

Original Research

Automatic eye blink artifact removal for EEG based on a sparse coding technique for assessing major mental disorders

Raheel Zafar¹, Abdul Qayyum^{2,*} and Wajid Mumtaz²¹Department of Engineering, National University of Modern Languages, Islamabad 44000, Pakistan²Centre for Intelligent Signal & Imaging Research (CISIR), Department of Electrical & Electronic Engineering, Universiti Teknologi PETRONAS, Perak, 32610, Malaysia*Correspondence: enqr.qayyum@gmail.com (Abdul Qayyum)DOI: [10.31083/j.jin.2019.03.164](https://doi.org/10.31083/j.jin.2019.03.164)This is an open access article under the CC BY-NC 4.0 license (<https://creativecommons.org/licenses/by-nc/4.0/>).

In the electroencephalogram recorded data are often confounded with artifacts, especially in the case of eye blinks. Different methods for artifact detection and removal are discussed in the literature, including automatic detection and removal. Here, an automatic method of eye blink detection and correction is proposed where sparse coding is used for an electroencephalogram dataset. In this method, a hybrid dictionary based on a ridgelet transformation is used to capture prominent features by analyzing independent components extracted from a different number of electroencephalogram channels. In this study, the proposed method has been tested and validated with five different datasets for artifact detection and correction. Results show that the proposed technique is promising as it successfully extracted the exact locations of eye blinking artifacts. The accuracy of the method (automatic detection) is 89.6% which represents a better estimate than that obtained by an extreme machine learning classifier.

Keywords

EEG; eye blink artifact; ICA; sparse representation; support vector machine; pattern recognition; neural computation

1. Introduction

The electroencephalogram (EEG) is a standard modality for the study of neural activity by direct measurement from the scalp. In 1929, Hans Berger used EEG in humans for the first time. EEG offers portability, low cost, and ease of availability, and relatively high temporal resolution. For these reasons, EEG is popular in different brain applications such as the brain-computer interface (BCI) (Nezamfar et al., 2011; Robinson et al., 2011), decoding (Crouzet et al., 2015; Zafar et al., 2017) and seizure detection (Cecotti and Graser, 2011; Zhou, 2014). However, EEG data is confounded with various artifacts that may lead to serious misinterpretations, particularly in clinical studies. An artifact is typically unwanted noisy data that must be removed before further processing.

In most cases this noise has a large amplitude that affects data such that no meaningful statistical analysis can occur before arti-

fact removal. Hence, artifact elimination or correction is an essential pre-processing step for EEG analysis. For example, eye blink activity is a common type of artifact that involves high voltage levels. These typical voltage changes propagate from the eyeball through the head. Another type of artifact that can obscure an EEG signal is due to muscular or myogenic activity produced by contraction or expansion of head muscles. During its recording and transmission, there are many points where EEG data can be contaminated. These "artifacts" are mostly biologically generated and produced outside of the brain. Artifacts in EEG recordings generally result from eye blinks, eye movements, breathing, heartbeat, muscular activity, and line noise.

The eye blink artifact is very common in EEG data records and has a higher amplitude when compared to the usual signal associated with a given task. For example, EEG signals typically range between 0.5-30 Hz, whereas an eye blink artifact can be as large as 200 Hz. Eye artifacts are normally measured via electrooculogram (EOG), where pairs of electrodes are placed above and around the eyes. It is not easy to evaluate EOG because the measurements are contaminated with the EEG signals which are of interest. Thus, it is not possible to subtract the EOG signal even when an exact model of it exists (Jung et al., 2000). Eye movement artifacts are generated by reorientation of the retinocorneal dipole (Overton and Shagass, 1969). This artifact has an even stronger effect on the EEG than the eye blink artifact.

EEG data can also be corrupted by strong signals from the power supply. This artifact is usually filtered by using a notch filter. Muscle artifacts are typically produced by activity face and neck muscles. Signals related to this artifact exhibit a wide range of frequencies and can be distributed between different electrodes. Heartbeat or pulse artifacts originate from electrodes placed near or on a blood vessel where the change in voltage occurs during the recording because of contraction and expansion of the vessel. This type of artifact appears as a smooth wave or sharp spike and is determined by the raw EEG dataset (Cardoso, 1999).

Several methods have been proposed to correct the distortion produced by artifacts; however, each method has its advantages and disadvantages. New and improved techniques can decrease

artifacts, particularly those that are externally generated. Common techniques used to remove artifacts in EEG data are principal component analysis (PCA) (Subasi and Gursoy, 2010), independent component analysis (ICA) (Subasi and Gursoy, 2010) and canonical correlation analysis (Safieddine et al., 2012). Delay methods can also be used to address non-instantaneous mixing of brain and artifact signal sources (Dhiman et al., 2010). Alternatively, there are deterministic approaches which include wavelet transform (WT) and empirical mode decomposition (Safieddine et al., 2012).

ICA is particularly useful for removing eye blink and muscle artifacts from EEG records (see Comon, 1994; Jutten and Herault, 1991). Eye blinks and movement can be removed by subtracting their respective independent components (ICs). Artifact detection and removal are two of the standard applications of ICA for EEG (Jung et al., 1998; Vigário, 1997). In this procedure, components responsible for the artifacts are set to zero while the remaining ICs are projected back onto the scalp. ICA is so prevalent because it is available in freely available EEGLAB software (Delorme and Makeig, 2004). ICA-based clustering algorithms can also be used to remove artifacts from raw EEG data (Zou et al., 2012). ICA and PCA approaches have also been used to remove artifacts from magnetoencephalography data (Barbati et al., 2004; Jun and Pearlmuter, 2005). PCA has been proposed by Berg and Scherg (1991) to remove eye blink artifacts, but is unable to completely separate the artifacts from brain signals, particularly when amplitude differences are small (Lagerlund et al., 1997).

To increase classification accuracy using content recognition, sparse coding (SC) was applied to local image feature representations. SC represents data having a strong activation of a small set of neurons. The basic function of SC is normally learned from natural images. In SC, for a single stimulus there are different subsets of available combinations, and there is less chance of interference when it is simultaneously presented because the representation grows exponentially with the signal to noise ratio. SC modeling has been successfully used in image and video-based classification problems but has never been used for artifact detection. The method can solve classical problems based on image noise reduction, super-resolution processing, and restoration. It also performs well in several pattern recognition problems based on signal and image processing applications. Image categorization has also been achieved by using sparse coding for image patches (Zhang et al., 2015), so it provides an effective feature selection tool. Sparse coding offers the following practical advantages: (a) large storage capacity for coded signals (b) associative memory capacity; means capacity of associated similar patterns (c) ease of calculation, (d) easily structure natural signals, (e) minimize energy utilization as a general economic principle incorporated by biologic evolution, and (f) meets requirements to conclude electrophysiological experiments. Electrophysiological experiments are those experiments which are designed to extract the relevant information from the brain. Sparse coding helps in electrophysiological experiments by proving the required results. As a specific image processing application, SC, therefore, offers many benefits.

More recently, the use of hybrid methods has most recommended, especially concerning blind source separation-based methods such as ICA (Jung et al., 2000; Safieddine et al., 2012).

ICA and PCA are considered the most robust methods for decomposition of data into the underlying and the noisy components. The development of automatic criteria for identification of components representing artifacts has increased their utility for real-time automatic applications. In this context, the development of hybrid methods involving adaptive noise cancellation (adaptive filtering methods) may provide better solutions.

Here, the primary objective is to automatically detect and remove artifacts from noisy raw EEG data and by providing a robust solution with the help of classification algorithms such as a support vector machine, random forest or extreme machine learning techniques for depression assessment. Currently, the hospital collected healthy and depressed persons dataset is being employed for automatic artifact removal and classification of healthy and major depressive disorder (MDD) patients.

