Improving classification accuracy using intra-session classifier training and implementation for a BCI based on automated parameter selection

Chanan S. Syan* and
Randy E.S. Harnarinesingh

Department of Mechanical and Manufacturing Engineering,
Faculty of Engineering,
The University of the West Indies,
St. Augustine, Trinidad & Tobago,
Email: chanan.syan@sta.uwi.edu
Email: randy.harnarinesingh@gmail.com
*Corresponding author

Ramaswamy Palaniappan

School of Computer Science and Electronic Engineering,
University of Essex,
Wivenhoe Park, Colchester,
CO4 3SQ, UK
Email: rpalan@essex.ac.uk

Abstract: Genetic Algorithms (GAs) were used in a previous study to automate parameter selection for an EEG-based P300-driven Brain-Computer Interface (BCI). The GA approach showed marked improvement over data-insensitive parameter selection; however, it required lengthy execution times thereby rendering it infeasible for online implementation. Automated parameter selection is retained in this work; however, it is achieved using the less computationally intensive N-fold cross-validation (NFCV). Additionally, this study sought to improve BCI classification accuracy using a training data collection and application protocol that the authors refer to as 'Intra-session classifier training and implementation'. Intra-session classifier training and implementation using NFCV-driven automated parameter selection yielded a classification accuracy of 82.94% compared to 45.44% for the inter-session approach using data-insensitive parameters. These findings are significant impact since the intra-session protocol can be applied to any P300-based BCI regardless of its application platform to obtain improved classification accuracy.

Keywords: P300; BCI; FLDA; genetic algorithm; N-fold cross-validation; automation.

Reference to this paper should be made as follows: Syan, C.S., Harnarinesingh, R.E.S. and Palaniappan, R. (2012) 'Improving classification accuracy using intra-session classifier training and implementation for a BCI based on automated parameter selection', *Int. J. Intelligent Systems Technologies and Applications*, Vol. 11, Nos. 1/2, pp.36–48.

Biographical notes: Chanan S. Syan is presently the Head of Production Engineering and Management, Leader of Graduate Programmes and Professor at the University of the West Indies. He obtained his graduation from the

University of Bradford, UK, in 1983 with a BEng (Hons) in Mechanical Engineering. In 1988 he obtained a doctorate from the University of Hull, UK. He has over 12 years of industrial experience and over 25 years in academia at all levels. Specialisations include – brain computer interfaces, manufacturing, design and manufacture and asset management. He has lead research, managed successful research projects and published widely internationally in all forms of learned publications.

Randy E.S. Harnarinesingh is currently pursuing the PhD in Mechanical Engineering from the University of the West Indies at St. Augustine, Trinidad. He obtained his graduation with Electrical and Computer Engineering (First Class Honours) from the University of the West Indies at St. Augustine, Trinidad, in 2008. His post-graduate research is in the area of brain-computer interfaces and autonomous robotics.

Ramaswamy Palaniappan is currently an Academic at the School of Computer Science and Electronic Engineering, University of Essex, UK. He is a senior member of the Institute of Electrical and Electronics Engineers and IEEE Engineering in Medicine and Biology Society, member in Institution of Engineering and Technology, and Biomedical Engineering Society. He also serves as editorial board member for several international journals. His pioneering work on BCI has received international recognition. His current research interests include biological signal processing, brain-computer interfaces (BCI), biometrics, artificial neural networks, genetic algorithms and image processing. To date, he has published over 100 papers in peer-reviewed journals, book chapters and conference proceedings.

This paper is a revised and expanded version of a paper entitled 'Automation of pre-processing and feature extraction parameter selection for a single-trial P300-based brain-computer interface using a genetic algorithm' presented at the '26th International Conference of CAD/CAM, Robotics & Factories of the Future (CARs&FOF)', Kuala Lumpur, Malaysia, 26–28 July 2011.

1 Introduction

A Brain-Computer Interface (BCI) is a system that permits users to control external devices using only their inherent brain activity (Kubler et al., 2006). In the BCI paradigm, users are mandated to perform a physical or mental task that signifies their intended command. For example, a BCI user may focus on a directional arrow presented on a computer screen in order to control a mouse cursor. The performance of representative mental and physical tasks in the BCI paradigm results in the distinguishable modulation of a user's brain activity features. The BCI, through a recording modality such as Electroencephalography (EEG), identifies this brain feature modulation and uses it to classify the user's command.

