Automated EEG artifact elimination by applying machine learning algorithms to ICA-based features

2017 J. Neural Eng. 14 046004

(http://iopscience.iop.org/1741-2552/14/4/046004)

OPEN ACCESS
IOP Publishing
Journal of Neural Engineering
J. Neural Eng. **14** (2017) 046004 (8pp)
https://doi.org/10.1088/1741-2552/aa69d1

Automated EEG artifact elimination by applying machine learning algorithms to ICA-based features

Thea Radüntz[1], Jon Scouten[1], Olaf Hochmuth[2] and Beate Meffert[2]

[1] Federal Institute for Occupational Safety and Health, Mental Health and Cognitive Capacity, Nöldnerstr. 40-42, 10317 Berlin, Germany
[2] Department of Computer Science, Humboldt-Universität zu Berlin, Rudower Chaussee 25, 12489 Berlin, Germany

E-mail: raduentz.thea@baua.bund.de

Received 16 September 2016, revised 26 March 2017
Accepted for publication 29 March 2017
Published 12 May 2017

Abstract

Objective. Biological and non-biological artifacts cause severe problems when dealing with electroencephalogram (EEG) recordings. Independent component analysis (ICA) is a widely used method for eliminating various artifacts from recordings. However, evaluating and classifying the calculated independent components (IC) as artifact or EEG is not fully automated at present. *Approach.* In this study, we propose a new approach for automated artifact elimination, which applies machine learning algorithms to ICA-based features. *Main results.* We compared the performance of our classifiers with the visual classification results given by experts. The best result with an accuracy rate of 95% was achieved using features obtained by range filtering of the topoplots and IC power spectra combined with an artificial neural network. *Significance.* Compared with the existing automated solutions, our proposed method is not limited to specific types of artifacts, electrode configurations, or number of EEG channels. The main advantages of the proposed method is that it provides an automatic, reliable, real-time capable, and practical tool, which avoids the need for the time-consuming manual selection of ICs during artifact removal.

Keywords: electroencephalography (EEG), biomedical signal processing, artifact elimination, independent component analysis, image filtering, support vector machines (SVM), artificial neural networks (ANN)

1. Introduction

An electroencephalogram (EEG) provides information about the neuronal activity in the brain, and thus the mental state of a person. EEG analysis is a widely used tool with a number of applications. EEGs are very important for investigating the functioning of the human brain, but users should be aware that they are affected greatly by artifacts. The causes of biological artifacts may include eye movements, blinks, muscle activity, and the heartbeat. Technically induced artifacts can be caused by faulty electrodes, high electrode impedance, line noise, and interference from electric devices. The correct interpretation of EEG traces is only possible if all of these artifacts are eliminated before their analysis.

The previously proposed methods for artifact rejection have various disadvantages. Discarding contaminated EEG segments based on visual and manual inspections can lead to severe losses of experimental data, and these approaches are completely inadequate when working with continuous EEG in brain-computer interface applications or online mental

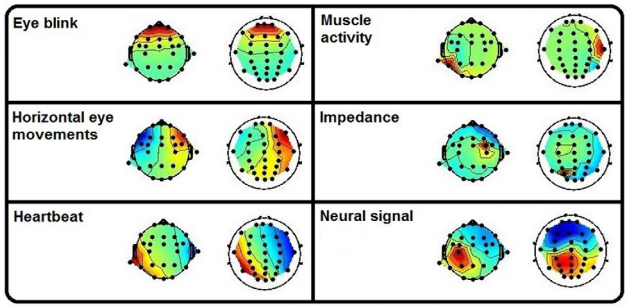

Figure 1. Typical topoplot image patterns of different artifact types.

state monitoring [1]. Independent component analysis (ICA) is a widely used method for eliminating various artifacts from recordings [2], where the artifact components are then discarded and the EEG components are projected back, thereby reconstructing an artifact-free signal. However, the process of evaluating and classifying the calculated independent components (ICs) as artifact or EEG is not fully automated at present. Indeed, the previously proposed methods can introduce new artifacts into EEG recordings [3–5] or concentrate on specific artifact types [6, 7]. Thus they are unsuitable for real-time applications [8].

To overcome most of these problems, we developed a new method for artifact elimination based on fully automated IC classification. This new method was inspired by observations of experts during visual examinations of topoplots, i.e. two-dimensional scalp map projections obtained after the application of ICA to EEG recordings. Hence, in order to evaluate the EEG quality after artifact removal, we compared the agreement between the component ratings given by experts and the automated ratings obtained using our method. Thus, we obtained direct information about the artifact removal quality of our method.

