

Classification of EEG Motion Artifact Signals Using Spatial ICA



Hsin-Hsiung Huang, Aubrey Condor, and Helen J. Huang

Abstract Using electroencephalography (EEG) data to extract information about cortical signals has become an increasingly explored task of interest in the field of computational neuroscience. In this paper, we proposed a novel procedure which reduce dimension by applying spatial Independent Component Analysis (SICA) on EEG motion artifact data and classify gait speed for a given subject by the projected EEG motion artifact signals. Whereas most applications of ICA in analyzing EEG data employ temporal ICA, we use SICA and Principal Component Analysis for dimension reduction before applying classifiers such as Support Vector Machines, Naive Bayes, and multinomial logistic regression to the extracted independent components. We evaluate and compare the classification models by using randomly selected channels from the multi-channel EEG motion artifact data as our test data. For practical application and interpretation, we treat the test channels as if they might come from a new trial for the given subject.

Keywords Classification · Brain signals · Time series · High-dimensional · Spatial dimension reduction

1 Introduction

Electroencephalography (EEG) has become more readily available as a method to analyze cortical activity due to its relatively low cost and high temporal resolution [10]. Although it can be convenient to collect EEG data, the dimensions of the

H.-H. Huang (✉) · A. Condor
Department of Statistics and Data Science, University of Central Florida, Orlando, FL, USA
e-mail: hsin.huang@ucf.edu; acondor99@Knights.ucf.edu

H. J. Huang
Department of Mechanical and Aerospace Engineering, University of Central Florida, Orlando, FL, USA
e-mail: hjhuang@ucf.edu

recorded electrical activity can become large very quickly, especially when using high-density EEG for trials recorded at lengthy time periods. Since EEG datasets are high-dimensional and noisy, thus it is difficult to use them for classifying subjects in terms of their movement characteristics. Therefore, it is important to develop methods for analyzing the EEG data with a reduced dimensional space where the variable information is still retained for classifying different gait movements for a given subject.

The purpose of this paper is to introduce a novel procedure which could classify movement characteristics with EEG motion artifact data projected by spatial Independent Component Analysis (SICA) for a given subject. The EEG motion artifact data did not actually contain electrophysiological signals from the body. Instead, the EEG motion artifact data consisted of signals that were recorded from an isolated conductive reference cap using an EEG system [17]. This EEG motion artifact recording method has enabled the development of new artifact rejection techniques to clean EEG signals [22] but these EEG motion artifact data could also potentially be used to classify movement characteristics.

ICA originates from a method to solve problems such as the “cocktail party” problem where one hopes to identify individual voices when many people are speaking simultaneously, recorded in multiple devices in different locations. In this case, the temporal ICA (TICA) algorithm assumes independence in time such that the original voices can be extracted from the mixtures [5]. Similar to the cocktail party problem, each electrode in EEG data is composed of the mixture of multiple electrophysiological signals that includes the true underlying cortical signals which are assumed to be temporally independent [6]. Therefore, TICA has been most commonly used to analyze EEG data [9]. Because TICA is most commonly used to analyze EEG data, the authors who recorded the first set of EEG motion artifact data in [17] applied TICA and source localization to the data and found that the independent components were mostly outside of the brain volume, which provided evidence that TICA could be used to partially distinguish motion artifact from electrophysiological signals [25]. However, the authors did not attempt to extract movement characteristics from the EEG motion artifact data, which inspired us to investigate that possibility in this paper.

In contrast to TICA, SICA assumes spatially independent components and has been used more commonly in literature which focuses on the analysis of functional magnetic resonance imaging (fMRI) data [7]. However, the temporal dimension of EEG data can be much larger than that of the spatial dimension in many cases [23]. Therefore, our aim is to reduce the temporal dimension and employ SICA on the reduced data instead of using TICA. Consequently, we propose a method which iteratively computes SICA on subsets of the partitioned data, and concatenates the independent components from each iteration.

The rest of this paper is organized as follows. We introduce the techniques of ICA, and then describe the examined data and the proposed data analysis procedure for dimension reduction. Furthermore, the classification methods: k -nearest neighbors, Support Vector Machines, Naive Bayes, and multinomial logistic regression are described. Finally, we discuss the results of our classification and further suggest the implications of our conclusions.