Common BCI application platforms are cursor control (Trejo et al., 2006), spelling systems (Wills and MacKay, 2006) and wheelchair navigation (Rebsamen et al., 2007; Pires et al., 2008). These applications highlight the practical utility of the BCI technology for the disabled community since they allow environmental interaction and communication without the need for pre-existing neuromuscular capabilities. The features of brain activity commonly used for BCI are Visually Evoked Potentials (VEP) (Lin et al., 2006;

Citi et al., 2008), sensorimotor rhythms (Peters et al., 2001; Krusienski et al., 2007) and Slow Cortical Potentials (SCP) (Birbaumer et al., 2003). VEPs can be classified into either the Steady State Visually Evoked Potential (SSVEP) or the P300 VEP. The SSVEP is manifested as a spike in the EEG's power spectrum at the flashing frequency of a stimulus to which the user is observing. The P300 appears as an electropositive peak in the EEG around 300–600 ms following the observation of a rare, target or deviant stimulus. In general, the P300 occurs in the 0–8 Hz portion of the frequency spectrum.

For the purposes of BCI, the P300 response is evoked using an oddball paradigm. The oddball paradigm is a stimulus presentation scheme in which a target stimulus is delivered amongst more frequently occurring non-target stimuli. In a P300-based BCI, the target stimulus is the stimulus which represents the user's command. The other stimuli, which represent other possible commands, are considered to be non-target. Therefore, the EEG segment following the presentation of the stimulus which encodes the user's command contains the P300. Consequently, the user's command can be identified through the inspection of post-stimulus EEG segments for the P300 response.

Pre-processing is implemented in BCI systems in order to attenuate unwanted signal artefacts that can degrade classification accuracy. Feature extraction is applied prior to classification to reduce the computational cost of classification and to avert the Hughes phenomena associated with high-dimensional training data. Both the pre-processing and feature extraction stages of BCI operation entail parameter selection, for example filter cut-off frequencies. Parameter selection is usually based on human expertise. However, the optimal parameter selection strategy for any given BCI arrangement depends on recording circumstances such as cap placement and subject mood which are difficult to measure and quantify. This has the ability to make a general parameter selection strategy made by an expert suboptimal. In this light, it is advantageous to select parameters in an unbiased manner without the associated subjectivity of a human expert. This study explores two avenue of data-driven parameter selection, namely Genetic Algorithms (GA) and N-fold Cross-Validation (NFCV).

Identifying the user's command from the collected EEG is achieved using pattern classification. Acquiring the calibration data necessary to tune a BCI to its user is done by means of a training session (Hoffmann et al., 2008). In a training session, the user performs a series of tasks that are selected by the BCI. This allows for the acquisition of labelled training examples which are used for the supervised training of a machine learning classifier. The BCI classifier obtained in this manner is implemented in a different session to identify subject commands. This classifier training protocol is hereafter referred to as Inter-session classifier training. In these subsequent sessions, the subjects commands are not known to the BCI and can therefore be used to obtain an unbiased measure of classifier accuracy. However, this approach is problematic.

A classifier obtained from a training session is not just a function of the underlying brain signals on which the BCI is based but also the experimental circumstances of that session such as electrode placement, electrode impedances, subject mood, subject attentiveness and external noise. These variables are difficult to measure, control and therefore repeat. Implementing a classifier derived from a training session in a subsequent session is suboptimal with regard to classification accuracy since the classifier free parameters are likely to be fitted to the experimental circumstances of the training session. One way to address this issue of suboptimality is to train and implement a classifier in the same session. This has the ability to significantly improve BCI classification accuracy. Command misclassifications are detrimental to BCI operation since they either result in the execution of the wrong command or they require corrective action which is time consuming. This paper investigates intra-session classifier training as a basis for improving P300 classification accuracy compared to the commonly used inter-session approach.

2 Materials and methods

Four male post-graduate students (S1–S4) between the ages of 18 and 25 each participated in four sessions of data collection. Subjects S1 and S3 were paid participants and subjects S2 and S4 were unpaid volunteers. Subjects S1, S2 and S3 attended the University of Essex at Colchester, UK. Subject S4 attended the University of the West Indies at St. Augustine, Trinidad.