In a previous study [9], we described some preliminary research into the extraction of ICA-component features using image processing algorithms, where we evaluated several operators in terms of their ability to characterize the typical topoplot image patterns of different artifacts (figure 1, see [2]). A comparison of 13 different image processing operators showed that the best performance was obtained using the so-called feature images computed by applying the horizontal Sobel operator (HSO), large horizontal gradient operator (LHG), range filter, second layer of the three-dimensional local binary pattern (3D LBP (2nd)), and Gaussian curvature.

Using these six preselected features we aimed to establish a complete system for artifact elimination, which is fully automatic with real-time capability. In this study, we describe our approach to achieving this goal. In the proposed method, we identify the most suitable feature image and combine it with the typical EEG frequency bands obtained from the power spectrum of the IC. We give details of the procedures required to select a classifier and we test and compare several machine learning algorithms for classifying artifacts. Finally, we identify the classifier with the highest agreement rates relative to a human expert. We benchmark our findings based on comparisons with the existing removal methods called ADJUST [10] and CORRMAP [11].

2. Material and experiments

The investigations were performed in the shielded laboratory at the federal institute for occupational safety and Health in Berlin. EEG traces were captured by 25 electrodes, which were arranged according to the 10–20 system, with reference to Cz and at a sample rate of 500 Hz. The recorded signal lengths varied between 1.5 and 20 min. The sample comprised 57 people (aged between 30 and 62 years, with 31 females and 26 males). During the experiment, the participants had to solve cognitive tasks with varying degrees of difficulty. We described the details of the experiment in a previous study [9].

To ensure a thorough validation, we also tested our system with two additional data sets: one similar to the setup described above, and another obtained with a substantially different electrode configuration and experimental design (www.baua.de/de/Forschung/Forschungsprojekte/f2247.html?nn=2799254). Table 1 provides an overview of the data sets used.

All of the studies were approved by the local review board at our institution and the experiments were conducted in accordance with the Declaration of Helsinki. All of the procedures were conducted with the adequate understanding and written consent of the subjects.

3. Methods

3.1. Principle

The pipeline for EEG artifact elimination comprises three main modules: pre-processing, feature generation, and classification.

Filtering and ICA are performed in the pre-processing module. We applied a band pass filter in the order of 100 to the raw signals. The cut-off frequencies were 0.5 and 40 Hz. We then decomposed the multi-channel EEG into ICs using the Infomax algorithm [2, 12]. All of the ICs (equal to the number of channels) were used in the following computations.

We should note one of the main concepts in our method, which is the interpolation of the ICA mixing matrix onto a 51×63 fixed-size grid using the inverse distance method described by [13]. This step aims to generate images with identical dimensions despite any original differences in the number of EEG channels and the positions of the electrodes. A different number of electrodes would actually result in variable column lengths of the mixing matrix and hence require retraining of the classifier. The projection of ICs onto the same grid allows the classification of the topoplot images without retraining the classifier in any further investigations using different number of channels and positions. Thus, our classifier always has the same number of inputs (i.e. pixel in the image) and classifies the image patterns of the topoplots.

In the next step, the topoplot images obtained together with the IC power spectra are used as inputs for the feature generation module. This module comprises image processing algorithms for feature extraction from the topoplots to obtain so-called feature images with the same size as the original topoplot images.

Table 1. Data sets used for classifier training and testing (arranged according to the 10–20 system).

	Subjects	Tasks	Electrodes	Reference	Sample rate
Data set 1 (training, testing)	57	11	25	Cz	500 Hz
Data set 2 (testing)	12	12	25	Cz	500 Hz
Data set 3 (testing)	10	1	63	FCz	500 Hz

In order to emphasize the key specific properties of the topoplots for analysis, we used six preselected operators. Thus, by applying HSO, range filtering, and LHG, we preserved the extent and strength of the gradients in the original topoplot in the feature image. The feature images obtained from the second layer of 3D LBP characterized the texture. Finally, the Gaussian curvature calculation yielded feature images containing information about geometrical forms. Furthermore, we used the raw topoplots in our analysis in order to benchmark our findings.

Starting with the same topoplot image (e.g. the topoplot at the top left in figure 2) each operator computed a different feature image, examples of which are illustrated in figure 2. Hence, one of our main aims was to define the most suitable operator for discriminating between the topoplot patterns according to their type.