2 Classification Methods

We use the EEG motion artifact data projected by Spatial ICA with the following classifiers. (1) k -nearest neighbor (k -nn), (2) Support Vector Machines, (3) Naive Bayes, and (4) multinomial logistic regression. A brief description of each method is introduced as follows. Notice that X is the matrix of predictors and Y is the response vector of classification labels in this section.

2.1 k -Nearest Neighbor (k -nn)

We used the k -nearest neighbour classification method with $k = 3$, which determines the class by majority voting of each point's k nearest neighbors. The knn was fulfilled by knn3() function in package 'caret' in R were used in this study.

2.2 Support Vector Machines

In an approach to solve multi-class pattern recognition, we can consider the problem as many binary classification problems [8, 27]. If we consider the case of K classes, K classifiers are constructed. Each classifier builds a hyperplane between itself and the $K - 1$ other classes [27]. If our response, or the two classes, are represented by $Y \in \{-1, 1\}$, we can use a Support Vector Machine (SVM) to construct a hyperplane to separate the two groups such that the distance between the hyperplane and the nearest point, or the margin, is maximized [8].

The optimization problem seeks to minimize: $\phi(w, \xi) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^n \xi_i$ with constraints $y_i((w \cdot x_i) + b) \geq 1 - \xi_i, i = 1, \dots, n$ and $\xi_i \geq 0, i = 1, \dots, n$. We can solve the optimization problem by solving the dual problem, which consists of minimizing $W(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n y_i y_j \alpha_i \alpha_j K(x_i, x_j)$ under the constraints $0 \leq \alpha_i \leq C, i = 1, \dots, n$ and $\sum_{i=1}^n \alpha_i y_i = 0$. The above gives the decision function: $f(x) = \text{sign} \left[\sum_{i=1}^n (\alpha_i y_i K(x, x_i)) + b \right]$ [27].

There exist many different types of kernel functions to use in Support Vector Machine classification. Support vector machines (SVM) [8] with a linear kernel was used in this study. The cost parameters C was tuned using cross validation [16] with package 'e1071' in R [21].

2.3 Naive Bayes

The Naive Bayes algorithm is a classification algorithm based on Bayes' rule and a set of conditional independence assumptions. Given the goal of learning $P(Y|X)$ where $X = X_1, \dots, X_p$, the Naive Bayes algorithm makes the assumption that each

feature X_i is conditionally independent of each of the other X_j s given $Y = k$, and also independent of each subset of the other X_j s given $Y = k$ [15]. Bayes' rule states that the probability of some observed data, $x = (x_1, \dots, x_p)$, belonging to class k is $P(Y = k|x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^K \pi_l f_l(x)}$, where $P(Y = k) = \pi_k$ and $f_k(x) = p(X = x|Y = k)$ is the probability density for X in class k . For a given class k , Naive Bayes classification makes the assumption that all of the features, or x_i s are independent, or $f_k(x) = P(x_1, x_2, \dots, x_p|Y = k) = \prod_{i=1}^p P(x_i|Y = k)$. Thus, the Naive Bayes classifier is as follows: $\arg \max_k \pi_k \prod_{i=1}^p f_k^i(x_i)$, where $f_k^i(x_i) = P(X_i = x_i|Y = k)$ [15]. We believe that, because we are reducing our data by Spatial ICA, Naive Bayes classification will provide the best results in terms of misclassification rate. We will use all other classification methods for a comparison to Naive Bayes.

2.4 Multinomial Logistic Regression

Logistic regression can be used as a method for modeling a categorical response variable by finding significant parameters. In the multinomial case, our y response variable represents more than two categories. It does not require the assumption of statistical independence of predictors unlike the Naive Bayes classifier, but assumes collinearity between predictors. We used package 'glmnet' in R [11] to fulfill the multinomial logistic regression.