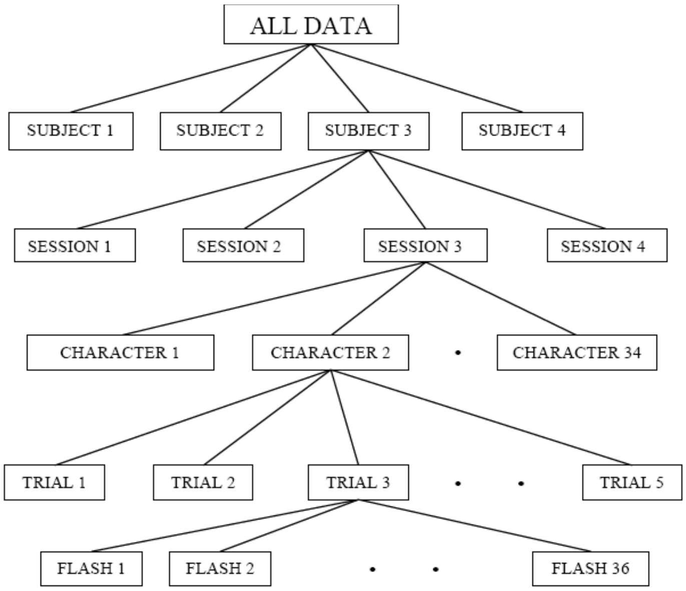
2.1 Data set description: stimulus presentation paradigm and recording circumstances

EEG was recorded at 32 sites using the BioSemi™ ActiveTwo system operating at a recording frequency of 256 Hz. The Graphical User Interface (GUI) used for the Stimulus Presentation Paradigm (SPP) was coded in MATLAB® 7.8.0.347.

The SPP is presented in Figure 1. The user is presented with a square matrix of 36 alphanumerical characters. This matrix is referred to as the P300 Speller Matrix (Farwell and Donchin, 1981). The user's task is to spell the phrase 'QUICK BROWN FOX JUMPED OVER THE LAZY DOG' in a sequential fashion using the Speller Matrix. The current character to be spelt, referred to as the target character, is explicitly identified to the user by being highlighted for 2 seconds. The user is instructed to count the number of times the target character is flashed in the subsequent sequence of stimulus presentation. Each character of the matrix is subsequently flashed in a random sequence for 100 ms. A 75 ms blank period, in which no stimulus is flashed, follows the highlighting of each stimulus. This 75 ms period is known as the Inter-Stimulus Interval (ISI). The flashing of all 36 elements of the matrix is known as a trial. Five trials of stimulus presentation are performed for each target letter. The presentation of the total 34 characters is referred to as a session. The hierarchical structure of the resulting total data set across subjects S1–S4 is expressed in Figure 2.

Figure 1 Single-character P300 speller stimulus presentation paradigm

Figure 2 Hierarchical structure of data set



2.2 Pre-processing and feature extraction

The 32 channel data set is reduced to an eight-channel subset using the eight EEG channels P7, P3, P4, P8, FZ, CZ, PZ and OZ. These channels are known to strongly contain the P300 signal (Hoffmann et al., 2008). Smaller channel sets are beneficial for BCI operation because they result in lower dimensional feature sets which lessen the computational load of classifier training and implementation in addition to making the onset of the Hughes Phenomena (Evangelista et al., 2006; Wu et al., 2009) less likely. Each EEG channel was forward-reverse filtered using an n-th order digital Butterworth filter with 3DB cut-off frequencies at 1.5 Hz and 12 Hz. 'n' is taken as 3 for the tests that require data-insensitive parameter selection however 'n' is treated as a free parameter to be optimised using NFCV for the automation of data-driven pre-processing parameter selection.

Each 500 ms post-stimulus interval is down-sampled by a factor of 'K' in order to produce a single temporal feature set per channel. 'K' is taken as 8 for the tests that require data-insensitive parameter selection however 'K' is treated as a free parameter to be optimised for data-driven feature extraction parameter selection. The temporal feature sets relating to each channel are then concatenated across the eight selected EEG channel to create a single spatiotemporal feature vector relating to each 500 ms post-stimulus EEG segment.