Our previous results [9] only indicated a weak influence of downsampling the feature images on the system's accuracy. Therefore, we used feature images with 372 pixels instead of the original 3213 in order to optimize the computational time.

Subsequently, we combined the feature images with the frequency bands obtained from the IC's power spectrum into feature vectors and used them as inputs for the classification module. The classification module only required the feature images and IC frequency bands as inputs to produce the classification result, which comprised either an artifact or EEG component as the output. Artifact components were discarded and EEG components were projected back, thereby reconstructing an artifact-free signal.

3.2. Classification

3.2.1. Classification methods. In general, machine learning can be described as the generation of knowledge from experience by a system. This means that the system receives examples of an actual situation and rather than merely memorizing them, it determines common principles among examples from the same class. After a training phase, the system is able to generalize and assign new events to the predefined classes.

In addition to using linear discriminant analysis (LDA) for selecting the most suitable features [9], we employed logistic regression (LgR) as a binary classifier as well as support vector machines (SVMs) and artificial neuronal networks (ANNs), which were implemented and trained as follows.

LgR. LgR is similar to LDA a linear classifier, i.e. it identifies a linear decision boundary between the data. In LgR, the outliers are only given a small loading, and thus they affect the rating little. Hence, LgR is assumed to be a more general approach. By contrast, LDA includes outliers when computing the covariance matrix, which makes this method more precise because of the additional information, but this also reduces its robustness against large outliers.

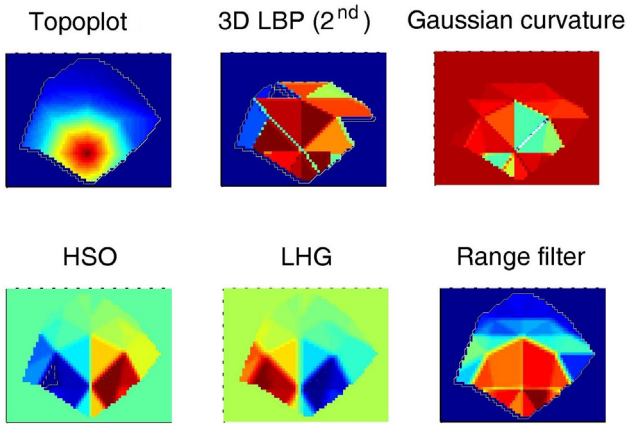

Figure 2. Examples of the feature images used. All of the feature images were generated from the top left topoplot image using several operators (horizontal Sobel operator (HSO), large horizontal gradient operator (LHG), range filter, second layer of the 3D local binary pattern (3D LBP (2nd)), and Gaussian curvature).

SVMs. SVMs are used widely as so-called large margin classifiers. A characteristic of SVMs is that they attempt to classify objects into classes with the maximum possible object-free area around them. Each object is represented by a vector in a vector space. An SVM searches for a hyperplane that separates the classes in this space. Depending on the SVM kernel used, the data separation process can be designed as linear or nonlinear. In nonlinear separation, the vector space and its objects are transformed into a higher dimensional space, which allows linear separation using a plane. After changing back to a lower dimensional space, the linear hyperplane becomes nonlinear and it can even be discontinuous [14, 15]. The preferred nonlinear kernel is the Gaussian radial basis function, which we used in this study.

ANNs. ANNs were first proposed in the early 1940s and they are very valuable in applications where there is little knowledge of the problem. The architecture of an ANN is defined by the number of layers, the number of particular neurons (nodes), and how they are connected with each other (edges). ANNs contain input and output layers as well as one or several hidden layers. The number of hidden layers is crucial for the network's structure. In this study, the ANN comprised a one-layer network trained by back-propagation [16] using 100 iterations.

