Published in final edited form as:

Expert Syst Appl. 2016 December 15; 65: 164–180. doi:10.1016/j.eswa.2016.08.044.

Mixture of autoregressive modeling orders and its implication on single trial EEG classification

Adham Atyabi[a,b,*], **Frederick Shic**[a], and **Adam Naples**[a]

[a]Yale Child Study Center, School of Medicine, Yale University, New Haven, Connecticut, United States of America

[b]School of Computer, Science, Engineering and Mathematics, Flinders University of South Australia, Australia

Abstract

Autoregressive (AR) models are of commonly utilized feature types in Electroencephalogram (EEG) studies due to offering better resolution, smoother spectra and being applicable to short segments of data. Identifying correct AR's modeling order is an open challenge. Lower model orders poorly represent the signal while higher orders increase noise. Conventional methods for estimating modeling order includes Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Final Prediction Error (FPE). This article assesses the hypothesis that appropriate mixture of multiple AR orders is likely to better represent the true signal compared to any single order. Better spectral representation of underlying EEG patterns can increase utility of AR features in Brain Computer Interface (BCI) systems by increasing timely & correctly responsiveness of such systems to operator's thoughts. Two mechanisms of Evolutionary-based fusion and Ensemble-based mixture are utilized for identifying such appropriate mixture of modeling orders. The classification performance of the resultant AR-mixtures are assessed against several conventional methods utilized by the community including 1) A well-known set of commonly used orders suggested by the literature, 2) conventional order estimation approaches (e.g., AIC, BIC and FPE), 3) blind mixture of AR features originated from a range of well-known orders. Five datasets from BCI competition III that contain 2, 3 and 4 motor imagery tasks are considered for the assessment. The results indicate superiority of Ensemble-based modeling order mixture and evolutionary-based order fusion methods within all datasets.

Keywords

Autoregressive analysis; Genetic algorithm; Particle Swarm Optimization; Electroencephalogram

*Corresponding author: Adham.Atyabi@Yale.edu (Adham Atyabi).

1. Introduction

Autoregressive (AR) models have established their value in a broad range of applications in digital spectral analysis. AR models ability in terms of handling short segments of data, offering better frequency resolution and smooth power spectra are advantageous in comparison to Discrete Fourier Transform (DFT) and Fast Fourier Transform (FFT) (Palaniappan, 2006b). The inherent computational efficiency of AR models is of advantage compared with alternatives such as Moving Average (MA) and Autoregressive Moving Average (ARMA) (Ning Bronzino, 1990). Accurate modeling order settings results in accurate representation of the underlying signal. Low modeling orders are known to be poor representatives of the signal properties while high modeling orders are likely to represent noise resulting in unreliable representation of the signal.

The recorded Electroencephalogram (EEG) reflects the unified action of many cortical areas and is further smeared by volume conduction of the signal in the brain, thus the activity at any one scalp electrode reflects the mixture of many spatially and temporally overlapping patterns of the brain activity. This acknowledges the argument that any representation of such pattern that favors a single aspect against the rest is likely to be incomplete representation of the underlying patterns. In light of AR-based EEG feature representation, this issue is indicative that any AR feature representation of an EEG signal that is originated from a fixed modeling order is only reflective of a sub-group of the underlying patterns that the EEG recording contains. Therefore, the conventional AR order estimation methods such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Final Prediction Error (FPE) will tend to identify that order by capturing the strongest underlying pattern in the EEG (in their view) irrespective to the completeness of the resultant pattern and its implication in terms of pattern classification. Although the idea of representing an EEG spectrum with AR features of more than one AR-order is novel within EEG and signal processing community, an equivalent (in some degree) study been conducted in statistical analysis of non-linear time series by Chun Shan Wong (2000) in which mixture of k stationary or non-stationary AR components are used. The study is focused on sub-selection of AR components for better representing the non-linear time series rather than identifying multiple AR-order candidates. AIC and BIC are considered for order estimation in the study. Authors claimed advantages such as "*more full range of shape changing predictive distribution and the ability to handle cycle and conditional heteroscedasticity in the time series*" (Chun Shan Wong, 2000). Authors also acknowledged the utility of mixture of AR models in handling conditional heteroscedastisity in their discussions (Chun Shan Wong, 2000).

Current study challenges the conventional operation of AR-order estimation in EEG-based BCI studies by assessing following hypothesis:

Hypothesis

Adequate mixture of AR features derived from various AR modeling orders is a better representative of the underlying signal compared with any AR-based representation that is derived from a single modeling order.

