research Network (CERNET) Innovation Project under Grant NGII20180708; and in part by the Program of Scientific and Technological Research of Henan Province under Contract 192102310215.

ABSTRACT Stroke is one of the leading causes of disability and incidence. For the treatment of prognosis of stroke patients, Motor imagery (MI) as a novel experimental paradigm, clinically it is effective because MI based Brain-Computer interface system can promote rehabilitation of stroke patients. There is being a hot and challenging topic to recognize multi-class motor imagery action classification accurately based on electroencephalograph (EEG) signals. In this work, we propose a novel framework named MRC-MLP. Multiple Riemannian covariance is used for EEG feature extraction. We make a multi-scale spectral division to filter EEG signals. They consist of different frequency bandwidths name sub-band. We concatenate and vectorize features extracted by Riemannian covariance on each sub-band. We design a fully connected MLP model with an improved loss function for motor imagery EEG classification. Furthermore, our proposed method MRC-MLP outperforms state-of-the-art methods and achieves approximately mean accuracy with 76%.

INDEX TERMS EEG, motor imagery, multi-layer perceptron, Riemannian covariance.

I. INTRODUCTION

Stroke is one of the leading causes of high incidence. It has a high rate of mortality and disability in the world [1]. Patients suffering from stroke have different degrees of limb dysfunction. There will be a neurological repair effect on stroke patients for long-term rehabilitation therapy [2]. Clinically, it is one of the main objectives for the treatment of stroke by rehabilitation training that needs help of the professional doctors. They will come to being burdens for physicians to make different plans for different stroke patients. Meanwhile they take plenty of medical resource. In recent years, with the development of artificial intelligence (AI), human-computer interaction (HCI) technology, brain-computer interface (BCI) technology, people try to design some intelligent rehabilitation training system that can reduce medical expenses and save cost of labor for physicians by providing one type of active rehabilitation.

In active rehabilitation training, BCI constructs a direct information transfer pathway, which converts brain activity signals into commands to external devices instead of traditional brain and muscle control pathways [3]. Brain activity signals consist of electroencephalogram (EEG), electrocorticogram (ECOG), near-infrared reflectance spectroscopy (NIRS), and functional magnetic resonance imaging (fMRI) monitored clinically. EEG signals are commonly applied in BCI systems due to the collection of non-intrusion and easy access. According to brain activity signals of interest and modes of operations based on the control paradigms, there are various types of EEG-based BCI systems such as visualize P300 [3], steady-state visual evoked potential (SSVEP) [4], and motor imagery (MI) [5]. Motor imagery is a cognitive process in which the object imagines the action rather than performing the actual action, e.g., right or left hand. Based on MI, we collect EEG signals that reflect the above actions. In MI-BCI system, it is important and difficult to recognize actions by analysis of EEG signals. It enables communication between the postoperative stroke and the control of a prosthesis [6], [8].

Recently, MI is being one of the most important research domains on BCI. It emerges many methods proposed to recognize actions of motor imagery. One is multi-layer

The associate editor coordinating the review of this manuscript and approving it for publication was Khalid Aamir.

perceptron (MLP) network that has achieved better results on EEG signals classification [7], [23]. On binary motor imagery EEG classification, researchers have achieved a high accuracy. But there is still a great challenge on multi-class motor imagery classification based on EEG signals. In this work, we aim to propose a hybrid framework to classify actions of MI based on EEG signals. We compose Riemannian covariance feature extraction method with a multi-layer perceptron (MLP) model. In the experiments, we conduct extensive experiments on the BCI competition IV dataset 2a. The experimental results demonstrate that the proposed method outperforms others on accuracy.

The main contributions of this work can be described as follows:

- We propose a novel connected way for the multi-layer perceptron model, in which we design an improved loss function.
- We conducted corresponding experiments to verify the effectiveness of the model from three aspects: different loss function, different scale spectral division, and different feature extractors.

The rest of this work is organized as follows: We summarize and analyze the previous works of motor imagery classification in section 2. Section 3 describes the materials, method of Riemannian covariance feature, the proposed multi-layer perceptron network structure and the improved loss function. In Section 4, we give the experimental setup and the analysis of the experimental results. Conclusions are presented in section 5.