3.2.2. Approach for classifier selection. A number of steps are required to select a classifier. Firstly, it is necessary to identify a suitable subset of features, before training the different

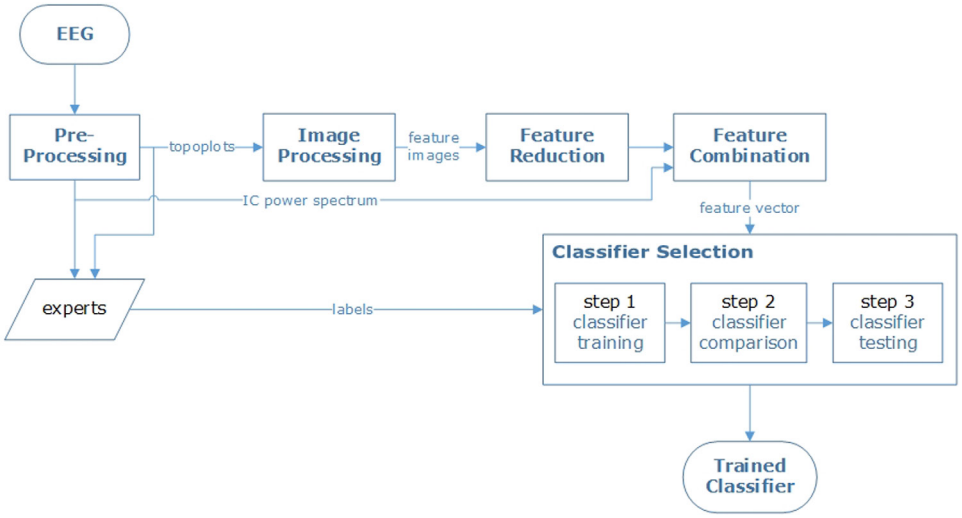

Figure 3. Approach for classifier selection.

classifiers, and then comparing and testing them to select the classifier with the best performance (figure 3).

Six suitable feature images were selected using LDA, as described in detail by [9]. Thus, LDA-based identification of the best feature images employed a data volume comprising 625 sets (subjects × tasks) with 25 topoplots in each set. Table 2 provides an overview of the data used. The six selected feature images (HSO, LHG, range filter, 3D LBP (2nd), Gaussian curvature, and raw topoplots) achieved an LDA accuracy rate of around 85%.

However, experts do not rely only on the patterns in the topoplots because they also consider the frequency bands of IC activation when making their judgments. Hence, it was necessary for the fully automated artifact elimination method to simulate the behavior of experts by integrating the IC power spectra in the classification process. For the features described in the following, we always refer to the feature images combined with their IC power spectra.

The selection of a classifier is divided into three steps: (1) classifier training and identifying the optimal parameters, (2) classifier comparison, and (3) classifier testing. In step 1, we used 60% of data set 1, and we used 20% in each of steps 2 and 3. The results obtained are described in the next section.

4. Results

We evaluated our artifact removal algorithm based on a thorough inspection of the classification results and the execution time. The classification results satisfied the standard for assessing classifier performance, thereby obtaining a good quality signal, but the execution time determines the real-time feasibility of the algorithm.

4.1. Classifier training

From among 60% of the subjects, we randomly selected a subset of 80% for training (determining the most suitable parameters) each classifier and 20% for testing. Step 1 involved tuning the classifiers. For each classification method, each feature and each parameter were subjected to a cross-validation (5 × 4). The average results obtained are presented in figure 4.

Table 2. Amounts of data used for selecting suitable feature images based on LDA.

	Proportion of total set size	Number of sets	Number of feature images
Training	60%	375	9375
Testing	40%	250	6250

Based on the curves obtained, we empirically selected the best parameter for each classification method and feature (table 3). The classifier tuning results are listed in table 4.

4.2. Comparison of classifiers

Based on the results obtained in step 1, the accuracy rates for all features were worse using LDA than the other classification methods (table 4), so we removed it from the computations in step 2.

The results obtained by LgR, SVM, and ANN for each feature are presented in table 5. The best recognition result of 95.85% was achieved using the range images combined with ANN, followed by the range images with SVM (94.04%), and the combination of ANN with the second layer of the 3D LBP (94.07%). LgR was particularly robust where the recognition rates for all features were between 92.7% and 93.4%. Thus, we hypothesize that the data including outliers should be given a smaller loading during training so they had less influence on the rating.

Thus, range images and classification with SVM and ANN were the most suitable for automated artifact elimination.

4.3. Classifier testing

In step 3, we tested the selected classifiers (SVM and ANN) with the selected feature (range image) using the remaining 20% of data set 1. We also evaluated our method in further tests with data set 2 and data set 3. To determine the real-time feasibility of the algorithm, we examined the execution time for each module. The results are described in the following.