3 The EEG Motion Artifact Signals Data and Spatial ICA Methodology

We used the EEG motion artifact signals data that was collected in [17] and were analyzed using TICA in [25]. The method of recording isolated motion artifact in EEG is described in detail in [17]. Briefly, the isolated motion artifact data were collected using a 256 channel EEG system (ActiveTwo, Biosemi) from ten young and healthy participants. A non-conductive silicone swim cap was placed on each subject's head to block true electrophysiological signals. A simulated scalp consisting of a short wig soaked in conductive gel was placed over the silicone layer, and the EEG cap and electrodes were placed over the simulated scalp. Subjects sat (0 m/s) and walked at four different speeds (0.4 m/s, 0.8 m/s, 1.2 m/s, 1.6 m/s) on a treadmill. Each trial was 10 min in duration, and data were recorded at 512 Hz [25]. Ten subjects with complete data sets were used in analysis. In this study, we pre-processed data by vertically concatenating each of the subject's five speeds into one data file while creating a speed label for each signal. In terms of the temporal dimension, each recording consisted of 300,000–310,000 points for the 10 min of recorded signal. However, we used only points 1 through 300,000 for consistency within and between the ten subjects. Therefore, for one of the ten subjects, we have

total 1280 EEG motion artifact signals with dimension $p = 300,000$ in five speeds and the sample size $n = 256$ given a speed.

Independent Component Analysis (ICA) belongs to a class of methods often referred to as “Blind Source Separation” which aim to extract certain quantities from a mixture of other quantities [26]. ICA-unlike other statistical methods of dimension reduction that find mutually de-correlated signals such as Principle Component Analysis (PCA) or Factor Analysis (FA)- is based on the assumption of statistical independence [26]. ICA decomposes the data such that we are left with maximally independent signals by maximizing non-Gaussianity. One important distinction of ICA is that there is no order or ranking of the extracted components. It is also notable that the components do not recognize the difference of signs [18]. Since the EEG signals in our dataset have heavy-tailed and multimodal distributions, it is inadequate to apply PCA, which can not recover statistically independent source signals [13, 14].

Let us denote the observed data as an n by p matrix, X where n the number of spatial voxels and p represents the number of time points. In Spatial ICA (SICA), we consider the n vectors containing each of the p instances to be our signals [3]. We can represent the SICA decomposition as follows. Assuming that X is a mixture signals matrix from sources matrix S , and let $r =$ the number of components, A is $n \times r$ and S is $r \times p$, and then $X = AS + E$, where E is defined using the smallest $(n - r)$ principal components (PCs). $S = WKX$ is the estimated $n \times m$ matrix source matrix, W is the estimated $m \times m$ un-mixing matrix, K is the estimated $p \times m$ pre-whitening matrix projecting data onto the first m principal components, where n is the number of observations and m is the number of independent components [3, 14]. In this study, the ordering of independent components is determined by Principal Component Analysis.

Before ICA is performed, it is necessary to first pre-process the data with reduction and whitening. For the purposes of data compression, SICA presumes that there are fewer independent sources than there are time points [7]. Reduction is first performed by PCA and the specified number of components are retained such that the maximum amount of variation is represented. ICA combined with PCA allows both whitening and achieving dimension reduction [23].

ICA supposes that the underlying sources are each not normally distributed. It follows that, sources can be extracted by making them as non-Gaussian as possible with the measure of negentropy. Given a covariance matrix, the distribution that has the highest entropy is the Gaussian distribution [4, 18]. Negentropy or differential entropy is a measure of deviation from normality expressed as

$$N(Z) = (EG(Z) - EG(Z_{Gaussian}))^2$$

where Z is an arbitrary multivariate random variable and $Z_{Gaussian}$ is a multivariate Gaussian random variable of the same covariance matrix as Z , and the contrast function $G(u) = -\exp(-u^2/2)$ was used in our analysis [18]. There are several different algorithms that employ methods to estimate the independent components. The FastICA algorithm maximizes negentropy $N(X)$ [23]. We use the FastICA

algorithm [13, 14] for SICA because it has been shown to outperform most other ICA algorithms in speed of convergence [23].