II. RELATED WORK

Previous works have proposed many methods on MI-based EEG tasks. They mainly include feature extraction and classification methods. For feature extraction methods, they aim to extract important features of MI from EEG signals. However, accurate classification of EEG is difficult due to its low signal-to-noise ratio, nonlinearity, weak intensity, and time-variant behaviour. It is a challenge to extract common features to get good classification results. Currently, most works mainly focus on three feature extraction methods for EEG signals of MI. One is common spatial pattern method (CSP). It is earliest introduced to the field of EEG signal analysis by Koles *et al.* [9]. For CSP, it finds a set of spatial filters, by which the variance of one type of signals is maximized and the other is minimized. Many variants of CSP had been devised to extract features from EEG signals. Novi *et al.* [10] proposed SBCSP that decompose the EEG signals into sub-bands by a filter bank named FBCSP. It is an improved variant of SBCSP by selecting corresponding of CSP features on each frequency bands automatically [11]. Jiao *et al.* [12] introduced a novel SGRM algorithm to exploit inter-subject information for constructing an efficient sparse group representation after CSP feature extraction. The second method is power spectral density (PSD). It extracts EEG signals by using time-frequency analysis methods. Meng *et al.* [14] used PSD to reveal the effect of caffeine and sugar intake on BCI online performance and resting brain signals. The common time-frequency analysis is short-time Fourier transform (STFT), Hilbert Huang transform (HHT), and wavelet transform (WT) [15], etc. STFT divides signals into some segments with an equal window length. And then it carries out Fourier transform (FT) on each segment. HHT consists of empirical mode decomposition (EMD) and Hilbert transform (HT). Compared with STFT, WT uses a variable window width and a wavelet function to conduct time-frequency analysis. It could obtain better time and frequency resolution than STFT. The last method is Riemannian covariance. It extracts features by constructing a covariance matrix based on Riemannian geometry [16]. Many studies used Riemannian covariance to extract the features on EEG signals and obtained good experimental results [17], [18]. In the part of classification, plenty of traditional machine learning methods were proposed to recognize MI on features from EEG signals. They include support vector machine (SVM) [19], Random Forest [20], Bayesian classifier, linear discriminant analysis (LDA) [23], etc. They were simple and useful, but there is still being a challenge to improve the classification accuracy. Because they highly depend on the quality of features extraction engineering. In recent years, with the development of deep learning, researchers have applied deep learning methods to classify MI action based on EEG signals [21]. One method is firstly to convert MI based EEG signals into time-frequency maps, followed by convolutional neural networks (CNN) that are used for further spatial feature extractions and classifications [24], [30]. Lawhern *et al.* [22] introduced a compact convolutional neural network named EEGNet that is a robust model for a range of BCI tasks. According to the temporal characteristics of EEG signals, researchers employed long short term memory (LSTM) models to classify EEG data [25]. Arce *et al.* [26] adopted two deep neural network models with different number of layers to recognize motor imagery EEG signals. They achieved an average accuracy of 85% on the 2-class classification. SBELM combined the advantages of ELM and sparse Bayesian learning to automatically control model complexity and eliminate redundant hidden neurons for motor imagery EEG classification [13].Though there are plenty of deep learning models proposed to classify EEG tasks. It is still a challenge to improve the model performance. In this work, we present a novel method MRC-MLP for MI action recognition based on EEG signals.

III. MATERIALS AND METHODS

In this work, we focus on multi-class classification problem of MI actions based on EEG signals. We propose a hybrid framework MRC-MLP shown in Figure 1. Firstly, we preprocess the multi-channel EEG signals. It is Multiple Riemannian Covariance (MRC) that represents multiple Riemannian covariance extractor with different frequency sub-bands used in the stage of feature extraction. It computes covariance matrix between channels to extract features. There are three steps for MRC. First, we make a multi-scale division on

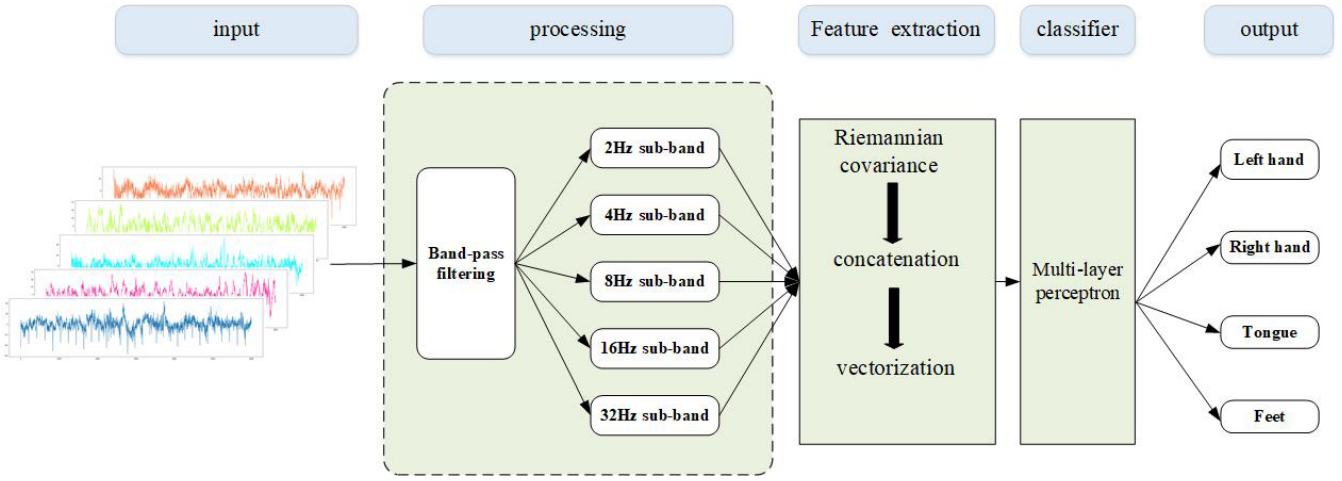

FIGURE 1. The overview of our proposed framework MRC-MLP for motor imagery classification.

spectral. Then we use the Riemannian covariance method to extract features for each frequency sub-bands. Third, we vectorize and concatenate sub-band features as the input of the classification model. Next, we construct a multi-layer perceptron model to classify MI EEG signals.