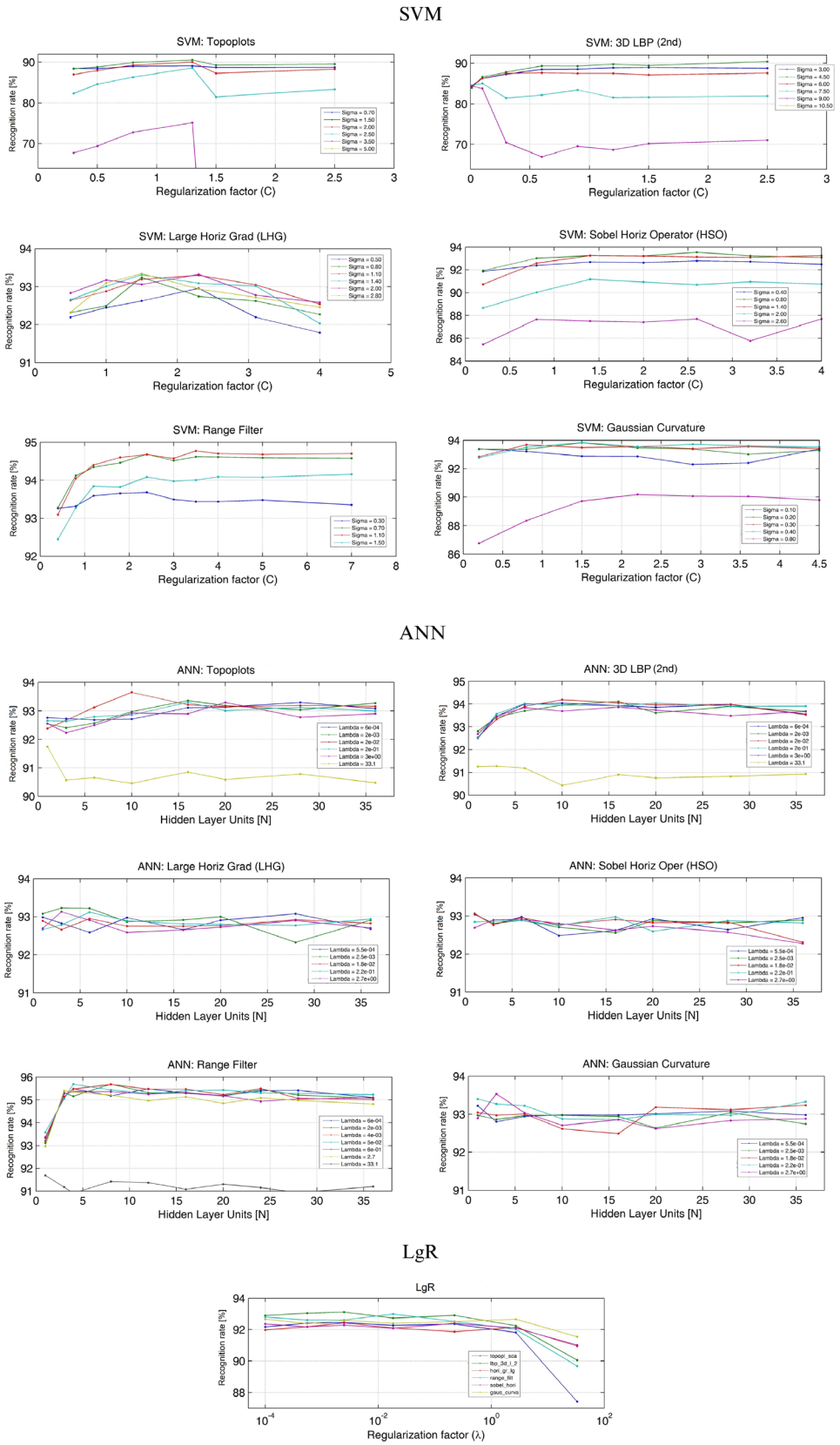

Figure 4. Recognition rates for the empirical selection of the optimal parameters with each classification method and feature.

4.3.1. Recognition rates. In step 3, the classifier results were compared with the ratings given by the expert, whose labels were used to train the classifiers (expert 1), but also with the ratings given by two additional experts (expert 2 and expert 3). The test results and the agreement among the experts are presented in table 6. As expected, the agreement between the machine ratings and the ratings given by expert 1 was high. The ANN algorithm achieved the best performance of

Table 3. Optimal parameter for each classification method and feature (N: number of eigenvectors, λ: regularization parameter, C: SVM regularization, σ: kernel width, and units: number of neurons in the hidden layer).