2.3 Classification

The collected spatiotemporal training examples belong to two categories: one contains spatiotemporal feature vectors relating to target characters; and other contains spatiotemporal feature vectors relating to non-target characters. This results in a two-class classification problem for training. The training examples are used to train a Fisher's Linear Discriminant Analysis (FLDA) classifier. Implementing the classifier on a trial of 36 spatiotemporal vectors yields 36 scalar values. The resulting 36 scalar values are summed across the five trials for each character. The target character is taken as that character whose spatiotemporal feature yields the largest resulting trial-wise sum of scalars on pre-multiplication with the FLDA optimal vector. The chance accuracy of correctly identifying a target character in this manner is 2.78%.

3 Inter-session and intra-session classifier training and implementation

As aforementioned, four subjects each participated in four sessions of EEG data collection. Each session entails the delivery of 34 target characters. Each recorded session was separated into distinct training and test sets. The spatiotemporal feature vectors corresponding to the first 20 characters of each session form a training set. The remaining 14 characters form an independent test set which is used to obtain an unbiased measure of classifier accuracy. Testing a classifier on a test set which is derived from the same session in which the training set is obtained yields an instant of intra-session training accuracy. Conversely, testing a classifier on a test set which was derived on a different set from which the training set is obtained yields an instant of inter-session training accuracy.

4 N-fold cross-validation

A GA approach for sequential data-driven parameter selection in the feature extraction and pre-processing stages of BCI operation was investigated. GAs are random search optimisation techniques inspired by the mechanics of natural selection (Syan and Harnarinesingh, 2010). The findings of this study are provided in Table 1.

Table 1 Comparison of two-stage genetic algorithm parameter selection for feature extraction and pre-processing parameters proposal to data-insensitive parameters

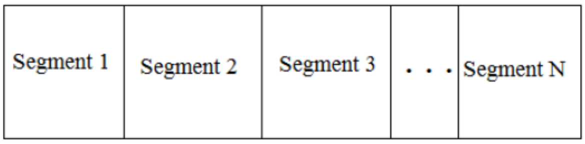
Subject	*4-channels*	*8-channels*	*16-channels*	*32-channels*	*Genetic-optimisation*		
					Stage 1	*Number of channels*	*Stage 2*
1	71.53%	64.96%	67.15%	59.12%	59.12%	19	57.66%
2	65.03%	62.94%	60.84%	61.54%	60.84%	17	55.24%
3	27.94%	25.74%	22.06%	24.26%	23.53%	20	22.06%
4	56.12%	43.88%	39.57%	38.85%	33.81%	21	28.06%
6	39.13%	27.54%	28.99%	26.09%	25.36%	14	18.84%
7	44.06%	44.76%	34.27%	32.87%	30.77%	15	38.46%
8	43.70%	24.44%	27.41%	28.15%	27.41%	20	20.74%
9	67.39%	63.04%	61.59%	65.94%	63.04%	18	61.59%
Average	51.86%	44.66%	42.74%	42.10%	40.49%	18	37.83%

Source: Harnarinesingh et al. (2011)

The two-stage GA approach to channel selection produced an average classification accuracy benefit of 4.27% whilst averting the need for human input in the parameter selection process. The GA approach therefore outperformed the general data-insensitive approach. However, the significant downside to the GA approach is its time-consuming nature. For example, a GA with a population size of 250 individuals that is set to terminate after 500 generations entails the evaluation of 125,000 fitness functions. The execution time of all of the fitness functions in addition to the computational overhead associated with the evolutionary heuristics of genomic trading, genomic mutation and sub-population migration is large for even computationally simple fitness functions. This renders GAs infeasible for online BCIs.

NFCV is utilised in the place of GAs for this work. NFCV is a popular approach in the BCI field (Lal et al., 2004; Citi et al., 2008). Consider the case where classification hyper-parameters need to be optimised for a training set. The training set can be split into a smaller training set and an independent test set. The optimal classification hyper-parameter can then be searched by obtaining the parameters that maximise classification accuracy on the test data. However, searching for a design parameter that maximises classification accuracy in this manner is likely to result in over-fitting the classifier on the training set. To address this problem, NFCV separates the entire data set into N segments and uses every possible segment in both the training and testing roles. This ensures that the trained classifier cannot be over-fitted to any particular segment of the training data.