4 Data Analysis Procedure

The challenge of the examining EEG signals data is its high dimension. Since the dimension comes from the multiple time moments which records the EEG signals of the individuals walking on a treadmill at a given speed, we assume that the high dimensional ($p > n$) EEG motion artifact signals used in this study can be decomposed as $m < n$ independent components. Note that SICA cannot perform dimension reduction directly. Therefore, PCA is applied as a pre-process to determine the ordering of the importance of the components by the magnitude of the eigenvalues of the correlation matrix of X . To reduce the high dimensional signals, we transform the signals by using only the first four (which is $K - 1$, K is the number of categories) independent components. We applied package FastICA in software R [20] for our data analysis. For each of the ten subjects, we split our data into training and testing subsets, such that we will use the training set to train our classification model, and the testing set to see how our model performs for new observations. We randomly select 256 channels out of the 1280 concatenated channels as our test set, and use the remaining 1024 channels as our training set. Although the channels for a given trial (or given speed) are receiving motion artifact signals simultaneously, we proceed in our analysis as if the selected test signals are recorded from another trial. The data analysis procedure is shown in the following algorithm. The proposed method is outlined, starting from the structure of the original data to the concatenated data, further into the training and testing split, and finishing with the SICA and PCA projected data.

For each of the ten subjects data, we first concatenate all records as a dataset with 1280 rows and 300,000 columns, and then partitioned the data as a training set (1024 records) and a test set (256 records) by random sampling. For each interval of 1000, we sampled a time point for both the training and test sets. The downsampling rate is equalled a duration of 1.95 s, since the original signals were collected by 512 Hz sampling rate (the unit of time is $1/512$ or 0.00195 s). Then we applied SICA to the downsampled training set using the FastICA algorithm on the training data, since the five categories can be represented in a 4-dimensional space, each signal is compressed to the first four independent components. We then project the test data set onto the space of the first four independent components obtained from the training data. The plots in Figs. 1 and 2 show that most of the four independent components for the test set each subject have obvious clusters corresponding to the walking-speed categories.

We used the k -nearest neighbors with $k = 3$, SVM, Naive Bayes, and multinomial logistic regression for classification modeling. The accuracy rate is computed as the number of correctly categorized signals, over the total number of classified signals. We use the accuracy rates and the multi-class area under the curve

Algorithm 1 Spatial ICA and classification of EEG motion artifacts signals

-
- 1: There are 10 subjects. Each subject has five EEG motion artifact signals datasets according to the five different walking speeds. Each dataset has 256 rows/space points (from 256 channels) and 300,000 columns/time points (recordings in sequential time points).
 - 2: **procedure**
 - 3: Concatenate all records of the five speeds EEG motion artifact datasets corresponding to the subject walking speeds.
 - 4: Downsample the time points by keeping the first sample and then every 1000th sample after the first.
 - 5: Partition the data as training (1024 records) X_{tr} and test sets (256 records) X_{te} by random sampling.
 - 6: Apply SICA to the training set and extract 4 independent components by fastICA with type ‘deflation’ and the exponential contrast function. The components are extracted one at a time.
 - 7: The SICA outputs a source matrix S_{tr} , pre-whitening matrix K , and un-mixing matrix W .
 - 8: Obtain the source matrix of the test set by projection $S_{te} = X_{te}KW$
 - 9: Build classification models on S_{tr} and evaluate by using S_{te}
 - 10: **end procedure**
 - 11: For comparisons, another analysis used randomly sampling 4 time points from X_{tr} and X_{te} , and build classification models on the training and evaluation on the test set.
-

(AUC) to evaluate the proposed method with comparisons of randomly selecting four time points. Next we trained our Naive Bayes model for classification. For each subject, we use the projected SICA test data to compare the classified results with the true classifications and output the accuracy rate and AUC. Package ‘naivebayes’ in R was used for the Naive Bayes Classification [19]. Finally, we trained our last classification model with the training data using multinomial logistic regression [28]. We do not evaluate individual parameters for significance but instead simply use the fitted model for prediction of the test data to obtain the accuracy rates for each subject. We also provide a and 3. We provide the Area Under the Curve (AUC) values as well to show classification performance. These values give the total area under the Receiver Operating Curve (AUC) values were calculated with package HandTill2001 in R [12]. Classification results are presented in Tables 1 and 2. Table 3 is the comparison of the proposed SICA method versus random sampling four time points on the simulation data, which was generated by adding noises into the signals of Subject 1. The noises were sampled from uniformly distributed random variables with the range of $(-a, a)$, where a is $\sqrt{3}$ signal-to-noise (SNR) ratio in order to make the simulated data have $SNR = 1$.