In this study, a hybrid method is proposed for the automatic detection and removal of only the eye blink artifact. To the author's knowledge, this is the first reported use of the proposed method to automatically detect this EEG artifact. The proposed approach is based on computing the similarity between sparse coefficients vectors using a Euclidian distance technique and saved the index or location of minimum value between sparse coefficients for the detection of the artifact. The proposed algorithm produced a better detection rate in terms of classification performance than currently existing artifact removal techniques. Based on the given dataset, various combinations have been used to test the proposed artifact detection method and validate the results by use of a classification technique for depression assessment. The various combinations are N1MDD1, N2MDD2, N3MDD3 and are discussed in result section. Eyeblink is one of the important artifacts that affect the data and detection of eye blink artifacts is not easy. Artifact removal is generally the most important step in the preprocessing of an EEG dataset. In this particular case, a more robust and accurate solution for the assessment of depression using healthy and major depressive disorder (MDD) patients was required. The proposed method automatically removes artifacts without any manual human intervention. It also employs classification techniques to validate the results for healthy and MDD patients. The techniques used for validation are support vector machine (SVM), random forest classifier (RF) and extreme learning machine (ELM) which are comprehensively discussed in result section.

2. Methodology

2.1 Dataset

This study involved two different subject groups (\pm standard deviation): 1) Thirty-two subjects with major depressive disorder (MDD mean age, 40 = \pm 12 years) and 2) Thirty-two age-matched healthy control subjects (mean age 38 \pm 15 years). Subjects were recruited from a clinic at the Hospital Universiti Sains Malaysia (HUSM), Malaysia. The MDD subjects were diagnosed using the Diagnostic and Statistical Manual (DSM-IV) (Barbati et al., 2004) which outlines the symptoms of depression. The diagnosis was made by an experienced psychiatrist from the HUSM clinic. Moreover, MDD subjects with psychotic symptoms, pregnant patients, alcoholics, smokers, and people with epilepsy were excluded. Subjects were asked to abstain from drugs and coffee for the duration of data recording. MDD severity was assessed

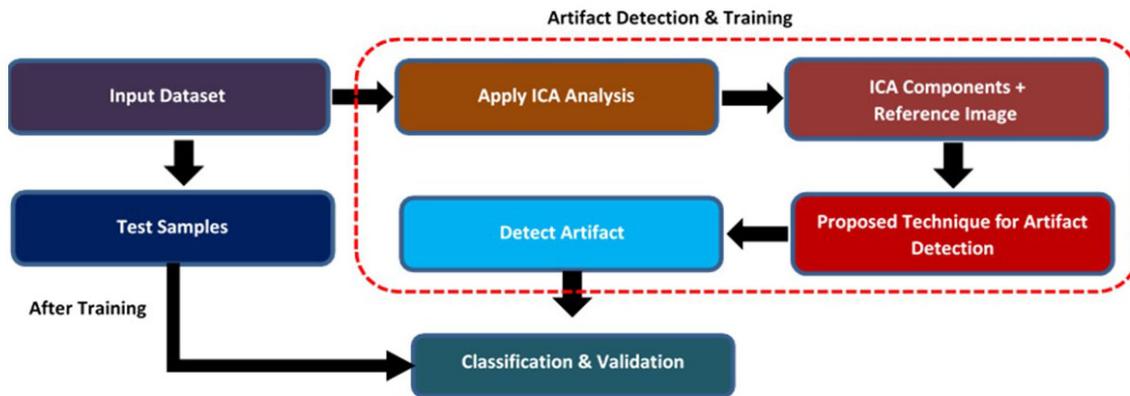


Figure 1. Sparse coding based method proposed for automatic artifact detection and classification of healthy and MDD EEG datasets. ICA is applied to input EEG data which is sent along with the reference image for artifact detection. The proposed technique is applied which detects the artifacts, and finally classification is done to check the validity of the proposed method.

based on two different clinical questionnaires: Beck's Depression Inventory-II (Davis et al., 1994; Mallat and Zhang, 1993) and the hospital anxiety and depression scale (Chen et al., 2001). Healthy controls were screened for physical and mental illness and were assessed to be disease-free.

The experimental design was approved by the HUSM ethics committee. Subjects signed consent forms and were fully aware of the experimental design. During recordings, EEG data acquisition involved vigilance-controlled monitoring: i.e., five minutes of both eyes closed and eyes open EEG data recording using a 19-channel EEG cap with linked-ear reference. The electrode placements followed the international 10–20 electrode placement standard, and the cap covered the scalp including the temporal, frontal, occipital, parietal, and central regions. The cap was attached to an amplifier (Brain Master Systems). A 50 Hz notch filter was used with a 0.5–70 Hz filter before the preprocessing. A sampling rate of 256 Hz was used to discretize data which were re-referenced to the infinity reference before analysis (Melgani and Bruzzone, 2004).

Thirty-two (32) subjects have participated in this study including twenty (20) subjects were male, and twelve (12) were female in the MDD group, and twenty-five (25) subjects were male, and seven (7) were female in healthy control group.

2.2 Method

The proposed eye-blink detection method involved a combination of ICA and sparse representation based on the proposed dictionaries. EEG data were subjected to the ICA using different ICA components, and the results were validated using topo map images. A reference topographic image eye-of blink activity was referred to as a reference input image. This image represented a standard eye-blink artifact in EEG data. Next, the proposed sparse coding technique was implemented by comparing a reference input image and the vector similarity between different numbers of components with the reference image. The proposed dictionaries stored the distinct features taken from the reference input data feature vector to compare with the ICA component feature vector. The comparison between two feature vectors was based on a computed similarity index value obtained from the reference input and

target images. A lower coefficient value indicates greater similarity between two feature vectors and the minimum similarity index value could be used for identification of the location of an artifact.

Following this comparison using the proposed sparse representation method, an index value referred to as the artifact location was stored. After artifact detection and removal, different machine learning algorithms are used to classify the healthy and MDD EEG datasets for assessment of the severity of depression.

Three different datasets were employed to validate the proposed method. The steps employed to detect and classify an artifact is given in Fig. 1.

In the proposed method, the sparse coding is first time implemented for this application. That is why the detail of the procedure is separately mentioned in the **Appendix**. The detail of sparse representation, sparse vector similarity measure, proposed hybrid dictionary, performance metrics and classification algorithms which includes support vector machine, random forest algorithm, and extreme learning machine is explained at the end of the paper in the **Appendix**.

3. Results

In this study, different EEG datasets were used to remove artifacts based on a sparse representation technique that employed 20 ICA components for their measurement and detection. Sparse representation is a powerful technique for extracting prominent features of a pattern from the ICA component. It uses a minimum distance technique to measure the similarity between reference data and the artefactual component. The results of 20 ICA components are given in Fig. 2. In this Fig., the detection of the artifact component is indicated with a circle. Results were similar to the artifacts detected in other datasets. Three datasets were based on healthy subjects, and three datasets were based on MDD patients. The location of the ICA components detected for the healthy and MDD datasets are given in Figs. 2-4 and 5-7, respectively.

The artifact detected from the first channel in the healthy state (N1) is shown in Fig. 2. The artifact detected at position 9 (channel 9) in the healthy state (N2) is shown in Fig. 3. In a healthy state (N3), the position of artifacts at 3 locations is shown in Fig. 4. Artifact location detection was also achieved using the proposed

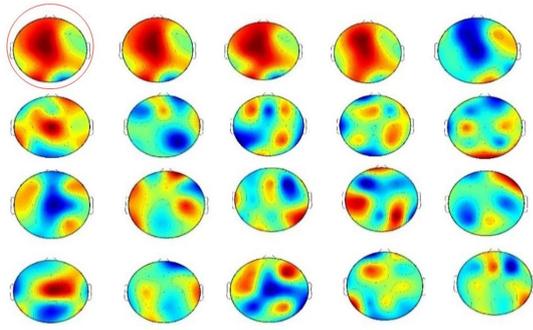


Figure 2. Visualization validation of artifact detection by the proposed model for healthy patient subject 1 (N1). The artifact detection is done on the first channel. The above result uses a minimum distance technique to measure the similarity between reference data and the artefactual component. The detection of the artifact component is indicated with a circle.

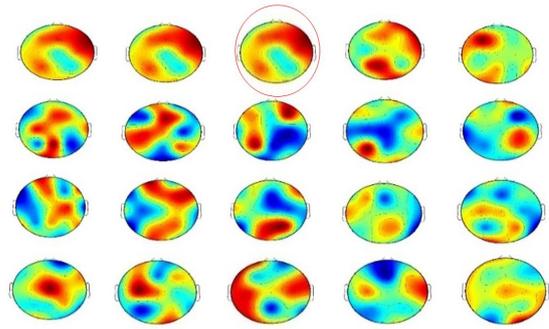


Figure 4. Visualization validation of artifact detection by the proposed model for healthy subject 3 (N3). The artifact detection is done at 3rd channel. The above result uses a minimum distance technique to measure the similarity between reference data and the artefactual component. The detection of the artifact component is indicated with a circle.