Figure 3 Data segmentation for N-fold cross-validation



The algorithmic representation of the determination of NFCV accuracy on a training set is presented below.

```
accuracy = 0
for i = 1 to N
train classifier on all segments but segment i
test trained classifier on segment i
        accuracy = accuracy + test_accuracy
end for
```

In this work, feature extraction and pre-processing parameters are discretely searched using NFCV sequentially in the feature extraction and pre-processing stages by obtaining the parameters that maximise the ‘accuracy’ variable in the preceding algorithm. Using NFCV in this manner with an FLDA classifier is a form of meta-optimisation.

4.1 Automated data-driven feature extraction

The only free parameter of the feature-extraction stage which can be tuned to individual subject data is the down-sampling factor used to down-sample each EEG channel in order to ultimately produce spatiotemporal features. Fivefold cross-validation is used on all 16 training sets to identify the most appropriate down-sampling factor (K) for the training data, i.e. the down-sampling factor that maximises the fivefold cross-validation accuracy. The best value of K as identified in the training set using NFCV is then applied to unseen test data to evaluate the performance of the data-driven parameter proposal. For subjects S1–S4, this amounts to 64 instances of parameter proposal testing.

4.2 Automated data-driven pre-processing

The free parameter of the pre-processing stage which is tuned to individual subject data is the order a Butterworth digital filter with 3 dB cut-off frequencies at 1 Hz and 12 Hz. Fivefold cross-validation is used on all 16 training sets to identify the most appropriate filter order (n) for the training data, i.e. the filter order that maximises the fivefold cross-validation accuracy. The down-sampling factor obtained in the first stage of data-driven parameter selection is used. The best value of n as identified in the training set using fivefold cross-validation is then applied to unseen test data to evaluate the performance of the data-driven parameter proposal. For subjects S1–S4, this amounts to 64 instances of parameter proposal testing.

5 Results

5.1 Comparison of inter-session classifier training and implementation to intra-session classifier training

The classification accuracies due to inter-session and intra-session classifier training and implementation are presented in Table 2. This analysis was conducted using constant and data-insensitive parameters for the feature extraction and pre-processing BCI stages (Hoffmann et al., 2008). This is done in order to isolate the performance benefit due to the training format from any performance benefit due to automated data-driven system parameter selection.

Table 2 Classification accuracies for subjects S1–S4 using both inter-session and intra-session classifier training

Training session	*Testing session*	*Classification accuracy*	*Classification accuracy*	*Classification accuracy*	*Classification accuracy*
		Subject S1	*Subject S2*	*Subject S3*	*Subject S4*
S1	S1	84.62%	84.62%	53.85%	100.00%
S2		23.08%	69.23%	38.46%	92.31%
S3		7.69%	53.85%	0.00%	92.31%
S4		7.69%	92.31%	38.46%	84.62%
S1	S2	46.15%	38.46%	69.23%	61.54%
S2		38.46%	38.46%	92.31%	100.00%
S3		46.15%	46.15%	53.85%	84.62%
S4		23.08%	53.85%	61.54%	76.92%

Table 2 Classification accuracies for subjects S1–S4 using both inter-session and intra-session classifier training (continued)

Training session	*Testing session*	*Classification accuracy*	*Classification accuracy*	*Classification accuracy*	*Classification accuracy*
		Subject S1	*Subject S2*	*Subject S3*	*Subject S4*
S1	S3	6.67%	92.31%	100.00%	73.33%
S2		40.00%	76.92%	100.00%	66.67%
S3		46.67%	84.62%	93.33%	100.00%
S4		33.33%	100.00%	60.00%	93.33%
S1	S4	40.00%	47.06%	33.33%	33.33%
S2		46.67%	64.71%	20.00%	40.00%
S3		33.33%	58.82%	6.67%	26.67%
S4		46.67%	70.59%	46.67%	80.00%

The average inter-session classification accuracies and the intra-session classification accuracies for subjects S1–S4 across the 64 training/test session combinations are shown in Table 3.