IC power spectra &	LDA (N-1)	LgR (λ)	SVM (C/σ)	ANN (λ/units)
Topoplots	376	$2.5 \cdot 10^{-3}$	1.3 / 1.5	$1.8 \cdot 10^{-2}$ / 10
3D LBP (2nd)	376	$2.5 \cdot 10^{-3}$	2.5 / 4.5	$1.8 \cdot 10^{-2}$ / 10
LHG	376	$2.5 \cdot 10^{-3}$	1.5 / 2.8	$2.5 \cdot 10^{-3}$ / 3
HSO	376	$2.2 \cdot 10^{-1}$	2.6 / 0.8	$5.5 \cdot 10^{-4}$ / 4
Range filter	376	$1.8 \cdot 10^{-2}$	3.5 / 1.1	$5.0 \cdot 10^{-2}$ / 3
Gaussian curvature	376	2.7	1.5 / 0.4	2.7 / 3

Table 4. Mean recognition rate (%) and standard deviation (in parentheses) for each feature obtained by classifier tuning to determine the optimal parameters.

IC power spectra &	LDA	LgR	SVM	ANN
Topoplots	86.9(1.1)	92.3(1.3)	90.4(0.8)	93.2(1.0)
3D LBP (2nd)	84.3(1.4)	92.9(0.8)	90.8(1.5)	94.1(0.7)
LHG	89.3(1.0)	92.3(1.0)	93.3(1.1)	92.8(1.2)
HSO	89.7(1.1)	92.2(0.8)	93.4(0.7)	92.7(0.9)
Range filter	90.9(1.3)	92.8(1.1)	94.6(1.0)	95.5(0.9)
Gaussian curvature	90.5(0.7)	92.5(1.3)	93.7(1.2)	93.1(0.9)

Table 5. Validation results (%) for the classifiers and each feature.

IC power spectra &	LgR	SVM	ANN
Topoplots	93.13	89.53	92.55
3D LBP (2nd)	92.65	89.13	94.07
LHG	92.91	92.62	92.76
HSO	93.16	93.78	92.91
Range filter	93.02	94.04	**95.85**
Gaussian curvature	93.42	93.31	93.45

95.42% compared with expert 3 and very good agreement of over 90% with expert 2. The level of agreement between the machine and expert 2 has comparatively more weight considering this experts agreement of 92.84% to expert 1 and 90.89% to expert 3. Furthermore, there was a small but consistent advantage of ANN compared with SVM.

Furthermore, we split the results obtained by our classifiers based on the visual determinations by expert 1 into four different artifact groups: eye blink, horizontal eye movements, heartbeat, and others. For convenience, the experts did not differentiate between impedance and muscle activity. Table 7 shows the classifier performance for each group, which shows that the performance was similar across groups, i.e. the result of our method is not dependent on the artifact type

To test the system with data set 2, expert 3 visually inspected and classified the ICs. The agreement rates between the machine and expert are listed in table 8. Both classifiers achieved recognition rates over 90%. Again, ANN was superior to the SVM with a recognition rate of 95.31%.

Table 6. Agreement (%) among the experts as well as between the experts and the two classifiers for data set 1.

	SVM	ANN	Expert 1	Expert 2	Expert 3
Expert 1	93.51	95.20	—	92.84	95.15
Expert 2	90.62	91.71	92.84	—	90.89
Expert 3	92.87	95.42	95.15	90.89	—

Table 7. Agreement (%) between an expert and two classifiers regarding different artifact types in data set 1.

	SVM	ANN
Eye blink	100	100
Horizontal eye movements	94.81	93.51
Heartbeat	96.7	93.51
Others	96.53	90.68
Neural signal	96.31	98.1

Table 8. Agreement (%) between an expert and two classifiers using data set 2.

	SVM	ANN
Expert	93.25	95.31

Table 9. Agreement (%) among the experts as well as between experts and the two classifiers for data set 3 (different electrode numbers and configurations).

	SVM	ANN	Expert 1	Expert 2	Expert 3
Expert 1	86.19	91.43	—	93.97	93.81
Expert 2	81.43	86.35	93.97	—	90.00
Expert 3	85.40	90.63	93.81	90.00	—

Finally, all three experts visually inspected and classified the ICs in data set 3. According to the results obtained using data set 2, we expected better performance with the trained ANN classifier. The agreement rates between the machine and experts are listed in table 9. The ANN achieved the best recognition rate of 91.43%. The agreement among the experts varied between 90.00% and 93.97%. The recognition rates using the SVM classifier for range images varied between 81.43% and 86.19%.