5 Classification Results

The aim of our study is to explore SICA as a method of dimension reduction for analyzing high dimensional EEG or EEG motion artifact datasets. We proposed an algorithm to downsample and perform SICA to the signals with a large number of time points such that sufficient information is still retained for classification. By using the first four independent components, the k -nn with $k = 3$, support vector

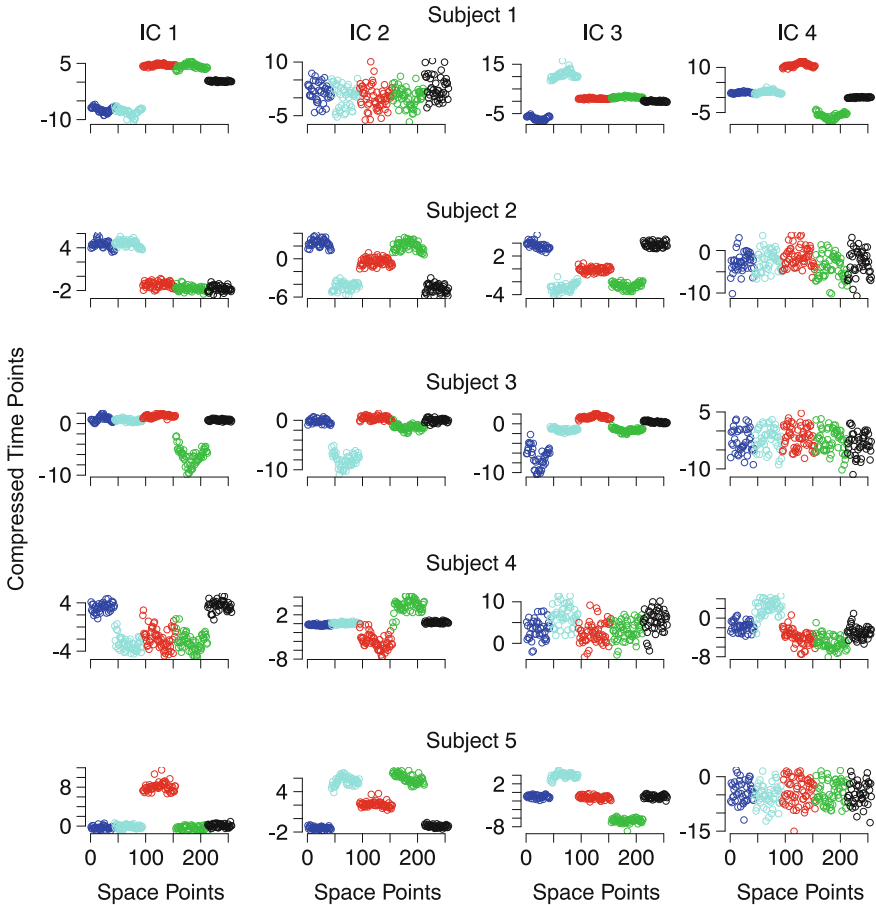


Fig. 1 The first four projected independent components of the test set of each of subjects 1–5. In each plot, the y-axis is the component and the x-axis represents the space points (electrode channels) of the EEG motion artifacts signals. There are total 256 space points and five speeds (five categories). On average, each category contains about 51 space points. For each subject, most of the first four independent components clusters according to the subject’s walking speed. The clusters are highlighted by different colored. The x-axis is the 256 space points in the test set, and the y-axis represents the values of the independent components, which are compressed time points

classifier with linear kernels, and multinomial logistic regression all successfully classify the EEG motion artifacts signals. The Naive Bayes method performed worse than the others. In contrast, the classification results are very poor when just using four randomly selected time points. The comparisons show that the proposed method effectively reduce the dimension of time with high classification accuracy. The scatter plots of the top four independent components (ICs) indicate that these ICs have different patterns with respect to their walking speeds (see Figs. 1 and 2).