Table 3 Average inter-session classification accuracy compared to intra-session classification accuracy for subjects S1–S4

Training format	*Subject 1*	*Subject 2*	*Subject 3*	*Subject 4*	*Average μ*	*SD σ*
Intra-session	54.11%	69.57%	71.54%	95.00%	72.55%	16.87%
Inter-session	29.49%	66.14%	48.46%	68.80%	53.22%	18.22%

The *t*-test statistic is obtained using the formula, $t = \mu_1 - \mu_2 \bigg/ \sqrt{\dfrac{\sigma^2}{N_1} + \dfrac{\sigma^2}{N_2}} = 1.5573$ (6 dof).

The critical one-tailed *t*-test statistic at the 90% level is 1.44.

5.2 Automated data-driven feature-extraction

The average classification accuracies for subjects S1–S4 due to NFCV parameter selection in both the inter-session and intra-session formats are presented in Table 4. The average classification accuracy due to parameter selection by random chance is given as well. No pre-processing is applied in this stage. Data-driven pre-processing and feature extraction parameter selection is done sequentially. It is more appropriate to perform data-driven feature extraction first since a pre-processing set-up is not required for data-driven feature extraction. However, a feature extraction stage is needed for pre-processing parameter selection.

The *t*-test statistic for the comparison of average classification accuracies in intra-session format due to automated feature extraction parameter selection is $t =\ t = \mu_1 - \mu_2 \bigg/ \sqrt{\dfrac{\sigma^2}{N_1} + \dfrac{\sigma^2}{N_2}} = 3.2856$ (30 dof). The critical one-tailed *t*-test statistic at the 99% level is 2.75.

Table 4 Average inter-session and intra-session classification accuracy on unseen test data using N-fold cross-validation of down-sampling factor

Subject	*Session*	*Average intra-session accuracy*		*Average inter-session accuracy*	
		N-fold cross-validation parameter proposal	*Chance parameter proposal*	*N-fold cross-validation parameter proposal*	*Chance parameter proposal*
1	1	69.23%	74.99%	35.04%	30.58%
	2	53.85%	38.46%	28.03%	30.17%
	3	13.33%	27.50%	19.49%	19.40%
	4	46.67%	40.00%	8.00%	8.00%
2	1	100.00%	92.31%	77.38%	73.70%
	2	61.54%	67.31%	85.67%	80.58%
	3	100.00%	100.00%	57.47%	62.48%
	4	82.35%	63.24%	66.67%	66.03%
3	1	92.31%	88.46%	65.99%	64.55%
	2	84.62%	88.47%	61.54%	61.30%
	3	100.00%	96.67%	28.21%	31.00%
	4	66.67%	62.50%	66.49%	63.23%
4	1	84.62%	81.73%	18.12%	31.92%
	2	100.00%	85.58%	47.86%	45.43%
	3	93.33%	85.83%	40.00%	46.45%
	4	80.00%	75.83%	75.72%	68.50%
Average, μ		76.78%	73.06%	48.86%	48.96%
SD, σ		24.06%	21.82%	24.02%	21.51%

5.3 Automated data-driven pre-processing

The average classification accuracies for subjects S1–S4 due to NFCV parameter selection in both the inter-session and intra-session are presented in Table 5. The average classification accuracy due to filter order selection by random chance is given as well.

Table 5 Average inter-session and intra-session classification accuracy on unseen test data using N-fold cross-validation of forward-reverse Butterworth filter order

Subject	*Session*	*Average intra-session accuracy*		*Average inter-session accuracy*	
		N-fold cross-validation parameter proposal	*Chance parameter proposal*	*N-fold cross-validation parameter proposal*	*Chance parameter proposal*
1	1	84.62 %	79.49%	39.83%	38.46%
	2	76.92%	66.67%	34.70%	35.10%
	3	33.33%	37.78%	23.93%	25.76%
	4	60.00%	55.56%	24.61%	25.53%
2	1	92.31%	87.18%	14.70%	13.50%
	2	61.54%	61.54%	14.70%	15.00%
	3	100.00%	92.31%	7.01%	6.38%
	4	82.35%	80.39%	12.13%	12.02%

Table 5 Average inter-session and intra-session classification accuracy on unseen test data using N-fold cross-validation of forward-reverse Butterworth filter order (continued)