A. DATASETS

In the experiments, datasets are from an open access dataset BCI competition IV dataset 2a [27]. It contains nine healthy subjects who are required to complete four different motor imagery actions. They consist of both feet, tongue, left hand, and right hand respectively. Each subject completes two sessions on different days, one is for training and the other is for testing. Every session includes 288 trials with the same amount of each MI task. We give a trial of the experimental deployment paradigm in Figure 2. In the beginning, there is a fixation cross on the black screen after the beep. Then from 2s-3.25s, a cue in the form of the arrow pointing either to the left, right, down and up appeared, prompting the subject to perform the desired motor imagery. The subjects are requested to carry out the motor imagery task until the fixation cross disappeared from the screen at 6s. Through the MI experiments, we collect signals that consist of 22 EEG channels and 3 EOG channels by sample frequency of 250 Hz. We use between 0.5Hz and 100Hz in the band, moreover, a notch filtering of 50Hz was enabled. In this work, we extract EEG data from 2.5-6s to make full use of relevant data. We utilize 22 EEG channels and exclude 9.41% of the trials in BCI competition IV dataset 2a like other works that caused by some artifacts based on EOG signals.

B. MULTI-SCALE SPECTRAL DIVISION

The amplitude of EEG signals in the cortical area changes regularly for one person that is a state of motor imagery. This phenomenon is known as event-related desynchronization (ERD) and event-related synchronization (ERS) [19]. For example, when one imagines a movement of left hand,

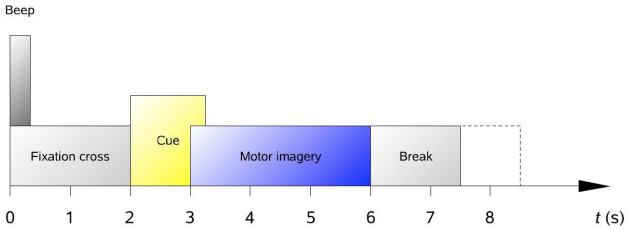

FIGURE 2. Time scheme of the paradigm from [27].

the amplitude of right cortical area decreases. Meanwhile, left cortical area increases. ERD and ERS are mainly proved in sensory-motor rhythms mu (8-13Hz) and beta (15-30Hz). However, the frequency band of ERD/ERS is extremely subject-dependent. It is hard to find the most discriminative features only in mu and beta rhythms. It is necessary to extend the frequency band of EEG signals and divide it into multi-scale spectral parts before the stage of extracting features [34]. The details of each sub-band after multi-scale spectral segmentation are shown in Table 1.

TABLE 1. Details about multi-scale spectral division.

Bandwidth/Hz	Number	Slide	Slide window/Hz
2	18	No	0
4	9	No	0
8	8	Yes	4
16	6	Yes	4
32	2	Yes	4

In this work, we choose 5 different types of frequency bandwidth (2Hz, 4Hz, 8Hz, 16Hz, 32Hz) for multi-scale spectral segmentation. Fifth-order Butterworth bandpass filter is used to set the wider range 4-40Hz. Then, we make

multi-scale spectral divisions and obtain different sub-bands. On 4-40Hz spectral, the number of 2Hz and 4Hz frequency sub-bands are 18 and 9 respectively. It is no doubt that redundant information will be introduced through further sliding windows. We do not use a slide window on 2Hz and 4Hz frequency sub-bands. In contrast, the number of 8Hz, 16Hz, 32Hz frequency sub-bands on 4-40Hz spectral is small. We set a 4Hz width slide window, and the effective features under the wide sub-bands can be further explored through overlapping spectral bands. Finally, we get 43 different frequency bandwidths on 4-40Hz spectral. Though multi-scale spectral division of EEG signals takes redundancy, it can capture the dynamic features of the EEG signals.

C. RIEMANNIAN COVARIANCE FEATURE EXTRACTION

Riemannian covariance method is based on Riemannian geometry [28], [29]. It describes the procedure of MI-EEG feature extraction as follows. Firstly, we calculate the spatial covariance matrix (SCM) for each trial of motor imagery EEG. Formula 1 gives the computation procedure.

$$P_i = \frac{x_i x_i^T}{t-1} \tag{1}$$

$x_i \in R^{N\times t}$ denote a selective segment of the i_{th} continuous EEG signals where n represents the number of electrodes and t represents the number of sampled points for the segment. Superscript T represents matrix transposition.

The spatial covariance matrix of P sampled two-dimensional data satisfies symmetry positive definite (SPD). The geodesic distance between two SPD matrices P_1 and P_2 in Riemannian space is defined as

$$\delta_R(P_1, P_2) = ||\log(P_1^{-1}P_2)||_F = [\sum_{i=1}^{n} \log^2 \lambda_i]^{1/2} \tag{2}$$

where λ_i represents the real eigenvalues of $P_1^{-1}P_2$.

Next, we need to obtain the mean matrix in Riemannian space. According the Riemannian framework, the Riemannian mean of $m > 1$ SPD matrix is defined as,

$$\theta(P_1, \ldots, P_m) = \arg\min_{P_{mean}} \sum_{i=1}^{m} \delta_R^2(P_{mean}, P_i) \tag{3}$$

After calculating the Riemannian mean matrix, we can obtain a new feature matrix M

$$M = \log(P_{mean}^{-1/2} P_i P_{mean}^{-1/2}) \tag{4}$$

At the last, we make a vectorization of feature matrix M as follows:

$$\vec{M} := vect(M) = \left[M_{1,1}; \sqrt{2}M_{1,2}; \ldots M_{n,n}\right] \in R^{(n+1)n/2} \tag{5}$$

$M_{n,n}$ represents an element in the nth row and nth column of the matrix M. Because covariance matrices M are symmetric, the off-diagonal elements are multiplied by $\sqrt{2}$ to preserve the norm.