4.3.2. Execution time. To determine the feasible real-time applicability of the proposed algorithm, we tested our system on an intel core i5-3320M processor (2.6 GHz) with 8 GB DDR3-SDRAM (2×4 GB). The input signal was a 25 channel EEG with a length of 94.34 s (500 Hz sample rate). We selected a short signal because the computational time required for ICA depends on the signal length and it is the most time-consuming operation in the processing pipeline. Hence, the system's execution time can be viewed as a bottleneck. In addition, it is necessary to consider that identifying N components requires more than kN^2 data sample points (N denotes the number of EEG channels, N^2 is the number of weights in the ICA unmixing matrix, and k is a multiplier that increases with the number of channels).

Hence, in our study, the pre-processing module required approximately 5 s for filtering, performing ICA, and generating a topoplot. The computation of one feature image

Table 10. Execution times for the processing pipeline.

Pre-processing	Band pass filter	0.164 s
	ICA	4.731 s
Feature generation	Topoplot	0.121 s
	Range filter	0.002 s
	Feature postprocessing	0.000 07 s
	IC power spectra	0.013 s
Classification (online testing)	SVM—range image	$5.6 \cdot 10^{-4}$ s
	ANN—range image	$9.6 \cdot 10^{-6}$ s

(here range filtering) and the corresponding IC power spectra required only 15.7 ms. The classification of one component using the SVM and ANN required less than 1 ms. Table 10 shows the execution time for each module, which demonstrates that the system is real-time capable. System training can be performed offline so it was not considered in the real-time performance evaluation.

5. Conclusions

In this study, we developed a new method for artifact elimination, which can reject any type of artifact from EEG traces. Our method is better than the currently available methods because it is not restricted to certain types of artifact, e.g. blinks, and it can run automatically without any user interaction. Furthermore, it is not limited to specific numbers and positions for the electrodes, and the system needs to be trained only once. Hence, it behaves similarly to human experts during the rating process of topoplots that is also independent of the electrode configuration because of the similar image patterns in the topoplots of each artifact type.

However, ICA does not require a specific montage, and it only demands the independence and linear co-dependence of the channels, but the localization and thus the topoplot patterns may be more accurate when the electrode coverage is more uniform [17]. This was obvious when experts were asked to rate topoplots obtained from a small number of electrodes or based asymmetric electrode configurations. The classification process by humans and machine lacked clarity. Thus, there may be more appropriate methods for artifact rejection with a smaller number of electrodes or asymmetric head coverage, e.g. [18–20].

Our novel approach for real-time and fully automated artifact elimination achieved recognition rates between 89.13% and 95.20%, where the best recognition performance was obtained for features derived from range images and IC power spectra combined with ANN.

To the best of our knowledge, no other method has comparable performance. Finally, in order to provide a relative performance rating for our new approach, we benchmarked our algorithm against the existing removal methods called ADJUST [10] and CORRMAP [11], which are implemented in EEGLAB.

We tested both methods with 250 ICs. CORRMAP only focuses on eye blinks and lateral eye movements, as mentioned by the authors [11], but it can be used in an automatic mode. The recognition rate obtained indicated 62% agreement with the ratings of our experts. ADJUST yielded an accuracy rate of 57.15% compared with the ratings of our experts. It may be assumed that these percentages correspond roughly to the percentage of eye blinks and lateral eye movements among the total artifacts.

Our proposed method could be improved by including further experts. For example, it might benefit from a crowd-sourced approach to collecting more IC labels for training the classifier more precisely. Recently, the swartz center for computational neuroscience launched an internet project asking researchers to label as many components as possible for a machine learning method. Including this information in our method could lead to more accurate results.

The real-time feasibility of our method is an additional advantage. The time-consuming computation of the ICA, which is the bottleneck in our system in terms of the execution time, could be improved by considering recent developments in online ICA [21–23].

Acknowledgments

We would like to thank Ms Xenija Weißbecker-Klaus, Mr Robert Sonnenberg, Dr Sergei Schapkin, and Ms Marion Freyer for conducting the laboratory experiments and acquiring the data. Furthermore, we would like to thank Ms Marion Freyer and Ms Friederice Schröder for daily operational support, Ms Dagmar Rahim for proofreading, and Dr Gabriele Freude for her general support.