Subject	*Session*	*Average intra-session accuracy*		*Average inter-session accuracy*	
		N-fold cross-validation parameter proposal	*Chance parameter proposal*	*N-fold cross-validation parameter proposal*	*Chance parameter proposal*
3	1	92.31%	82.05%	80.00%	76.57%
	2	84.62%	89.75%	72.65%	67.01%
	3	100.00%	97.78%	27.86%	27.12%
	4	80.00%	64.45%	87.52%	72.71%
4	1	92.31%	94.87%	58.63%	58.52%
	2	100.00%	100.00%	68.55%	69.29%
	3	100.00%	100.00%	72.65%	67.86%
	4	86.67%	82.22%	87.52%	87.64%
Average, μ		82.94%	79.50%	45.44%	43.65%
SD, σ		18.14%	17.85%	29.16%	27.10%

The t-test statistic for the comparison of average classification accuracies in intra-session format due to automated pre-processing parameter selection is $t = \mu_1 - \mu_2 \Big/ \sqrt{\dfrac{\sigma^2}{N_1} + \dfrac{\sigma^2}{N_2}} = 4.3677$ (30 dof) The critical one-tailed t-test statistic at the 99.9% level is 3.646.

6 Discussion

Intra-session classifier training and implementation was found to provide a classification accuracy benefit of 19.33% across four subjects in 16 sessions over 64 training/test set combinations. The performance increase was significant at the 90.00% level of statistical confidence. NFCV was subsequently used to derive feature extraction and pre-processing parameters using training data sequentially in two stages. The NFCV parameter proposals were tested in both the intra-session and inter-session settings. Initially, no pre-processing was implemented and NFCV was used to proposal feature extraction parameters. The results of this analysis are summarised in Tables 4 and 5.

The feature extraction parameters proposed by NFCV yielded an increase in classification accuracy in the intra-session format by 4.23%–76.78%. The accuracy due to parameter selection by random chance is 73.06%. In the inter-session format, NFCV parameter selection yielded a classification accuracy of 48.86% or a decrease in accuracy by 4.36%. This is consistent with theoretical expectations since tuning system parameters to recorded data in the inter-session format is likely to result in over-fitting. The accuracy due to parameter selection by random chance is 48.96%.

The feature extraction parameters proposed by NFCV on the training data were used for the second stage of pre-processing parameter selection. The parameters proposed by NFCV for the pre-processing stage yielded a further increase in classification accuracy in the intra-session format by 6.16%–82.94%. The accuracy due to parameter selection by random chance is 79.50%. In the inter-session format, NFCV parameter selection yielded a classification accuracy of 45.44% or a further decrease in accuracy by 3.42%. This is in line with theoretical expectations as tuning system parameters to recorded data in the inter-session setting are likely to result in over-fitting. The accuracy due to parameter selection by random chance is 43.65%.

The average classification accuracy due to data-driven system parameter selection in the intra-session classifier training format is 82.94% compared to 53.22% due to inter-session classifier training for constant and data-insensitive system parameters. This represents an average accuracy benefit of 29.72%. This performance increase is statistically significant at the 99.9% level of confidence.

7 Conclusion

A genetic algorithm approach to automating BCI parameter selection in the pre-processing and feature extraction stages was investigated. The approach yielded an average performance benefit of 4.27% compared to the data-insensitive approaches whilst averting the need for subjective human expertise in the parameter selection process. GAs however are time consuming and infeasible for online utilisation. In this light, NFCV was investigated as a basis for automating parameter selection. NFCV provided a performance benefit of 1.77% whilst significantly reducing the time required for searching design parameters. NFCV executes in a matter of minutes compared to hours for GA. Additionally, intra-session classifier training and implementation was investigated as a means of improving BCI classification accuracy. Using data-insensitive system parameters, the intra-session protocol provided an average performance benefit of 19.33% compared to the traditional inter-session approach. The inclusion of NFCV as a basis for automated data-driven parameter selection improved the performance benefit to 29.72%. This result is significant at the 99.9% level of statistical significance. This finding is significant as intra-session P300-BCI classifier training can be applied to any P300-based BCI regardless of the existing application platform to obtain improved classification performance. Furthermore, NFCV automated parameter proposal reduces the need for subjective human expertise in the parameter selection process whilst also lending to a marginal improvement in classification accuracy. This contributes to making BCI development less time consuming.

References

Birbaumer, N., Hinterberger, T., Kübler, A. and Neumann, N. (2003) 'The thought-translation device (TTD): neurobehavioural mechanisms and clinical outcome', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 11, No. 2, pp.120–123.