In the context of expert systems in general and Brain Computer Interfacing (BCI) in specific, proving this hypothesis provides the opportunity of extracting more veridical spectral patterns from EEG recordings that are more distinguishable from each other and are better representatives of the performed tasks by participants. Given the vast application of AR coefficients in BCI studies, such improvement can lead to the generation of systems that are much more responsive to the ongoing changes in the spectral information of participating subjects and are more accurate in terms of identifying the intention of their users.

In order to assess the hypothesis the study proposes two mechanisms for automatically identifying the correct mixture of AR modeling orders. These mechanisms are as follows:

1. Evolutionary-based fusion of AR features obtained from a collection of modeling orders (ranging between 2 to 30).
2. Ensemble-based modeling order mixture.

Three sets of experiments addressing different conventional methodologies for selecting AR-order are considered in order to provide base-line performance for evaluating the feasibility of the Evolutionary-based and Ensemble-based mechanisms proposed in this study. These conventional methodologies are as follows:

- **Investigator's intuition**: It is a common practice in EEG community to choose the modeling order based on suggestion's made by the literature on similar experimental conditions or to use the researcher's intuitions. Reported effective AR-orders in EEG spectrum analysis includes 2, 4 , 6, 8, 10, 16 and 30.
- **Conventional estimation approaches**: The most commonly considered order estimation methods includes Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Final Prediction Error (FPE) Vedavathi et al. (2014); Peiyang et al. (2015); Eriolu Gnay (2010); Kotkatvuorirnberg (2016); Wang et al. (2010); Dirgenali Kara (2006).
- Blind Mixture of AR features obtained from a collection of well-known modeling orders (ranging between 2 to 30) as proposed by Fitzgibbon (2007).

Five well-known datasets from BCI competition III that contain 2, 3 and 4 class problems (motor imagery tasks) are employed for assessment of the considered approaches.

The outline of the study is as follows. A brief introduction to Autoregressive analysis and its recent applications in single trial EEG studies are presented in Section 2. Section 3 introduces the approaches that are being used in the study. Section 4 presents datasets, data restructuring, and preprocessing approaches utilized in the study. Experimental results are discussed in Section 5. Sections 6 and 7 presents discussion and conclusion respectively.

2. Background

2.1. Autoregressive (AR) analysis

Autoregressive (AR) method can be considered as a linear filter that can be used for analysis of signals that are corrupted by white noise. Xu Song (2008) introduced following equation for AR:

$$x(n) = -\sum_{i=1}^{p} a_p(i)x(n-i) + \varepsilon(n) \quad (1)$$

In here, p is the AR order, $x(n)$ is the input signal and n refers to sample point. $a_p(i)$ is the AR coefficient and ε is the white noise. The purpose is to estimate AR coefficients. Even though AR is computationally efficient and its simplicity is its advantage compared to other techniques, the model order (p) plays a major role in the accuracy of the resulting features. In AR model, the use of a high p value provides spurious peaks while a low p value results in smooth spectrum (Xu Song, 2008). introduced following methods for estimating the modeling order, i) Akaike Information Criterion (AIC), ii) Final Prediction Error (FPE), iii) Minimum Description Length (MDL), vi) Criterion Autoregressive Transform (CAT), v) Hannan and Quinn (HQ), and vi) Residual variance (RV). Between these methods, AIC, BIC and FPE are common to be used in EEG studies.

Inoue et al. (2003) investigated the use of AR method in single trial motor imagery tasks. The study showed over 90% classification performance in right and left motor imagery tasks. Kus et al. (2006) employed a short time directed transform function (STDF[1]) based on AR model to distinguish actual and imaginary finger movement from each other. The study reported the im- pact of finger movement and imagination on beta and gamma frequency bands. Kus et al. (2006) demonstrated that beta and gamma rhythms have different synchronization properties (in a motor control task, an increase in gamma may be accompanied with decrease in beta frequency band). They also found out that gamma rhythm has a high contribution in imagery activities. Nagata et al. (2006) employed AR for feature extraction in a motor imagery problem based on distinguishing three tasks. In their study, EOG and EMG signals are used to detect and remove artifacts. Tsoi et al. investigated the combination of AR as feature extractor and Multilayer Perceptrons (MLP) as classifier in EEG classification problem (Tsoi et al., 1993). Wolpaw et al. (2000) proposed the use of spatial filter with AR method for BCI studies that contain components/activities that their motion needs to be controlled by participants (focusing on sensorimotor rhythms, beta and mu). The proposed spatial filter is considered in a way to match the user's β and μ rhythms. Wolpow considered AR as a better choice compared to FFT for short time segments due to its ability to provide higher resolution.