D. MULTILAYER PERCEPTRON CLASSIFICATION MODELS

Multilayer perceptron (MLP) network is one of the most common feed-forward artificial neural networks, which is inspired by the way the human brain processes information. The MLP consists of input layer, hidden layers, and output layer, and each hidden layer is fully-connected to the next layer. We employ the hidden layer to achieve nonlinear mapping on the input space and take the output layer to obtain the classification result. The output of each neuron in the hidden layer is the weighted summation that input signals

multiply them by their respective connection weights after the operation of an activation function. It is described as follows

$$y_j = f(\sum w_{ij} x_i + b_i) \tag{6}$$

where f is an activation function, b_i denotes a bias term, and w_{ij} represents the connection weight from the i^{th} neuron of the previous layer to the i^{th} neuron of the layer at hand.

To achieve better classification results, we get inspiration from the encoding part of autoencoder (AE) model, we present a stacked layers MLP model and more details about the network architecture are shown in Figure 3, in which the number of neurons in each layer decrease by degrees. That is, MLP goes deeper, the number of neurons in the hidden layer is decreased. The motivation for this design is that the encoding part of the AE model can capture significant feature information by using few neurons, which not only can reduce the dimension of input vectors, but also reduce feature redundancy. Compared to other deep learning methods, it can obtain abstract features and realize accurate classification of high dimensional features extracted by Riemannian covariance through simple operations between layers.

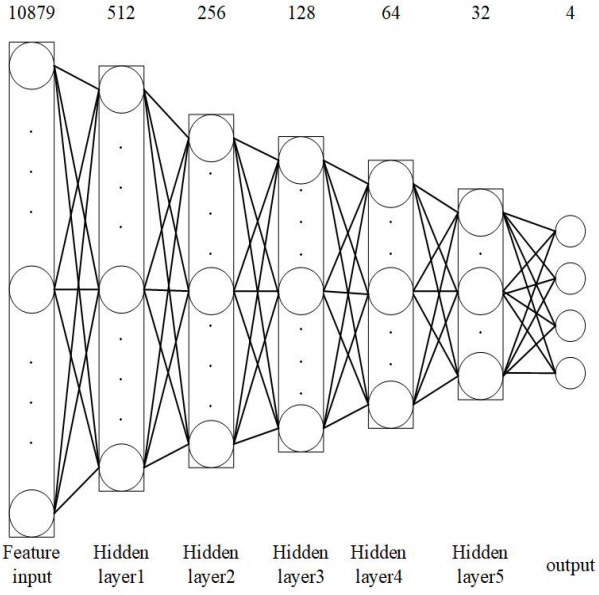

FIGURE 3. The proposed MLP network. At the top is the number of neurons in each layer, and the bottom is the type of each layer.

For the MLP model, the length of input is 10879 for the input layer. It is calculated as follows:

$$l_{input} = Num_{sub} \times \frac{n\,(n+1)}{2} \tag{7}$$

where Num_{sub} is the number of frequency sub-bands. n represents the number of EEG channels. The number of neurons is set to 512 in the first hidden layer, and the number of neurons is reduced by factor of two in the following four hidden layers. The last layer is classification layer, and SoftMax function is used as classifier in this work. As for activation functions, we use the most applied type which is a hyperbolic tangent function in past studies. It is defined as follows:

$$\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \tag{8}$$

Another difficulty is to update weights of network after the model of MLP has been determined. In the training phase, the purpose of changing the weight is to reduce the error between predicted labels and the true labels. In order to accelerate model convergence and reduce training time, we use Adam optimizer to update weights. Furthermore, the loss function has great effects on classification result performance of the proposed MLP. The traditional cross entropy loss cannot handle hard and imbalance samples. To address above problem, we present an improved loss function in this work and it will be introduced in next section.

E. IMPROVED LOSS FUNCTION

Loss function is very important to optimize models. It is known that the classic focal loss function can resolve effectively class imbalance problem. In this work, we design an improved loss function to alleviate the over-fitting and improve the accuracy based on focal loss [31]. Firstly, we describe the definition of focal for binary classification loss as formula 9.

$$FL = \sum_{t=0,1} -\alpha_t (1-p_t)^{\gamma} \log(p_t) \tag{9}$$

For a binary classification, where p_t is the estimated probability, and γ represents focusing parameter. $(1-p_t)^{\gamma}$ is called modulating factor, which is influenced by setting different focusing parameter. Coefficient α_t can be used to control the weight of positive and negative samples for the total loss. The motivation of focal loss is to increase the proportion of predicted error sample value and decrease the proportion of predicted correct sample value in total loss.

For improved loss function, we add l_2 regularization based on focal loss function. In this work, the categories of the motor imagery EEG data set are balanced, we set the class weight coefficient α_t to 1. The improved loss function for multi-class classification is presented as formula 10.

$$loss = \sum_{i=0}^{3} -(1-\hat{p}_i)^{\gamma} \log \hat{p}_i + \lambda \|W\|^2 \tag{10}$$

$\hat{p}_t$ denotes the estimated probability for class i which is computed by SoftMax function. γ is the focusing parameter. $\|W\|^2$ is the l_2 regularization of weight parameters for classification models. λ is the regularization coefficient, which is a trade-off between training errors and regularization term. In our proposed improve loss function, we use l_2 regularization on weight parameters to increase the stability of the proposed model and improve the overall classification accuracy. It is an implementation on structural risk minimization instead of experiential risk minimization. Experiments show that it can resolve the small-sample on EEG signal classification effectively.