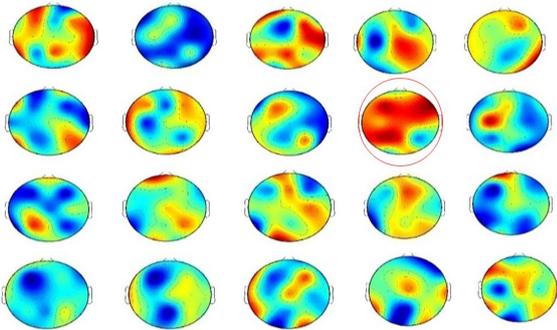


Figure 3. Visualization validation of artifact detection by the proposed model for healthy subject 2 (N2). The artifact detection is done on the 9th channel. The above result uses a minimum distance technique to measure the similarity between reference data and the artefactual component. The detection of the artifact component is indicated with a circle.

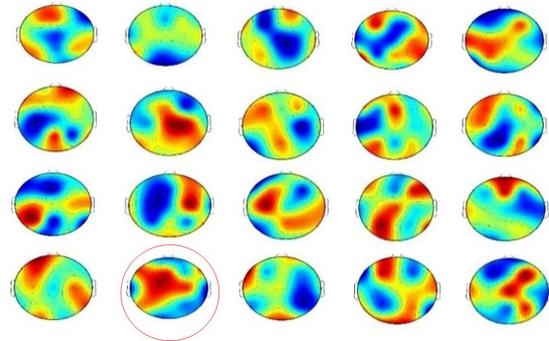


Figure 5. Visualization validation of artifact detection by the proposed model for MDD Subject 1. The artifact detection is done at 17th channel. The above result uses a minimum distance technique to measure the similarity between reference data and the artefactual component. The detection of the artifact component is indicated with a circle.

technique for the MDD dataset. Here, an artifact occurs at location 17 as shown in Fig. 5. The location of the artifact detected at location 4 in the MDD2 dataset is shown in Fig. 6. The artifact detected at location 4 in the MDD3 dataset is shown in Fig. 7. The topographic map visualizes activation in healthy and MDD subjects. The more activated colormap shows the artifact existing at a particular location in each MDD case. The heat map plots of some patients determined by the proposed model were validated by manual artifact detection. For most MDD cases this detection matched automatic detection as shown in Figs. 5, 6, 7.

Three classifiers were used to evaluate the detection rate of the proposed algorithm based on the different number of dictionaries used. The most well-known classifiers were SVM, RF, and ELM. The details of each classifier were discussed above. The different performance metrics were evaluated based on ground truth, and artifact location estimates using 50 iterations were based on different combinations of datasets. The dataset of six patients (three healthy and three MDD) was used for the validation/testing of artifact location detection which classified these artifacts based on the

proposed classifiers. These combinations were chosen using one healthy subject, and one MDD1 patient dataset in a combination referred to as "N1MDD1". Similarly, a combination can be made using other healthy and MDD datasets, referred to as "N2MDD2" and "N3MDD3". Classifiers were applied in these combinations to measure performance metrics of accuracy, precision, and recall using various dictionaries. The adaptive dictionary (KSVD) adaptive dictionary-based performance metrics for SVM, RF and ELM are shown in Fig. 8 (A, B, C). Performance metrics based on the discrete-time Tchebichef transform (DTT) dictionary are shown in Fig. 8 (D, E, F) using the SVM, RF and ELM classification algorithms. Similarly, for the DRT dictionary, performance metrics are given in Fig. 8 (G, H, I). These metrics were computed by taking the average value between healthy and MDD subjects. The highest accuracy produced by the proposed dictionary was for the N3MDD3 combination. The comparison between dictionaries is given in Fig. 9. The discrete ridgelet transform (DRT) was more accurate when employing L3MDD3 data when com-

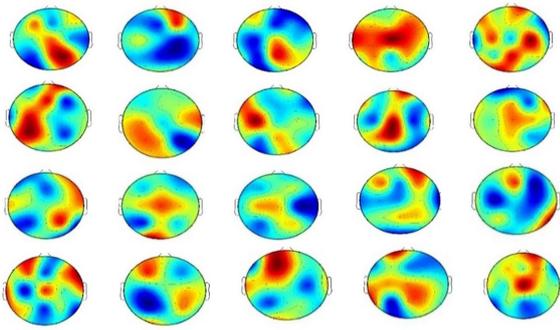


Figure 6. Visualization validation of artifact detection by the proposed model for MDD Subject 2. The artifact detection is done at 4th channel. The above result uses a minimum distance technique to measure the similarity between reference data and the artefactual component. The detection of the artifact component is indicated with a circle.

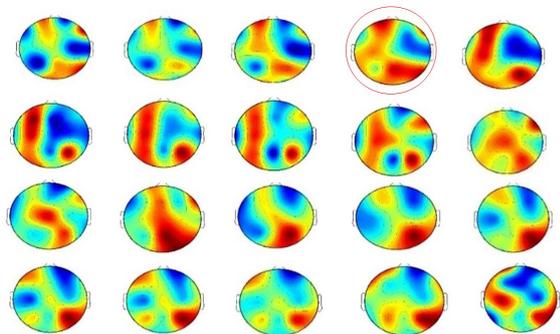


Figure 7. Visualization validation of artifact detection by the proposed model for MDD Subject 3. The above result uses a minimum distance technique to measure the similarity between reference data and the artefactual component. The artifact detection is done at 4th channel. The detection of the artifact component is indicated with a circle.

pared with other dictionaries and captured more prominent features during matching of reference and target data. The DTT dictionary also detected artifacts and comparatively produced higher accuracy when compared with the KSVD dictionary as shown in Fig. 9. For validation, a combination of healthy and MDD subjects was chosen for classification of a two-class sample. The healthy samples from first-class are denoted as one and the samples from MDD. The other class is denoted as zero. For validation/testing, we have passed the sample from healthy and MDD patients to the trained classifiers and compute the average accuracy, precision, and recall between two samples (healthy and MDD). These performance metrics (accuracy, precision, and recall) are taken as average between healthy and MDD patients' data samples. The best accuracy, precision, recall score based on various proposed dictionaries are shown in Fig. 9.

In machine learning and statistics, the ROC (receiver operating characteristic) curve has been used to check the performance of binary or multiclass classification problems graphically by providing the various discrimination threshold. It is used to test the evaluation based on various curves generated based on distinct num-

ber of threshold. The ROC used the true positive rate and false-positive rate for plotting different curves based on various cut-off points' parameters. The sensitivity used in machine learning to check the classifier capability and find detection rate and it is also called the true-positive rate. The probability of false alarm can be computed based on the false-positive rate, and it is also known as specificity of the system. In ROC curve, each point shows the specific sensitivity/specificity pair value and these values show the discrimination between classes based on parameter optimization for every decision threshold. The area under the ROC determines the distinction between groups of classes for a parameter. The classification threshold produced various curves for the sensitivities and specificities, and these values cannot maximize together at a certain point in the ROC plane for every parameter.

The ROC curves determined the number of true positives and false positives based on precision and recall values. The number of 50 iterations were used to determine the ROC using precision and recall metrics and is shown in Fig. 10. Each ROC shows that our proposed method based on the N3MMD3 dataset produced better results compared to other datasets by optimizing the single parameter using KSVD dictionary as shown in Fig. 10 (A, B, C). Similar each ROC produced based on the DTT dictionary provided comparatively better performance using N2MMD3 dataset. Moreover, each ROC shows that the SVM classifier produced better performance as compared to RF and ELM. It is concluded that the ROC performance is better with the adaptive dictionary (KSVD) and produced better performance by optimizing the single-precision parameters. The KSVD dictionary-based proposed model produced truer positive using healthy3 and MDD3 dataset and is shown in Fig. 10 (A). There is a slight difference between true positive and false-positive rates for the three chosen patient datasets. However, these ROCs curves indicate that the proposed model achieves optimal performance for the healthy and MDD datasets. This proposed model showed consistent results for all healthy and MDD cases, which indicates that the proposed model could be used for real-time artifact detection and further assessment and classification of depression.