Citi, L., Poli, R., Cinel, C. and Supulveda, F. (2008) 'P300-based BCI mouse with genetically-optimized analogue control', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 16, No. 1, pp.51–61.

Evangelista, P.F., Embrechts, M.J. and Szymanski, B.K. (2006) 'Taming the curse of dimensionality in kernels and novelty detection', *Advances in Intelligent and Soft Computing*, Vol. 34, pp.425–438.

Farwell, L. and Donchin, E. (1981) 'Taking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials', *Electroencephalography and Clinical Neurophysiology*, Vol. 70, pp.510–523.

Harnarinesingh, R.E.S., Syan, C.S. and Palaniappan, R. (2011) 'Automation of pre-processing parameter selection for a single-trial P300-based brain-computer interface using a genetic algorithm', *Proceedings of the 26th International Conference on CAD/CAM, Robotics and Factories of the Future (CARS&FOF 2011),* Kuala Lumpur, Malaysia, 26–28 July.

Hoffmann, U., Vesin, J., Ebrahimi, T. and Diserens, K. (2008) 'An efficient P300-based brain-computer interface for disabled subjects', *Journal of Neuroscience Methods*, Vol. 167, No. 1, pp.115–125.

Krusienski, D.J., Schlak, G., McFarland, D.J. and Wolpaw, J.R. (2007) 'A μ-rhythm matched filter for continuous control of a brain-computer interface', *IEEE Transactions on Biomedical Engineering*, Vol. 54, No. 2, pp.273–280.

Kubler, A., Mushahwar, V.K., Hochberg, L.E. and Donoghue, J.P. (2006) 'BCI meeting 2005-workshop on clinical issues and applications', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 14, pp.131–134.

Lal, T.N., Schröder, M., Hill, N.J., Hinterberger, T., Mellinger, J., Wetson, J., Bogdan, M., Hofmann, T., Birbaumer, N. and Schölkopf, B. (2004) 'Support vector channel selection in BCI', *Proceedings of the 22nd International Conference on Machine Learning*, pp.465–472.

Lin, Z., Zhang, C., Wei, W. and Gao, X. (2006) 'Frequency recognition based on canonical correlation analysis for SSVEP-based BCIS', *IEEE Transactions on Biomedical Engineering*, Vol. 53, No. 12, pp.2610–2614.

Peters, B.O., Pfurtscheller, G. and Flyvbjerg, H. (2001) 'Automatic differentiation of multichannel EEG signals', *IEEE Transactions on Biomedical Engineering*, Vol. 48, No. 1, pp.111–116.

Pires, G., Castelo-Branco, M. and Urbano, N. (2008) 'Visual P300-based BCI to steer a wheelchair: a Bayesian approach', *Proceedings of 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Vancouver, Canada, pp.658–661.

Rebsamen, B., Burdet, E., Guan, C., Zhang, H., Teo, C.L., Zeng, Q., Laugier, C. and Ang Jr., M.H. (2007) 'Controlling a wheelchair indoors using thought', *IEEE Transactions on Intelligent Systems*, Vol. 22, No. 2, pp.18–24.

Trejo, L., Rosipal, R. and Matthews, B. (2006) 'Brain-computer interfaces for 1-D and 2-D cursor control: designs using volitional control of the EEG spectrum or steady-state visual evoked potentials', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 14, No. 2, pp.225–229.

Syan, C.S. and Harnarinesingh, R.E.S. (2010) 'Genetically-optimised spatial filtering of EEG data for a P300-based brain computer interface', *Proceedings of the 25th International Conference on CAD/CAM*, Robotics and Factories of the Future (CARS&FOF 2010), Pretoria, South Africa, 13–16 July.

Wills, S.A. and MacKay, D.J.C. (2006) 'DASHER – an efficient writing system for brain-computer interfaces?', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 14, No. 2, pp.244–246.

Wu, B., Zhou, Y., Yan, L., Yuan, J., Li, R. and Wang, D. (2009, March) 'Object detection from HS/MS and multi-platform remote sensing imagery by the integration of biologically and geometrically inspired approaches', *Proceedings of the ASPRS 2009 Annual Conference*, Baltimore, Maryland.