IV. EXPERIMENTS

A. EXPERIMENT SETUP

We designed three sets of experiments to validate the model. The purpose of the first group of experiments is to explore the classification effect of motor imagery under different feature extractors. Previous studies have shown that CSP and Riemannian covariance feature extractors are widely used. We combined multiple sub-bands filtered EEG signals and use 7-layer MLP model with improved loss function to conduct experiments.

For the second group of experiments, we experiment 5 different types of frequency bandwidth (2Hz, 4Hz, 8Hz, 16Hz, 32Hz) for multi-scale spectral segmentation and combined multiple sub-bands. To verify the classification accuracy under combined multi-scale spectral division is better than each sub-band. We explore the classification performance with the same feature extractor Riemannian covariance and 7-layer MLP model with improved loss function.

In the last group of experiments, we aim to verify the validity of the improved loss function. We experimented on common loss functions for comparison based on the same feature extractor method and the same MLP parameters. In addition, we also compare our method with other literatures.

In this paper, all experiments were carried out using TensorFlow platform and python. The codes run on an Intel Core i5-7500 CPU, 8GB RAM and Windows environment. In the experiments, we use a 5-fold cross-validation to choose the optimal hyper-parameters. Such as epochs is 200 and learning rate is 0.001 and λ is 0.0005. In addition, according to the design principle introduced in the classifier section, we tuned the multi-layer perceptron models with different hidden layers from random subjects. Based on the trade-off between the accuracy and complexity of the model, we adopted an MLP model with 5 hidden layers. The mean accuracy of nine subjects in the MLP model with different hidden layers are shown in Figure 5.

In this work, we use the classification accuracy as a metric. The classification accuracy of each subject is calculated by the number of mis classifications divided by the number of total test trials. To visualize the performance of our prediction model, we give the confusion matrix of each subject in Figure 4. Each entry in a confusion matrix represents the

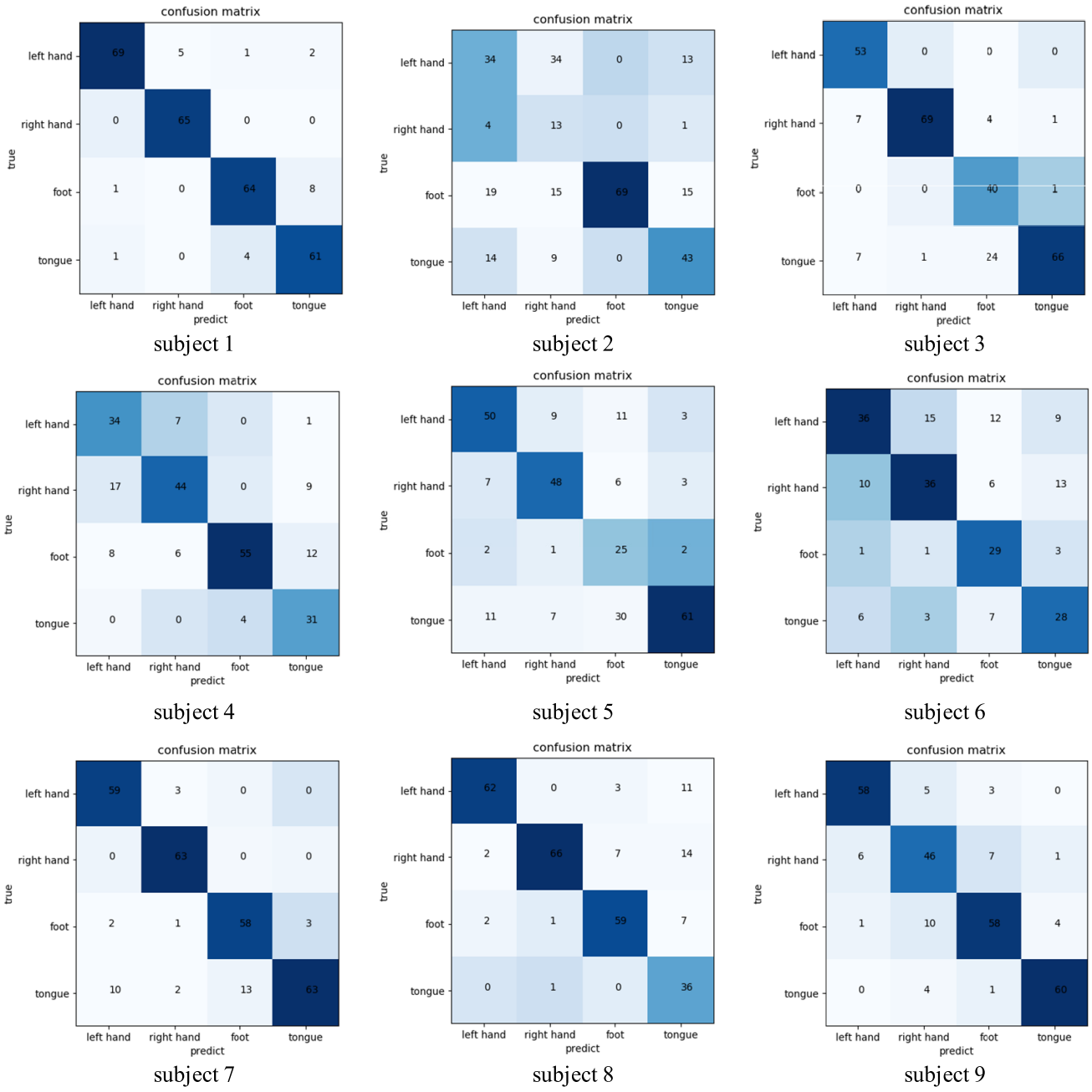

FIGURE 4. Confusion matrices of classification accuracy for the proposed MRC-MLP model.

number of predictions that were made by the model where it classified the classes correctly or incorrectly.