3.1 Comparison with existing automatic artifact removal methods

Most EEG applications process information automatically in real-time. However, manual identification of the artifact component is time-consuming and may not provide an efficient solution for multi-channel EEG data sequences. However, *a priori* information based on statistical characteristics is required for artifact detection by many signal processing techniques. The methods based on ICA/PCA techniques somehow provided semi-automated artifactual identification system that required some training parameters, however few automated artifact detection techniques require training samples for supervised classification. That is why in this study, few techniques based on ICA/PCA have been compared with our proposed method in Table 1. Automatic detection methods based on ICA require a further stage to make the system fully automated. The computational complexity of the proposed method is only slightly greater, based on real-time implementation takes five minutes compared with an ICA based method which takes three minutes for the detection of an equivalent artifact.

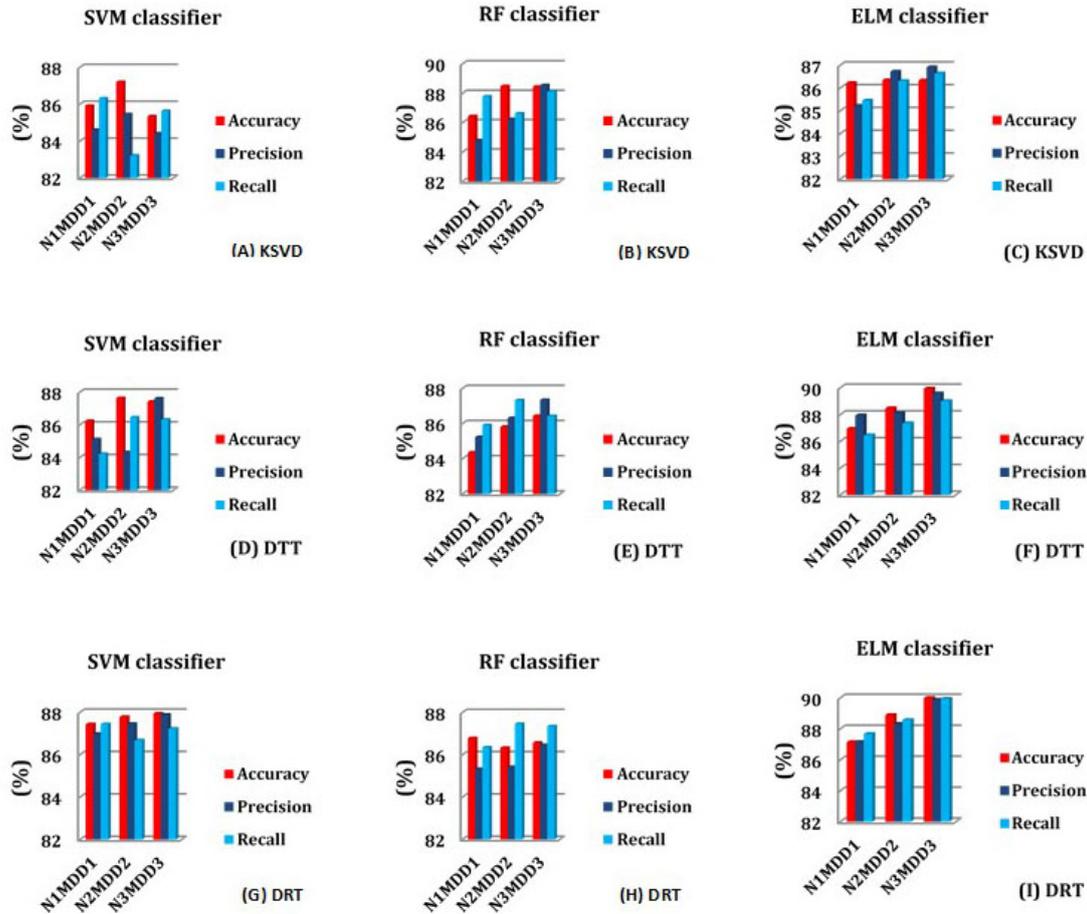


Figure 8. The first row represents accuracy, precision, and recall using KSVd dictionary (A, B, C) based on SVM, RF, and ELM. The second row represents accuracy, precision, and recall using DTT dictionary (D, E, F) based on SVM, RF, and ELM. The third row represents the accuracy, precision, and recall using DRT dictionary based on SVM, RF, and ELM (G, H, I).

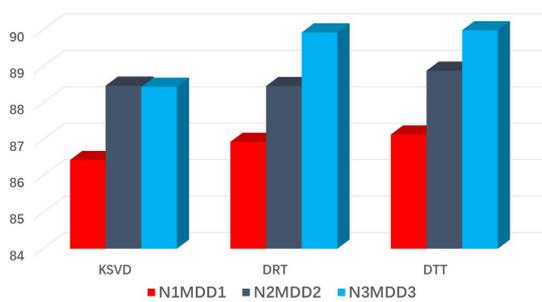


Figure 9. The comparison of the accuracy between different dictionaries based on healthy and MDD dataset using ELM classifier. These performance metrics (accuracy, precision, and recall) are taken as average between healthy and MDD patients' data samples. The best accuracy, precision, recall score based on various proposed dictionaries is shown.

4. Discussion

Eyeblink artifacts are the main source of contamination in EEG signals, so they should be removed before further analysis. The purpose of the current study was to design a framework that could automatically detect and correct eye blink artifacts by replacing them with more accurate values. Different methods of detection such as wavelet analysis (Pesin, 2007), ICA (Flexer et al., 2005), semi-automatic (Flexer et al., 2005) and fully automatic methods (O'Regan et al., 2013; Zou et al., 2012) have been described in the literature. However, in this study a sparse coding technique was used for the first time for the automatic detection of eye blink artifacts.

Sparse representation is a powerful technique used to automatically extract prominent features in a compressed form based on a proposed dictionary technique. These dictionaries capture features using basis functions and provides feasible and stable features. The dictionaries compress these features in feature space by extracting unique information from the data. Sparse spaces are linearly independent and provided discriminative information for purposes of classification and detection. The sparse technique de-