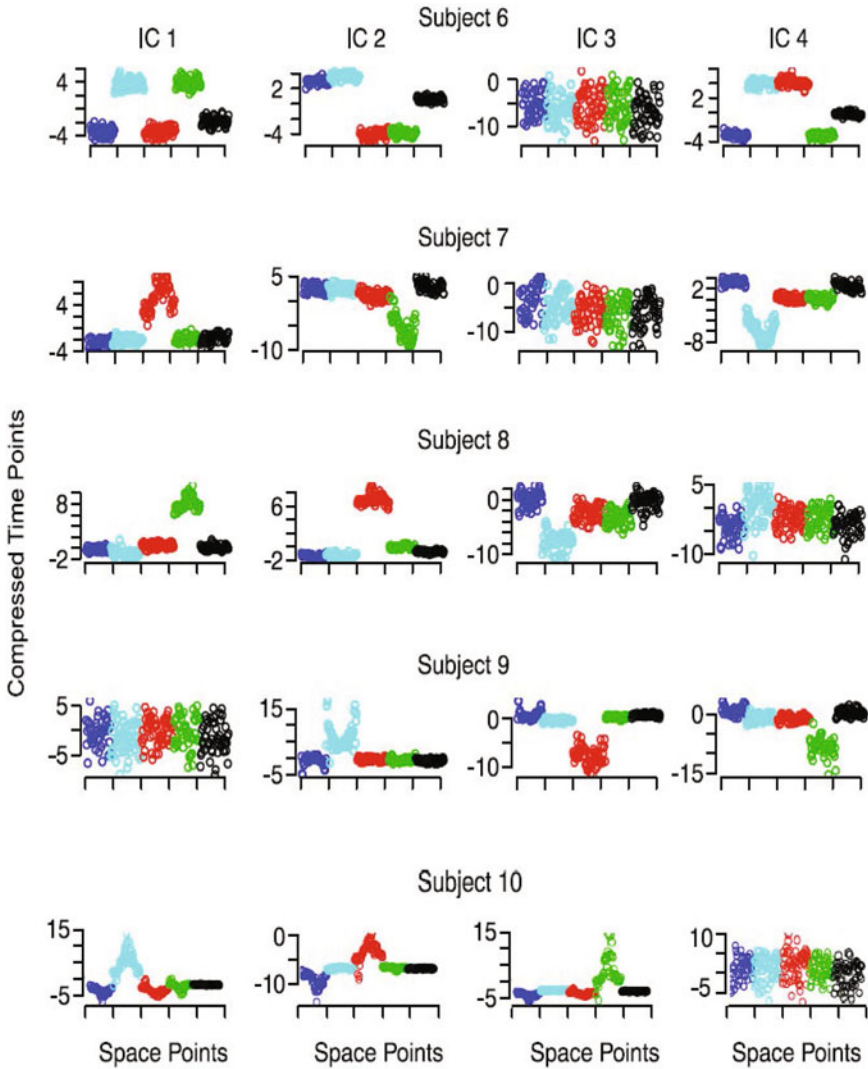


Fig. 2 The first four projected independent components of the test set of each of subjects 6–10. In each plot, the y -axis is the component and the x -axis represents the space points of the EEG motion artifact signals. There are total 256 space points and five speeds (five categories). On average, each category contains about 51 space points. For each subject, most of the first four independent components clusters according to the subject’s walking speed. The clusters are highlighted by different colored. The x -axis is the 256 space points in the test set, and the y -axis represents the values of the independent components

Table 1 Model comparisons in terms of accuracy rates

Methods	knn ($k = 3$)	SVM (linear kernel)	Naive Bayes	Multinomial logistic
<i>SICA with four components</i>				
Subject 1	0.9922	1.0000	0.8008	0.9961
Subject 2	0.9219	0.9759	0.8242	0.9492
Subject 3	0.8906	0.8789	0.7266	0.8945
Subject 4	0.7891	0.9609	0.8056	0.9967
Subject 5	0.8359	0.8906	0.8516	0.9180
Subject 6	0.8320	0.8320	0.7578	0.8320
Subject 7	0.9688	0.9141	0.9063	0.9531
Subject 8	0.8125	0.8594	0.7539	0.9219
Subject 9	0.7891	0.7852	0.8164	0.8359
Subject 10	0.8984	0.8867	0.8320	0.8984
<i>Random sampling four time points</i>				
Subject 1	0.1211	0.1367	0.1406	0.1445
Subject 2	0.1953	0.2148	0.2031	0.2188
Subject 3	0.1133	0.1055	0.1563	0.0938
Subject 4	0.2109	0.1445	0.1875	0.1758
Subject 5	0.3047	0.3594	0.3281	0.3359
Subject 6	0.2148	0.2305	0.2813	0.2500
Subject 7	0.3164	0.3047	0.2578	0.2969
Subject 8	0.2695	0.2422	0.2227	0.2383
Subject 9	0.2930	0.3008	0.2656	0.3125
Subject 10	0.1172	0.1055	0.1797	0.1055