B. RESULTS AND ANALYSIS

1) COMPARISONS ON FEATURE EXTRACTION

The results are shown in Figure 6. Although the CSP method has an increase of 5%-6% accuracy over the Riemannian covariance on subject 3 and subject 7, it is obvious that the accuracy of the latter is significantly higher than the former on other subjects. The mean accuracy of the CSP feature extraction is 68.18%, 7% lower than the Riemannian covariance feature extraction. In conclusion, the Riemannian covariance feature extractor is more effective than CSP based on the multi-scale spectral division of MI EEG.

2) SELECTION ON FREQUENCY OF SUB-BANDS

It can be seen from the result in Table 2 that the sub-band of 4Hz achieves the best mean accuracy of 73.43% than other sub-bands. From Table 3, the accuracy of the concatenated sub-bands is 76.05%. Accordingly, the concatenated features after multi-scale spectral division can improve the recognition of MI EEG. The width of the frequency sub-band

TABLE 2. The performance of different frequency sub-bands.

Subjects	2Hz (%)	4Hz (%)	8Hz (%)	16Hz (%)	32Hz (%)
1	90.03	88.96	87.90	88.25	81.13
2	54.41	55.47	55.83	55.47	48.40
3	76.92	79.12	81.31	83.51	82.78
4	71.05	68.85	60.08	53.07	46.92
5	46.30	64.85	59.78	39.49	29.71
6	53.48	55.81	58.13	57.20	53.48
7	83.03	85.19	84.11	84.11	78.33
8	79.33	81.18	80.81	79.70	78.22
9	81.43	81.43	79.54	78.78	75.37
mean	72.89	73.43	71.95	68.85	63.82

TABLE 3. Accuracy of the proposed loss function with common loss functions.

Subjects	FL (%)	CE (%)	MSE (%)	CE_l2(%)	MSE_l2(%)	Our proposed (%)
1	88.25	87.90	86.47	87.90	87.61	92.17
2	56.89	38.86	46.28	55.47	52.71	56.18
3	82.41	74.72	75.45	74.35	78.12	83.51
4	57.85	55.26	39.47	71.92	72.24	71.92
5	59.05	36.95	27.53	63.04	63.13	66.66
6	59.06	42.32	46.97	54.41	52.28	60.00
7	76.53	70.03	74.00	78.70	84.49	87.72
8	81.18	58.67	69.74	83.39	80.18	82.28
9	78.40	68.56	70.07	80.68	81.95	84.07
mean	71.08	59.25	59.55	72.21	72.25	76.05

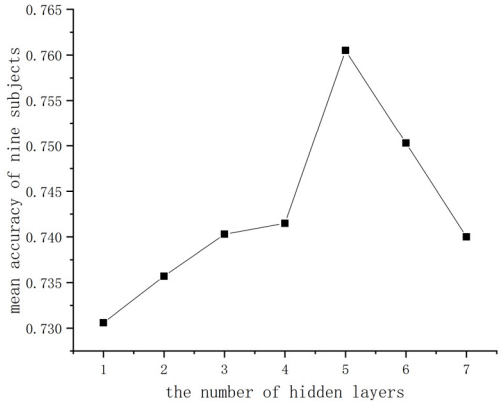

FIGURE 5. Mean accuracy results of different hidden layers in MLP.

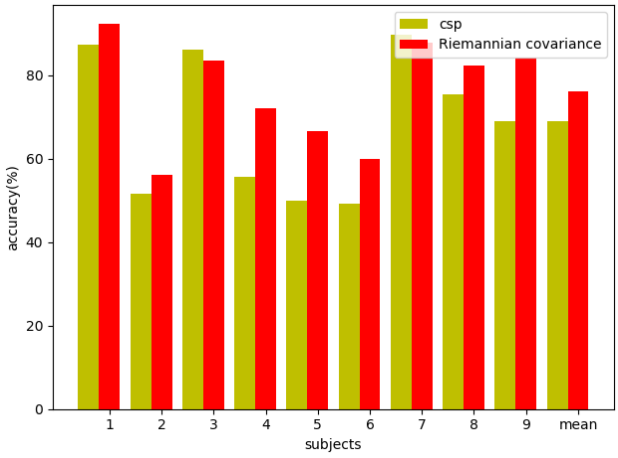

FIGURE 6. Accuracy results of two common feature extractors.

increases from 4Hz, it can be concluded that the mean accuracy of classification decreases gradually. But the sub-bands of 8Hz, 16Hz, 32Hz achieve better accuracy on subject 3.