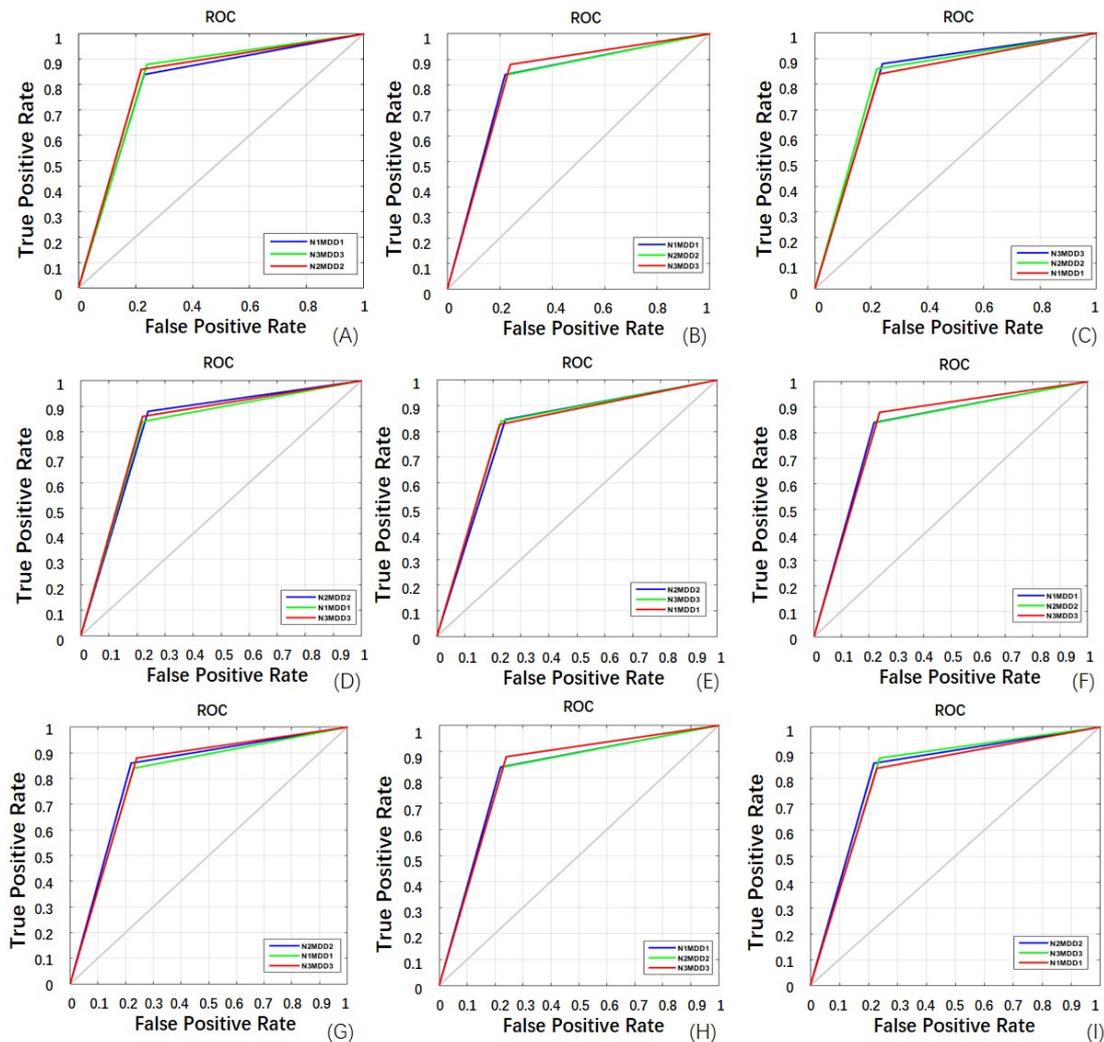


Figure 10. ROC curves for KSVd using SVM, RF, and ELM machine learning algorithms (A, B, C). ROC curves for DTT using SVM, RF and ELM machine learning algorithms (D, E, F). ROC curves for DRT using SVM, RF and ELM machine learning algorithms (G, H, I). The number of 50 iterations were used to determine the ROC using precision and recall metrics. Each ROC shows that N3MDD3 dataset produced better results compared to others.

composes a signal into linear combinations of dictionary atoms and stores the discriminative values of that signal in sparse space. This is why this technique captured only the unique and prominent features of the input signal in sparse feature space. Such features are uniquely represented in sparse space, are linearly independent of other class data and are also more stable and effective for automatic artifact detection and classification.

To validate the performance of the proposed method, different existing techniques are applied to the test datasets. The comparison between existing techniques and the proposed method was assessed in terms of accuracy, specificity, and sensitivity. Results are summarized in Table 1. This table shows that the proposed algorithm was more accurate in the automatic detection of artifacts. The use of the proposed DRT dictionary and the associated analytic method provided comparatively better accuracy, precision and recall of values as compared with existing automatic artifact detection methods. When employed with a pre-existing KSVd based dictionary the approach described here produced a better

detection rate in terms of accuracy, precision and recall as compared to existing methods; however, detection results were comparatively less accurate when compared with the proposed method when it employed the DRT dictionary.

In the study, 20 ICA was used during analysis to measure, whereas, accuracy was assessed using three different classifiers, i.e. SVM, RF, and ELM. Moreover, three different dictionaries were used to capture the most prominent features during the matching with reference data and DTT was found to be better when compared with DRT and KSVd.

To validate the proposed method, it was compared with different existing methods in terms of accuracy and features, and these methods included wavelet analysis and semi- and fully automatic approaches. For variability and reliability of the study, data of two different states were used during the analysis, i.e. three healthy and three MDD subjects. One of the advantages of the proposed method over wavelet analysis and semi-automatic approaches is that it detects artifacts automatically. In comparison to other au-

Table 1. Comparison of automatic artifact detection methods (Proposed and extant methods)

Methods	Accuracy (%)	Precision (%)	Recall (%)
Irene Winkler (Winkler et al., 2014)	81.9	79.45	80.93
K-means with Similarity (Qi et al., 2009)	86.56	85.95	85.77
Yuan Zou (Zou et al., 2016)	88.35	86.35	87.35
Auto-mutual information (Nicolaou and Nasuto, 2007)	84.9	84.78	85.68
Proposed Method with DRT	89.63	89.1	88.89
Proposed Method with KSVD	88.56	87.91	88.45

automatic detection techniques, a fixed dictionary approach was employed. It was based on sparse representation and detected and corrected eye blink artifacts automatically. The method produced superior results in terms of accuracy, specificity, and sensitivity. According to the author's knowledge, this is the first report of this type of application, and it produced better detection and correction accuracy (> 89%).

5. Conclusion

Here, sparse representation with fixed dictionaries was used to detect and correct artifacts from an EEG dataset. The dataset was composed of two distinct groups, i.e., healthy controls and MDD subjects. Results showed that the proposed algorithm provided automatic and accurate detection of eye blink artifacts as compared with existing techniques. The proposed technique involved capturing prominent features of a data set based on pre-specified dictionaries and comparing the obtained patterns with reference images to measure the similarities between the reference and ICA-based components. This method demonstrated a robust response against artifacts in which detection is easy, feasible, precise and automatic compared with existing artifact removal methods, although due to the dictionary record the computational time may be greater than for existing techniques.

Acknowledgment

This research is supported by the Ministry of Education (MOE), Malaysia, under the Grant of Higher Institution Centre of Excellence (HICoE) for CISIR (Ref: 0153CA-002).

Conflict of interest

The authors declare no conflict of interest.

Submitted: May 16, 2019

Accepted: July 23, 2019

Published: September 30, 2019

Appendix

A.1 Sparse representation

A sparse coding or sparse representation is widely used in neuroscience, machine learning, and numerous other signal and image processing applications (Zhang et al., 2015). The purpose of sparse modeling is to find a method that represents signals as a linear combination of a few typical patterns called "atoms", which are extracted from a dictionary. For a given image or signal $y \in R^n$ and a dictionary matrix $D \in R^{n \times k}$ that contains K atoms as column vectors $d_j \in R^n$, $j = 1, \dots, K$, the sparsest vector is defined in such a way that $x \in R^k$ and that $y \cong Dx$. The problem is solved by using

the optimization problem as shown in Eqn. 1:

$$\min \|x\|_0 \text{ subject to } \min \|y - Dx\|_2 \leq \epsilon \quad (1)$$

where ϵ is the reconstruction error of the signal y using the dictionary D and the sparse code x . Alternatively, based on low sparseness, this optimization problem has been solved by:

$$\min \|y - Dx\|_2 \text{ subject to } \min \|x\|_0 \leq \rho \quad (2)$$

where ρ is a specified sparsity level. The vector $x \in R^k$ comprises the representation coefficients of the signal y with respect to the dictionary D . Compared with other methods, such as PCA, sparse coding computes the vector with the smallest number of nonzero coefficients. The sparse coefficients formulation usually uses the l_0 -norm (pseudo-norm), which counts the nonzero entries of a vector. This formulation is an NP-hard problem (Mallat and Zhang, 1993), but can be solved using optimization greedy algorithms. These algorithms are called matching pursuit (MP) (Mallat and Zhang, 1993) or orthogonal matching pursuit (OMP) algorithms (Davis et al., 1994). The basis pursuit (BP) is the second class of algorithm based on relaxation where the l_0 -norm is replaced by a l_1 -norm. This converts the optimization problem into a convex problem that can be solved efficiently. Such methods are referred to as basis pursuit (BP) (Chen et al., 2001). The key point is that all these techniques are used to efficiently obtain the sparsest coefficients using the dictionary elements. The challenging step is construction of the dictionary. After this, dictionary choice is another key step that provides an optimal solution for the sparse coefficients. A predetermined dictionary enables a fast and efficient solution for the sparse representations used in image classification. The extant dictionaries are either a discrete cosine transform (DCT) or a discrete wavelet transform (DWT). The underlying functions originating from the discrete ridgelet transform (DRT) and the discrete-time Tchebichef transform (DTT) are used to construct the proposed dictionaries. Thus, it is more efficient to learn the dictionary from a given set of training data (Mallat and Zhang, 1993).