6 Discussion

Before classification, the independent components data for each subject was of dimensions 256 by 4 whereas the original test set started as 256 by 300,000. Hence, the temporal dimension of our data was reduced by 75,000 folds. The four independent components successfully retained a sufficient amount of information about the EEG motion artifact signals in order to successfully classify a subject's walking speed. However, it is important to take caution in interpreting the independent components. With TICA, it can be assumed that the independent components represent the unmixed cortical signals. SICA has been commonly applied to functional MRI data, where time points correspond to input dimensions and voxels are samples. In contrast, TICA for EEG assumes that sensors constitute input dimensions and time-points are samples [1]. We used SICA for EEG motion artifact signals with the assumption that sensors constitute input samples and time points are dimensions. Consequently, the EEG motion artifact signals observed at different time points are assumed to be linear sums of the source signals and maximizes spatial sparsity alone [2].

Table 2 Model comparisons in terms of multi-class areas under the ROC (AUC)

Methods	knn ($k = 3$)	SVM (linear kernel)	Naive Bayes	Multinomial logistic
<i>SICA with four components</i>				
Subject 1	0.9977	1.0000	0.6174	0.9997
Subject 2	0.9631	1.0000	0.8128	1.0000
Subject 3	0.9472	0.9998	0.4480	1.0000
Subject 4	0.8861	0.9974	0.8220	0.9894
Subject 5	0.9644	0.9695	0.3556	0.9820
Subject 6	0.9251	0.9997	0.4075	1.0000
Subject 7	0.9914	0.9977	0.7357	0.9977
Subject 8	0.9462	0.9896	0.4486	0.9892
Subject 9	0.9381	0.9551	0.5486	0.9718
Subject 10	0.9501	0.9891	0.6709	0.9902
<i>Random sampling four time points</i>				
Subject 1	0.4575	0.5503	0.4841	0.5433
Subject 2	0.5315	0.5721	0.5243	0.5792
Subject 3	0.4249	0.3431	0.4778	0.3463
Subject 4	0.5166	0.5014	0.4483	0.5216
Subject 5	0.5968	0.6533	0.6764	0.6708
Subject 6	0.5016	0.4673	0.5232	0.4965
Subject 7	0.5898	0.5670	0.5453	0.5549
Subject 8	0.5502	0.5369	0.5843	0.5164
Subject 9	0.5713	0.5727	0.4970	0.5581
Subject 10	0.4419	0.3814	0.5056	0.3651

Table 3 Model comparisons of simulation data with 100 repetitions in terms of average accuracy rates and AUC

Methods	knn ($k = 3$)	SVM (linear kernel)	Naive Bayes	Multinomial logistic
<i>SICA with four components</i>				
Accuracy	0.9577(0.0081)	0.9547(0.0120)	0.8215(0.0360)	0.9286(0.0204)
AUC	0.9908(0.0025)	0.9990(0.0010)	0.4473(0.0586)	0.9999(0.0004)
<i>Random sampling four time points</i>				
Accuracy	0.2310(0.1100)	0.2126(0.1109)	0.0039(0.0387)	0.2205(0.1233)
AUC	0.5267(0.0891)	0.5342(0.1271)	0.0069(0.0069)	0.5382(0.1328)

Numbers in the parenthesis are the standard deviation

It is evident that for a given subject, we are able to successfully classify walking speed with EEG motion artifact signals. The classification results (Tables 1 and 2) show that except for the Naive Bayes classifiers, k -nn, SVM, and multinomial logistic regression all have high classification accuracy and area under the Receiver Operating Curve (AUC). The Naive Bayes classifier assumes that every two predictors are mutually independent given the class. The classification results indicate that the statistically independent components obtained by SICA do not the class information, so that the assumption of Naive Bayes may not be satisfied by using

independent components as predictors. It is also well-known that independence does not imply conditional independence generally. Future studies may use supervised ICA [24] and might be performed to create gait movement profiles across different subjects. Such that, if information existed about a group of subjects and their raw uncleaned EEG signals that includes cortical signals and motion artifact signals for a given movement, a new subject's raw uncleaned EEG signals could be used to classify the new subject's movement.