3) PERFORMANCE OF DIFFERENT LOSS FUNCTIONS

Table 3 gives the classification results from different loss functions such as mean squared error (MSE) [32], cross entropy (CE) [33], focal loss (FL), mean squared error with l_2 regularization (MSE_l2), and cross entropy with l_2 regularization (CE_l2). It is observed that the improved loss function is superior to others. From Table 3, our proposed loss function achieves a mean accuracy of 76.05% for the nine subjects. On the one hand, compared with common loss functions such as CE and MSE without l_2 regularization, the corresponding mean accuracy increases by about 16%. It can be concluded from the experimental results, the improved loss function with l_2 regularization can help to enhance the classification performance of the motor imagery classification task. On the other hand, the FL without l_2 regularization achieves 71.08%, which indicates that the FL is potent on the recognition of MI EEG features.

4) EVALUATION ON MODEL PERFORMANCE

To prove the performance of the proposed method MRC-MLP, we provided a comparison among our proposed method and other existing methods. Table 4 shows the mean accuracies of different methods based on the BCI competition IV dataset 2a. Comparing the different methods, on the one hand, it can be seen from the second column of the Table 4

TABLE 4. Mean accuracy comparisons with other published results.

References	Feature extractor	Classifier	Mean accuracy
H. Yang et al. [35]	CSP	CNN	69%
A. Barachant et al. [34]	Riemannian covariance	LDA	70%
M. Hersche et al. [36]	CSP	SVM	73.70%
	Riemannian covariance	SVM	74.77%
Our proposed	CSP	MLP	68%
	Riemannian covariance	MLP	76%

that the Riemannian covariance matrix is a more effective feature extraction method for motor imagery EEG, achieving a significant improvement in accuracy relative to other feature extraction methods. On the other hand, from the last five rows of the Table 4 it can be concluded that the classifier proposed in this paper is superior to the traditional SVM and LDA algorithm. From the Table 4, it could be seen that our proposed model gets better results than others. However, our model requires more time and memory than SVM. The number of parameters used by our model is 5.7M. Our model consumes three times as much time as the SVM classifier by experiments. In addition, we also conducted experiments on the feature of combining multi-spectral and multi-temporal, and obtained a mean accuracy of 73%. It is not feasible for MLP because multi-temporal features bring dimensional disaster.

V. CONCLUSION AND FUTURE WORK

In this work we combine Riemannian covariance feature extractor method based on different frequency sub-bands and multi-layer perceptron model to classify motor imagery EEG signals. According to the method adopted in each step of motor imagery recognition process, we conduct corresponding experiments to verify our methods. In terms of feature extraction method, we compare the results of CSP and Riemannian covariance, and find that the latter is more effective. For the multi-scale spectral division, we compare the results of each frequency sub-band, and the experiment results demonstrate that combining the features of each sub-band is useful to improve the accuracy. Compared with other common loss functions, we propose an improved loss function which achieves convincing classification results. To our knowledge, our method achieves state-of-the-art classification results compared with the existing methods.

Our MRC-MLP model should be further explored considering the long training time and much parameters. And the future work will more focus on the feature selection and channel selection methods. Specifically, in order to improve the computational efficiency, we can select features or channels by using genetic algorithm to obtain more efficient features and reduce the dimension of input feature vector.

REFERENCES

[1] V. L. Feigin, B. Norrving, and G. A. Mensah, "Global burden of stroke," *Circulat. Res.*, vol. 120, no. 3, pp. 439–448, 2017.

[2] V. Kehayas and A. Holtmaat, "Rejuvenating brain plasticity," *Science*, vol. 356, no. 6345, pp. 1335–1336, Jun. 2017.

[3] E. Donchin, K. M. Spencer, and R. Wijesinghe, "The mental prosthesis: Assessing the speed of a P300-based brain-computer interface," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 174–179, Jun. 2000.

[4] G. R. Muller-Putz and G. Pfurtscheller, "Control of an electrical prosthesis with an SSVEP-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 361–364, Jan. 2008.

[5] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123–1134, Jul. 2001.

[6] T. M. Vaughan, W. Heetderks, L. Trejo, W. Rymer, M. Weinrich, M. Moore, A. Kübler, B. Dobkin, N. Birbaumer, and E. Donchin, "Brain-computer interface technology: A review of the second international meeting," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 94–109, Jun. 2003.

[7] L. G. Hernández and J. M. Antelis, "A comparison of deep neural network algorithms for recognition of EEG motor imagery signals," in *Pattern Recognition*. Cham, Switzerland: Springer, 2018, pp. 126–134.

[8] D. Cheng, Y. Liu, and L. Zhang, "Exploring motor imagery eeg patterns for stroke patients with deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 2561–2565.

[9] Z. J. Koles, M. S. Lazar, and S. Z. Zhou, "Spatial patterns underlying population differences in the background EEG," *Brain Topogr.*, vol. 2, no. 4, pp. 275–284, 1990.

[10] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band common spatial pattern (SBCSP) for brain-computer interface," in *Proc. 3rd Int. IEEE/EMBS Conf. Neural Eng.*, May 2007, pp. 204–207.

[11] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw., IEEE World Congr. Comput. Intell.*, Jun. 2008, pp. 2390–2397.

[12] Y. Jiao, Y. Zhang, X. Chen, E. Yin, J. Jin, X. Wang, and A. Cichocki, "Sparse group representation model for motor imagery EEG classification," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 631–641, Mar. 2019.

[13] Z. Jin, G. Zhou, D. Gao, and Y. Zhang, "EEG classification using sparse Bayesian extreme learning machine for brain–computer interface," *Neural Comput. Appl.*, vol. 32, pp. 6601–6609, Oct. 2018.