The MP algorithm has been used in over-complete dictionaries to construct the best matching atoms for each iteration based on the signal residual approach and produce the best matching atoms for sparse representation until all atoms reach a defined stopping criterion. The OMP algorithm is efficient when compared to the MP algorithm for sparse signal representation in classification problems. It produces sparse coefficients efficiently by using an orthogonal projection for each direction between dictionary elements, uses fewer iterative steps and achieves better performance

Table 2. Comparison of automatic artifact detection methods (Proposed and extant methods)

<p>Algorithm. Implementation steps for the OMP algorithm.</p> <p>Task: Approximation of sparse coefficients: $\alpha_1 = \operatorname{argmin}_{\alpha_1} \cdot \ \alpha_1\ _0 \text{ s.t. } B_1 = D\alpha_1$</p> <p>Input: Input sample B_1, Dictionary matrix D, coefficients vector for sparse representation α_1</p> <p>Initialization: $t = 1, r_0 = B_1, D_0 = \text{empty}$, index set $\Lambda_0 = \varnothing$. where \varnothing is empty set, τ is a constant value that is very small. d_j is the dictionary elements concatenated column-wise to construct the dictionary.</p> <p>While $\ Y_t\ > \tau$ do</p> <p>Step 1: Calculate the matching elements using the maximum separation between residue and dictionary elements and construct the bigger inner product between these vectors, i.e., the and () by manipulating $\lambda_t = \operatorname{argmax}_{j \notin \Lambda_{t-1}} \langle Y_{t-1}, d_j \rangle$.</p> <p>Step 2: Update the index set for every iteration $\Lambda_t = \Lambda_{t-1} \cup \lambda_t$ and reconstruct the dataset $D_t = [D_{t-1}, d_{\lambda_t}]$.</p> <p>Step 3: Calculate the sparse coefficients based on the least square algorithm $\check{\alpha}_1 = \operatorname{argmin} \cdot \ B_1 - D_t \check{\alpha}_1\ _2^2$</p> <p>Step 4: Update the representation residual using $Y_t = B_1 - D_t \check{\alpha}_1$</p> <p>Step 5: $t = t + 1$</p> <p>End</p> <p>Output: D</p>
--

when compared to the MP algorithm. The number of iterations used by the OMP algorithm is given in Table 2. The OMP greedy approach used a l_0 -norm minimization approach to approximate the sparse coefficient. An algorithm that employs the OPM optimization technique is given in Table 2; the OMP algorithm is efficient and fast. OMP is a basic and simple method that has been used to solve convex optimization problems and non-homogenous and non-linear problems. BP is another optimization technique; however, it generated greater computationally complexity when compared to OMP.

The OMP used dictionary atoms to compute the sparse coefficients for sparse representations. The similarity between sparse coefficients has been estimated using the sparse representation approach shown in Table 2.

A.2 The sparse vector similarity measure

The similarity between two images (left and right images in stereo imaging) could be measured based on minimum distance approaches such as cosine similarity or dot product vector method. The disparity range is a displacement value between left and right stereo images. The sparse coefficients y extracted from the left image using the proposed sparse modeling technique measure similarity with sparse coefficients y'_i extracted from right image based on sparse modeling within a certain disparity range $d = [d_{min}, d_{max}]$. The disparity value or range in our experiment was (0, 12). The angle θ between the vectors y and y'_i can be computed for each i . The scalar product between vectors is zero if these vectors are perpendicular to each other otherwise a non-zero value is generated by the product of these vectors. The dot product between vectors is computed by Eqn. 3:

$$\langle y, y'_i \rangle = |y| |y'_i| \cos \theta, \forall i \quad (3)$$

$$F_i = \cos \theta = \frac{\langle y, y'_i \rangle}{|y| |y'_i|} \quad (4)$$

where $|y| = \sqrt{\langle y, y \rangle}$, $\langle y, y'_i \rangle = \sum_j (y_i, y_j)$ and

$$A(d) = \max(F_i) \quad (5)$$

The angle θ between the vectors y and y'_i is chosen such that the two vectors are not perpendicular. This method provides a similarity between two sparse vectors. F_i is the cosine similarity value computed using two vectors and $A(d)$ gives the maximum value obtained from the cosine similarity value F_i .

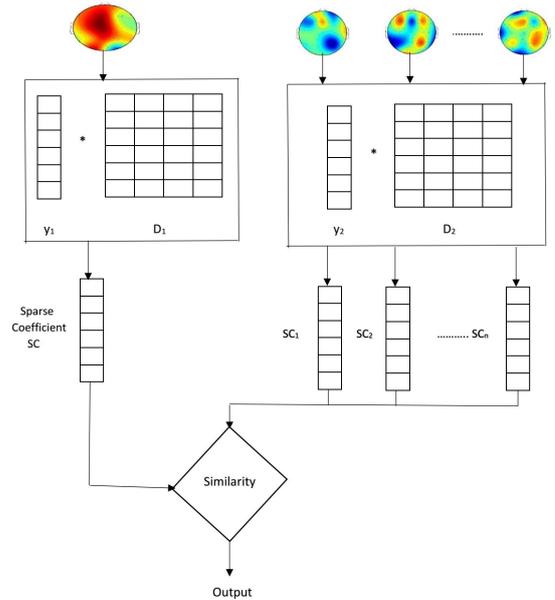


Figure 11. Proposed sparse representation algorithm using dictionary and the minimum distance technique for automatic artifact detection. Sparse coding computes the vector with the smallest number of nonzero coefficients. The primary goal is to obtain the sparse coefficients using the dictionary elements, which is followed by the similarity measured based on the minimum distance approach.

A.3 Proposed hybrid dictionary (DRT)

Many tasks in image processing applications depend on the type of dictionary used for sparse representation. Among others, these include de-noising, super-resolution, and fusion. Hence, the selection of an over-complete dictionary plays a fundamentally

Table 3. Algorithm for construction of a hybrid dictionary (DRT) for sparse coding location matching

Algorithm. Dictionary-based on hybrid wavelet Basis Function	
1	D is dictionary size, S is scale factor, T is translation parameter, θ is the rotation parameter, M_1 is the number of dictionary atoms, N_1 is the size of dictionary, M is the constant translation parameter length. M is the translational length.
2	FOR eachScale $S = 1 : M_1$ do
3	FOR eachtranslation $T = 1 : M$ do
4	FOR eachrotation $\theta = -\pi : \pi$ do
5	FOR each $n_1 = 1 : M_1$ do
6	FOR each $n_2 = 1 : N_1$ do
7	$Temp(n_1, n_2) = \sqrt{S} \times \sin(n_1 \times \cos \theta + n_2 \times \sin \theta - T) e^{-t/2}$
8	$D = temp(\cdot)$;
9	ENDFOR END FOR
10	$storeD(:, count) = temp(\cdot)$
	END FOR
11	END FOR
	ENDFOR
	Select bases have greater variance than a certain threshold $T_1 = M^{2 \times 0.05}$
12	Select size of dictionary $D_1(:, count) = D_1(:, sizeofdictionary)$
	$D = D_1$
	END FOR

significant role for the sparse recovery atoms. Sparse representation was employed here to approximate the basis element functions responsible for image classification, for which the ridgelet-based over-complete dictionary proved the better choice for the given sparsified image classification task. Moreover, we the Ricker wavelet—the negative normalized derivative of the Gaussian function—was used to create ridgelet-based elements for the hybrid dictionary.

Results showed superior performance for the ridgelet-based dictionary compared to various other over-complete dictionaries based on transform functions, such as DCT, wavelets, and curvelets. 2D ridgelets were defined by a wavelet-type function:

$$\psi_{x,y,\theta}(t_a, t_b) = x^{-1/2} ((t_a \cos \theta + t_b \sin \theta - x) / a) \quad (6)$$

where $\psi(\cdot)$ is a wavelet function. t_a and t_b are the line coefficients, x is the intercept, and a is the scaling factor used to compute the basis function.

The Ricker wavelet function is given by:

$$\psi(t) = \frac{2}{3} \pi^{1/4} \left(1 - \frac{t^2}{\sigma^2} \right) e^{-\frac{t^2}{\sigma^2}} \quad (7)$$

Ridgelet bases were obtained by selecting different values for x , y , and θ in Eqn. 6. These bases were prepared as vectors for insertion into the hybrid ridgelet-based dictionary. Pi and sigma are constants which produced richer wavelet coefficients when based on the Ricker wavelet function.