Acknowledgement We thank NIH for grant 1R01AG054621-01 which partially supported this study.

References

1. Anemüller, J., Duann, J. R., Sejnowski, T. J., & Makeig, S. (2006). Spatio-temporal dynamics in fMRI recordings revealed with complex independent component analysis. *Neurocomputing*, 69, 1502–1512
2. Barua, S., & Begum, S. (2014). A review on machine learning algorithms in handling EEG artifacts. In *Proceeding of the Swedish AI Society (SAIS) Workshop (SAIS 14)* (pp. 22–23).
3. Bell, A., & Sejnowski, T. (1995). An information maximization approach to blind separation and blind deconvolution. *Advances in Neural Information Processing Systems*, 7, 1129–1159.
4. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Berlin: Springer.
5. Brown, G. D., Yamada, S., & Sejnowski, T. J. (2001). Independent component analysis at the neural cocktail party. *Trends in Neuroscience*, 24(1), 54–63.
6. Calhoun, V. D., Adali, T., Hansen, L. K., Larsen, J. J. & Pekar, J. J. (2003). ICA of functional MRI data: an overview. In *In Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation* (pp. 281–288).
7. Calhoun, V. D., Adali, T., Pearlson, G. D., & Pekar, J. J. (2001). Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Human Brain Mapping*, 13, 43–53.
8. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
9. Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS One*, 7(2), e30135.
10. Enders, H., & Nigg, B. M. (2015). Measuring human locomotor control using EMG and EEG: Current knowledge, limitations and future considerations. *European Journal of Sport Science*, 16(4), 1–11.
11. Friedman, J., Hastie, T., Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22.
12. Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*, 45(2), 171–186.
13. Hyvärinen, A., & Oja, E. (1997) A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7), 1483–1492.
14. Hyvärinen, A., & Oja, E. (2000) Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4–5), 411–430.
15. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: Springer.
16. Keerthi, S., & Lin, C. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation*, 15(7), 1667–1689.

17. Kline, J., Huang, H., Snyder, K., & Ferris, D. (2015). Isolating gait-related movement artifacts in electroencephalography during human walking. *Journal of Neural Engineering*, *12*(4), 046022.
18. Langlois, D., Chartier, S., & Gosselin, D. (2010). An introduction to independent component analysis: InfoMax and FastICA algorithms. *Tutorials in Quantitative Methods for Psychology*, *6*(1), 31–38.
19. Majka, M. (2017) *naivebayes: High performance implementation of the Naive Bayes algorithm*. R package, 2017, version 0.9.1.
20. Marchini, J. L., Heaton, C., & Ripley, B. D. (2017) *Package fastICA*. R package, version 1.2-1.
21. Meyer, D., Dimiriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., et al. (2017) *e1071: Misc functions of the Department of Statistics, Probability Theory Group*. R package, version 1.6-8.
22. Nordin, A., Hairston, W., & Ferris, D. (2018). Dualelectrode motion artifact cancellation for mobile electroencephalography. *Journal of Neural Engineering*, *15*, 056024.
23. Oja, E., & Yuan, Z. (2006). The FastICA algorithm revisited: Convergence analysis. *IEEE Transactions on Neural Networks*, *17*(6), 1370–1381.
24. Sakaguchi, Y., Ozawa, S., & Kotani, M. (2002) Feature extraction using supervised independent component analysis by maximizing class distance. *IEEJ Transactions on Electronics Information and Systems*, *124*(1), 2502–2506.
25. Snyder, K., Kline, J., Huang, H., & Ferris, D. (2015). Independent component analysis of gait-related movement artifact recorded using EEG electrodes during treadmill walking. *Frontiers in Human Neuroscience*, *9*, 639.
26. Stone, J. (2005). Independent component analysis. *Encyclopedia of Statistics in Behavioral Science*, *2*, 907–912.
27. Weston, J., & Watkins, C. (1998). *Multi-class support vector machines*. Royal Holloway University of London, *2*, 907–912.
28. Yee, T. (2017). *VGAM: Vector generalized linear and additive models*. R package, version 1.0-4.