[14] J. Meng, J. H. Mundahl, T. D. Streitz, K. Maile, N. S. Gulachek, J. He, and B. He, "Effects of soft drinks on resting state EEG and brain–computer interface performance," *IEEE Access*, vol. 5, pp. 18756–18764, 2017.

[15] J. Liang, S. R. Chaudhuri, and M. Shinozuka, "Simulation of nonstationary stochastic processes by spectral representation," *J. Eng. Mech.*, vol. 133, no. 6, pp. 616–627, Jun. 2007.

[16] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 26, no. 3, pp. 735–747, Jan. 2005.

[17] S. Guan, K. Zhao, and S. Yang, "Motor imagery EEG classification based on decision tree framework and Riemannian geometry," *Comput. Intell. Neurosci.*, vol. 2019, pp. 1–13, Jan. 2019.

[18] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry," *Expert Syst. Appl.*, vol. 95, pp. 201–211, Apr. 2018.

[19] S. Chen, Z. Luo, and H. Gan, "An entropy fusion method for feature extraction of EEG," *Neural Comput. Appl.*, vol. 29, no. 10, pp. 857–863, May 2018.

[20] R. Zhang, X. Xiao, Z. Liu, W. Jiang, J. Li, Y. Cao, J. Ren, D. Jiang, and L. Cui, "A new motor imagery EEG classification method FB-TRCSP+RF based on CSP and random forest," *IEEE Access*, vol. 6, pp. 44944–44950, 2018.

[21] X. Zhang, L. Yao, X. Wang, J. Monaghan, D. Mcalpine, and Y. Zhang, "A survey on deep learning based brain computer interface: Recent advances and new frontiers," 2019, *arXiv:1905.04149*. [Online]. Available: http://arxiv.org/abs/1905.04149

[22] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.

[23] A. J. Villar, "Comparative study of robust methods for motor imagery classification based on CSP and LDA," in *Proc. 7th Latin Amer. Congr. Biomed. Eng. (CLAIB)*. Bucaramanga, Colombia: Springer, 2017, pp. 126–129.

[24] Z. Pang, J. Li, Y. Sun, H. Ji, L. Wang, and R. Lu, "Classifying motor imagery EEG signals using the deep residual network," in *Proc. Int. Symp. Comput. Sci. Comput.* Cham, Switzerland: Springer, 2018, pp. 64–68.
[25] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "LSTM-based EEG classification in motor imagery tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 11, pp. 2086–2095, Nov. 2018.
[26] F. Arce, E. Zamora, G. Hernández, J. M. Antelis, and H. Sossa, "Recognizing motor imagery tasks using deep multi-layer perceptrons," in *Machine Learning and Data Mining in Pattern Recognition*, vol. 10935, P. Perner, Ed. Cham, Switzerland: Springer, 2018, pp. 468–482.
[27] C. Brunner, R. Leeb, G. R. Muller-Putz, and A. Schlogl. *BCI Competition 2008—Graz Data Set A*. Accessed: 2015. [Online]. Available: http://bnci-horizon-2020.eu/database/data-sets
[28] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *Neurocomputing*, vol. 112, pp. 172–178, Jul. 2013.
[29] F. Yger, M. Berar, and F. Lotte, "Riemannian approaches in brain-computer interfaces: A review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1753–1762, Oct. 2017.
[30] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals," *J. Neural Eng.*, vol. 14, no. 1, Feb. 2017, Art. no. 016003.
[31] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
[32] O. Köksoy, "Multiresponse robust design: Mean square error (MSE) criterion," *Appl. Math. Comput.*, vol. 175, no. 2, pp. 1716–1729, Apr. 2006.
[33] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 19–67, Feb. 2005.
[34] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 920–928, Apr. 2012.
[35] H. Yang, S. Sakhavi, K. K. Ang, and C. Guan, "On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2620–2623.
[36] M. Hersche, T. Rellstab, P. D. Schiavone, L. Cavigelli, L. Benini, and A. Rahimi, "Fast and accurate multiclass inference for MI-BCIs using large multiscale temporal and spectral features," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 1690–1694.

PENGPENG YANG received the B.S. degree in information engineering from Sichuan Normal University, Chengdu, China, in 2015. He is currently pursuing the M.S. degree in computer science and technology with Zhengzhou University. His primary research interests include machine learning and medical data analysis.

JING WANG received the B.S. degree in information engineering from Henan Normal University, Xinxiang, China, in 2015, and the M.S. degree from Zhengzhou University, Zhengzhou, China, in 2018, where she is currently pursuing the Ph.D. degree in software engineering. Her research interests include image processing and data mining.

HONGLING ZHAO received the B.S. and M.S. degrees from Zhengzhou University, Zhengzhou, China, in 2001 and 2014, respectively. He is currently a Lecturer with the School of Distance Learning, Zhengzhou University. His research interests include big data, machine learning, and virtual reality.

RUNZHI LI (Member, IEEE) received the B.S. degree from the Shenyang University of Technology, Shenyang, China, in 2000, and the M.S. and Ph.D. degrees from Zhengzhou University, Zhengzhou, China, in 2003 and 2011, respectively. She is currently an Associate Professor with the Cooperative Innovation Center of Internet Healthcare, Zhengzhou University. Her research interests include medical data analysis and the prediction of diseases.