For these reasons, the ridgelet transform was proposed to model sparse signals and measure sparse coefficients for images. The ridgelet transform is an efficient 2D transform that can be used to store multi-dimensional signals or images. For example, in Fourier and wavelet analysis, ridgelet analysis is also used to approximate non-linear signals. Enhanced approximations can be constructed by using a simple algorithm for ‘ N ’ ridgelet functions

(Feng et al., 2015). Ridgelet analysis for object representation is extremely effective due to objects with singularities along lines by means ridgelets as a way of concatenating 1D wavelet transform along lines. This use of ridgelet transforms in image processing is very attractive because singularities are frequently joined together along the edges of an image. Thus, the proposed hybrid dictionary becomes the better choice for the construction of an overcomplete dictionary that gives better approximations for sparse representation.

We have extracted the feature vector (x) from input EEG data and applied D dictionary basis function on the input EEG features. The y vector is extracted from the inner product of D and x in a sparse representation way using OMP basis function in Eqns. 1 and 2. Eqns. 3 to 5 shows how to compute similarity measured between vector y and y'_i based on dictionary D and input EEG vector (x) and also measured maximum value based on index of similarity between y and y'_i and Eqns. 6 and 7 represented how we can compute the dictionary D based on basis functions $\psi(t)$.

Table 3 gives the algorithm used to construct the hybrid dictionary with ridgelet-based functions. Dictionaries comprising ridgelet-based functions are over-complete and constructed by different scaling factors and basis functions that employ Ricker wavelet functions. Each loop shown in Table 3 illustrates the different scaling, translation, and rotation parameters used to select the basis for DRT dictionary functions. The *temp* function was used to store the basis of the dictionary for each column, and the threshold for variance was set at 0.05 to normalize and scale dictionary atoms.

A.4 Dictionary construction for sparse representation

The use of fixed dictionaries was proposed because they are extremely fast and reliable (accurate). Pre-specified dictionaries are based on discrete cosine, wavelet, ridgelet, and Tchebichef transform basis functions. For comparison, the KSVD (K means-singular value decomposition) adaptive dictionary is implemented

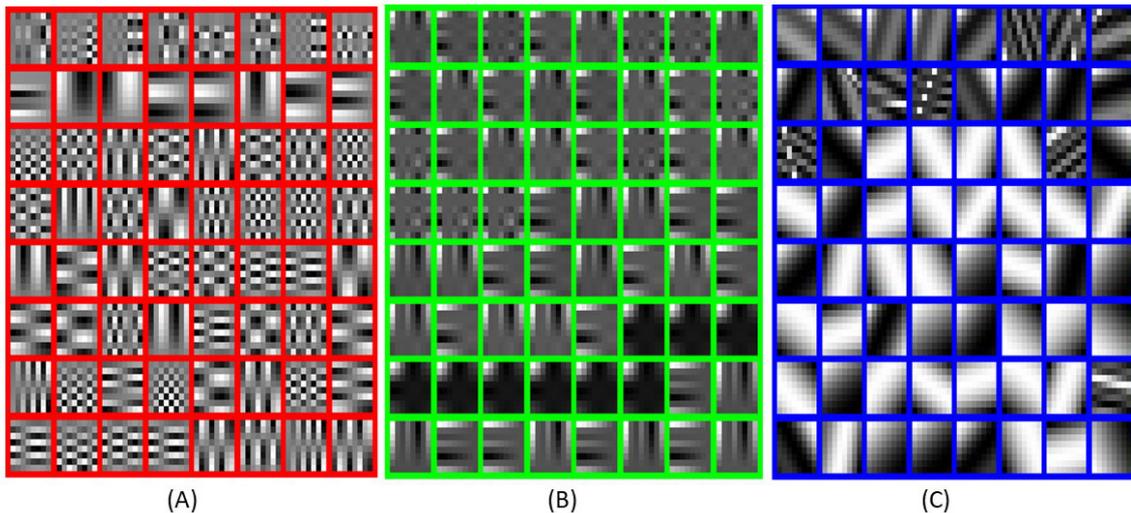


Figure 12. Basis functions used by (a) KSVD, The KSVD-based dictionary elements (or 'atoms') mostly capture singular image points (b) DRT, The dictionary stores image points in well-ordered forms that also capture structural patterns. Moreover, they performed well with smooth and regular patterns. (c) DTT, It also stored image structures well with regular patterns for all dictionary elements using an 8×8 EEG signal patch.

and compared with the dictionaries proposed here (DRT and DTT). The KSVD-based dictionary elements (or 'atoms') mostly capture singular image points (see Fig. 12 (A)). These values are scattered over different rows and columns compared to initial cells in a dictionary. The dictionary stores image points in well-ordered forms that also capture structural patterns. Moreover, they performed well with smooth and regular patterns. The DRT-based dictionary captured multiple irregular patterns concurrently, as shown in Fig. 12 (B). Similarly, the DRT dictionary demonstrated a good capture structure when compared with other fixed dictionaries such as DCT, DWT, and DTT and also stored image structures well with regular patterns for all dictionary elements. The DTT dictionary has the same cell pattern structure and captured good edge points in regular forms as shown in Fig. 12 (C). The patch size used for all dictionaries was (12×12) , which is optimal and captures well-structured patterns in an image. (When dictionary size increased, computational complexity also increased).

A.5 Performance metrics

Precision and recall are two of the most commonly used evaluation metrics in pattern recognition and information retrieval. Precision-recall used the relevance classification criteria. The precision can be defined as the ratio of the number of retrieved elements to the total number of relevant elements in an instance.

A.6 Classification algorithms

A.6.1 Support vector machine

Support vector machines (SVMs) have been extensively used in image processing and pattern recognition applications. They use a hyperplane for separation of training data using multidimensional training values of a fixed number of different classes. It minimizes the objective function by maximizing the learning $(x) = \langle w \cdot x \rangle + b$ from the sample data $(x) = \langle w \cdot x \rangle + b, \{x_i, y_i, i = 1, \dots, N\}$ among the closest sample data and the hyperplane x_i is an n -dimensional feature vector, where $y_i = \pm 1$, and minimize the $\frac{\|w\|_2^2}{2}$ using the constraint $y_i (\langle w \cdot x_i \rangle + b) \geq 1$ under maximization criteria.

The SVM classifier is introduced as a binary classifier that is converted into a multiclass classifier using two strategies. These strategies are called one against one (OAO) and one against all (OAA) (Manikandan and Venkataramani, 2009). The OAO classifier technique classifies every pair of classes while using the most common label for each pixel. The OAA technique classifies each class against the rest, and it chooses the label with largest confidence for each pixel. This strategy performs better when the number of classes is small (usually < 10) (Manikandan and Venkataramani, 2009). Here, OAA was employed due to the small number of classes (less than five).

A.6.2 Random forest algorithm

The random forest (RF) can be used for image classification in remote sensing applications due to its superiority and robustness to noise compared with other classifiers (Gislason et al., 2006). In Feng et al. (2015) proposed RF based on an ensemble learning technique. It required fewer parameters while running compared with other machine learning classifiers (SVM, ANN). The popularity of RF increased gradually due to achieving equal or higher accuracy in the field of remote sensing when compared with SVM for image classification (Martin et al., 1998; Pal, 2005). Random forest is based on an ensemble of independent individual classifications and depends on a regression tree (CART). The RF has the final response for calculating all the decision tree's output. There are two steps involved in the selection of this process for the evaluation of RF classifiers. The bootstrap approach selects 70 percent training samples randomly for the decision tree in the first step, and the second step uses the remaining 1/3 samples for out-of-bag (OOB) data, which in RF is used during cross-validation for evaluation of the classification accuracy. The OOB is very sensitive and may cause over-fitting to the training data (Pal, 2005). The advantage of the selection of random subset predictor variables is that it provides a better generalization capability and less correlation between trees. The main advantage of RF is that it gives the

contribution of each variable to the classification accuracy. There are two parameters used in RF: the number of trees, denoted as *ntree*, and number of selected random predictor variables (Feng et al., 2015), denoted as *mty*. Usually OOB error has an inverse relation with *ntree*: by increasing *ntree*, the OOB error is decreased within a certain threshold. There are two methods used for calculating *mty* RF: one-third and square root. RF is insensitive to outliers and changing of hyperparameters during training phase; it produces less computational burden and makes it easy to determine hyperparameters during training. There are fewer issues of over-fitting due to an individual decision tree in RF.