More information about the project that acquired our EEG data can be found at www.baua.de/de/Forschung/Forschungsprojekte/f2312.html?nn=2799254.

Author contributions

TR initiated the project and was responsible for the overall conception of the method. JS provided computational support for signal processing, classifier training, and testing. The study was supervised by TR. Data analysis and interpretation were performed by TR, OH, and BM. The manuscript was written by TR. Final critical editing was performed by OH and BM

References

[1] 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 *Psychophysiol.* **37** 163–78

[2] Makeig S, Bell A J, Jung T P and Sejnowski T J 1996 Independent component analysis of electroencephalographic data *Advances in Neural Information Processing Systems* (Cambridge, MA: MIT Press) pp 145–51

[3] Weerts T C and Lang P J 1973 The effects of eye fixation and stimulus and response location on the contingent negative variation (CNV) *Biol. Psychol.* **1** 1–19

[4] Oster P J and Stern J A 1980 Measurement of eye movement electrooculography *Techniques in Psychophysiology* ed I Matin and P H Venables (New York: Wiley) pp 275–309

[5] Peters J F 1967 Surface electrical fields generated by eye movements *Am. J. EEG Technol.* **7** 27–40

[6] Gao J F, Yang Y, Lin P, Wang P and Zheng C X 2010 Automatic removal of eye-movement and blink artifacts from EEG signals *Brain Topogr.* **23** 105–14
[7] Gao J, Lin P, Yang Y, Wang P and Zheng C 2010 Real-time removal of ocular artifacts from EEG based on independent component analysis and manifold learning *Neural Comput. Appl.* **19** 1217–26
[8] Fatourechi M, Bashashati A, Ward R K and Birch G E 2007 EMG and EOG artifacts in brain computer interface systems: a survey *Clin. Neurophysiol.* **118** 480–94
[9] Radüntz T, Scouten J, Hochmuth O and Meffert B 2015 EEG artifact elimination by extraction of ICA-component features using image processing algorithms *J. Neurosci. Methods* **243** 84–93
[10] Mognon A, Jovicich J, Bruzzone L and Buiatti M 2011 ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features *Psychophysiol.* **48** 229–40
[11] Viola F C, Thorne J, Edmonds B, Schneider T, Eichele T and Debener S 2009 Semi-automatic identification of independent components representing EEG artifact *Clin. Neurophysiol.* **120** 868–77
[12] Delorme A and Makeig S 2004 EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics *J. Neurosci. Methods* **134** 9–21
[13] Sandwell D T 1987 Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data *Geophys. Res. Lett.* **14** 139–42
[14] Burges C J C 1998 A tutorial on support vector machines for pattern recognition *Data Min. Knowl. Discovery* **2** 121–67
[15] Bishop C M 2006 Pattern recognition and machine learning *Information Science and Statistics* ed M Jordan *et al* (Berlin: Springer) ISBN: 978-0-387-31073-2
[16] Rumelhart D E, Hinton G E and Williams R J 1986 Learning representations by back-propagating errors *Nature* **323** 533–6
[17] Akalin Acar Z and Makeig S 2013 Effects of forward model errors on EEG source localization *Brain Topogr.* **26** 378–96
[18] Graichen U, Eichardt R, Fiedler P, Strohmeier D, Zanow F and Haueisen J 2015 SPHARA—a generalized spatial fourier analysis for multi-sensor systems with non-uniformly arranged sensors: application to EEG *PLoS One* **10** 1–22
[19] Gao J, Yang Y, Sun J and Yu G 2010 Automatic removal of various artifacts from EEG signals using combined methods *J. Clin. Neurophysiol.* **27** 312–20
[20] Gao J, Zheng C and Wang P 2010 Online removal of muscle artifact from electroencephalogram signals based on canonical correlation analysis *Clin. EEG Neurosci.* **41** 53–9
[21] Pion-Tonachini L, Hsu S H, Makeig S, Jung T P and Cauwenberghs G 2015 Real-time EEG source-mapping toolbox (REST): online ICA and source localization *37th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* pp 4114–7
[22] Hsu S H, Mullen T, Jung T P and Cauwenberghs G 2014 Online recursive independent component analysis for real-time source separation of high-density EEG *36th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society* pp 3845–8
[23] Akhtar M T, Jung T P, Makeig S and Cauwenberghs G 2012 Recursive independent component analysis for online blind source separation *IEEE Int. Symp. on Circuits and Systems* pp 2813–6