A.6.3 Extreme learning machine

The extreme learning machine (ELM) is a basic and new learning algorithm that has been developed from a single hidden layer feedforward neural networks (SLFN) (Marc'Aurelio Ranzato et al., 2007). It is a very time-consuming process to adjust the input weights and hidden layer bias for all feed-forward neural networks. To minimize or overcome these problems using traditional gradient-based learning algorithms, Huang et al. (2015) proposed an SLFN by randomly choosing input weights and hidden layer biases for an infinite activation function in the hidden layer. The SLFN can be observed as a linear system, and the determination of outputs weights are updated analytically. ELM is described in great detail by Huang et al. (2015). Based on input N data samples $\langle x_i, t_j \rangle$, where $x_j = [x_{j1}, x_{j2}, \dots, x_{jn}]^T$ is the j th sample with n -dimensional features, $t_j = [t_{j1}, t_{j2}, \dots, t_{jm}]^T$ characterize the actual labels of x_j standard SLFN with M hidden neurons.

References

Barbati, G., Porcaro, C., Zappasodi, F., Rossini, P. M. and Tecchio, F. (2004) Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals. *Clinical Neurophysiology* **115**, 1220-1232.

Berg, P. and Scherg, M. (1991) Dipole models of eye movements and blinks. *Electroencephalography and Clinical Neurophysiology* **79**, 36-44.

Cardoso, J. F. (1999) High-order contrasts for independent component analysis. *Neural Computation*, **11**, 157-192.

Cecotti, H. and Graser, A. (2011) Convolutional neural networks for P300 detection with application to brain-computer interfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**, 433-445.

Chen, S. S., Donoho, D. L. and Saunders, M. A. (2001) Atomic decomposition by basis pursuit. *SIAM Review* **43**, 129-159.

Comon, P. (1994) Independent component analysis, a new concept. *Signal Processing* **36**, 287-314.

Crouzet, S. M., Busch, N. A. and Ohla, K. (2015) Taste quality decoding parallels taste sensations. *Current Biology* **25**, 890-896.

Davis, G. M., Mallat, S. G. and Zhang, Z. (1994) Adaptive time-frequency decompositions. *Optical Engineering*, **33**, 2183-2192.

Delorme, A. and Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* **134**, 9-21.

Dhiman, R., Saini, J. and Priyanka, A. (2010) Artifact removal from eeg recordings—an overview. *National Conference on Computational Instrumentation*, CSIO Chandigarh, India, 19-20 March 2010, 62-66.

Feng, Q., Liu, J. and Gong, J. (2015) UAV remote sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sensing* **7**, 1074-1094.

Flexer, A., Bauer, H., Pripfl, J. and Dorffner, G. (2005) Using ICA for removal of ocular artifacts in EEG recorded from blind subjects. *Neural Networks* **18**, 998-1005.

Gislason, P. O., Benediktsson, J. A. and Sveinsson, J. R. (2006) Random forests for land cover classification. *Pattern Recognition Letters* **27**, 294-300.

Huang, G., Huang, G. B., Song, S., and You, K. (2015) Trends in extreme learning machines: a review. *Neural Networks* **61**, 32-48.

Jun, S. C. and Pearlmuter, B. A. (2005) Fast robust subject-independent magnetoencephalographic source localization using an artificial neural network. *Human Brain Mapping* **24**, 21-34.

Jung, T. P., Humphries, C., Lee, T. W., Makeig, S., McKeown, M. J., Iragui, V. and Sejnowski, T. J. (1998) Extended ICA removes artifacts from electroencephalographic recordings. *Advances in Neural Information Processing Systems* **1998**, 894-900.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V. and Sejnowski, T. J. (2000) Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* **37**, 163-178.

Jutten, C. and Herault, J. (1991) Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing* **24**, 1-10.

Lagerlund, T. D., Sharbrough, F. W. and Busacker, N. E. (1997) Spatial filtering of multichannel electroencephalographic recordings through principal component analysis by singular value decomposition. *Journal of Clinical Neurophysiology* **14**, 73-82.

Mallat, S. G. and Zhang, Z. (1993) Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing* **41**, 3397-3415.

Manikandan, J. and Venkataramani, B. (2009) Design of a modified one-against-all SVM classifier. *IEEE International Conference on Systems, Man and Cybernetics* **2009**, 1869-1874.

Marc'Aurelio Ranzato, F. J. H., Boureau, Y. L. and LeCun, Y. (2007) Un-supervised learning of invariant feature hierarchies with applications to object recognition. *Proc. Computer Vision and Pattern Recognition Conference (CVPR '07)*. IEEE Press.

Martin, M., Newman, S., Aber, J. and Congalton, R. (1998) Determining forest species composition using high spectral resolution remote sensing data. *Remote Sensing of Environment* **65**, 249-254.

Melgani, F. and Bruzzone, L. (2004) Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing* **42**, 1778-1790.

Nezamfar, H., Orhan, U., Purwar, S., Hild, K., Oken, B. and Erdogmus, D. (2011) Decoding of multichannel EEG activity from the visual cortex in response to pseudorandom binary sequences of visual stimuli. *International Journal of Imaging Systems and Technology* **21**, 139-147.

Nicolaou, N. and Nasuto, S. J. (2007) Automatic artefact removal from event-related potentials via clustering. *The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology* **48**, 173-183.

O'Regan, S., Faul, S. and Marnane, W. (2013) Automatic detection of EEG artefacts arising from head movements using EEG and gyroscope signals. *Medical Engineering & Physics* **35**, 867-874.

Overton, D. and Shagass, C. (1969) Distribution of eye movement and eyeblink potentials over the scalp. *Electroencephalography and Clinical Neurophysiology* **27**, 546-546.

Pal, M. (2005) Random forest classifier for remote sensing classification. *International Journal of Remote Sensing* **26**, 217-222.

Pesin, J. (2007) Detection and removal of eyeblink artifacts from EEG using wavelet analysis and independent component analysis. MSc. Thesis. *RIT Rochester Institute of Technology*.

Qi, H., Zhu, Y., Ming, D., Wan, B., Wang, Y. and Zhang, R. (2009) Independent component automatic clustering and its application on multi-trials imaginary hand movement related EEG. *Virtual Environments, Human-Computer Interfaces and Measurements Systems. IEEE International Conference* **2009**, 149-153.

Robinson, N., Vinod, A., Guan, C., Ang, K. K. and Peng, T. K. (2011) A Wavelet-CSP method to classify hand movement directions in EEG based BCI system. *Information, Communications and Signal Processing, 8th International Conference* **2011**, 1-5.

- Safieddine, D., Kachenoura, A., Albera, L., Birot, G., Karfoul, A., Pasticu, A., Biraben, A., Wendling, F., Senhadji, L. and Merlet, I. (2012) Removal of muscle artifact from EEG data: comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches. *EURASIP Journal on Advances in Signal Processing* **2012**, 127.
- Subasi, A. and Gursoy, M. I. (2010) EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert Systems with Applications* **37**, 8659-8666.
- Vigário, R. N. (1997) Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalography and Clinical Neurophysiology* **103**, 395-404.
- Winkler, I., Brandl, S., Horn, F., Waldburger, E., Allefeld, C. and Tangermann, M. (2014) Robust artifactual independent component classification for BCI practitioners. *Journal of Neural Engineering* **11**, 035013.
- Zafar, R., Dass, S. C., Malik, A., Kamel, N., Rehman, J., Ahmad, R. F., Abdullah, J. M. and Reza, F. (2017) Prediction of human brain activity using likelihood ratio based score fusion. *IEEE Access* **5**, 13010-13019.
- Zhang, Z., Xu, Y., Yang, J., Li, X. and Zhang, D. (2015) A survey of sparse representation: algorithms and applications. *IEEE Access* **3**, 490-530.
- Zhou, J. (2014) Multicatalyst system in asymmetric catalysis. *John Wiley & Sons*.
- Zou, Y., Hart, J. and Jafari, R. (2012) Automatic EEG artifact removal based on ICA and Hierarchical Clustering. *IEEE International Conference on Acoustics, Speech and Signal Processing* **2012**, 649-652.
- Zou, Y., Nathan, V. and Jafari, R. (2016) Automatic identification of artifact-related independent components for artifact removal in EEG recordings. *IEEE Journal of Biomedical and Health Informatics* **20**, 73-81.