As mentioned earlier, identifying optimal modeling order of AR features has a direct impact on how well the AR features represent the performed tasks by participants. Jansen et al. (1981) claimed model order of 10 as the optimal modeling order for EEG. Vaz et al. (1987) identified AR-order 5 as the optimal estimate when working with rhythmic and non-"featureless background" EEG. Krusienski et al. (2006) argued that the criteria considered in Sensory Motor Rhythms (SMR) -based BCI studies by literature does not necessarily result in optimal AR modeling order estimation. That is, Krusienski et al. suggested that in order to control BCI system, both the rhythmic and "featureless

[1]STDF is a modified version of Direct Transform Function (DTF) that operates based on the ensemble averaging paradigm by calculating the estimation of windowed data with respect to the existent of multiple realization of the process (Kus et al., 2006).

background" components of the EEG need to be captured by AR spectral estimation. The study confirmed this argument by comparing the performance of AIC with fixed modeling orders of 6, 10, 16 and r^2. Krusienski et al. concluded that higher model orders are expected to generate more accurate spectral estimation when working with SMR-BCI systems. Engin et al. (2001) suggested that the model order p has direct relation with sampling rate of EEG. They stated that "the order estimation for the data with lower sampling frequency results in lower orders" (Engin et al., 2001). The study investigated this issue with 5s EEG data sampled at 50Hz and 100Hz and considered model orders in the range of 1 to 25. Palaniappan Raveendran (2001) investigated feasibility of various conventional model estimation approaches including Akaike Information Criterion, Final Prediction Error, Residual Variance, Minimum Description Length, Criterion Autoregressive Transfer and Hannan-Quinn in addition to modeling order 6. Fuzzy ARTMAP neural network is employed for evaluating the performance. The study failed to identify a clear performance advantage across these methods when a dataset with multiple subjects is used. That is, neither of the conventional approaches performed consistently across subjects (Palaniappan Raveendran, 2001).

2.2. Applications of AR on EEG studies: survey of the state-of-the-art

Autoregressive models have many applications in EEG signal analysis varying from estimation of spectral characteristics of EEG signal to artifact rejection and stationary signal discrimination. This section covers a range of recently published articles that features the application of AR in EEG analysis.

Peiyang et al. (2015) proposed a modified autoregressive model for power spectrum analysis of the resting state EEG signal in the presence of artifacts and outliers. The study proposed the use of Lp norm ($p \leq 1$) in stead of conventional L2 norm configuration introduced by Yule-Walker[2]. P300 evoked potential responses to oddball paradigm from P3 electrode is used for assessing the feasibility of the approach. 4 to 20 outliers are randomly injected to the first 4s of the recording. The feasibility of the Lp-norm AR is assessed on the basis of its ability to fit the artifactual data with different degrees of outliers in which it managed to outperform both Yule-Walker and Burg AR methods.

Camilleri et al. (2015b) proposed a semi-supervised autoregressive switching multiple model (AR-SMM) for segmentation of EEG data in a BCI study. The study used a single model in the beginning (one labeled instance) and segmented the EEG data whenever detected a new model that did not fit to the characteristics of the previously identified segments. Total of 22 trials of eye-close and eye-open over O2 channel from a single participant is considered in the study.

Vijayan et al. (2015) investigated AR modeling in an EEG-based emotion recognition and classification study. The novelty of this approach lied within the use of statistical measures as a weighting system on AR features. DEAP benchmark emotion based EEG database is utilized. The database contained 4 emotional states recorded from 32 electrodes with 32

[2]L2-norm-based methods are known to be effective in representing spatially extended sources while L1-norm-based approaches are known to better estimate focal and sparse sources (Liu et al., 2015)

participants. The procedure considered in the study is based on fitting a higher order autoregressive model to the extracted Shannon-Entropy features and using a support vector machine (SVM) for classifying the resulting models. Yule-Walker AR coefficient estimation is used. The exact modeling order used in the study is not specified.

Shahabi Moghimi (2016) used multivariate AR (MVAR) modeling for extracting the connectivity patterns between EEG electrodes under different frequency bands in a music study. The study is focused on the emotional responses of 19 participants to a set of Iranian and classical music. The correlation between inter and intera regional connectivity and the assessments of musical selections are considered. An SVM is used for classifying the two conditions of joyful and neutral. Ar-order is set to 6 based on the estimation gained from AIC.

Pippa et al. (2015) proposed a new approach for automatic estimation of correct modeling order of AR features in an epileptic seizure detection study. Combinations of statistical features and regression analysis are used for identifying the optimal AR-order. Recordings from 10 epileptic participants are used for assessing the feasibility of the approach. The results indicated mean absolute error of 4 units in estimation of optimal AR-order. The authors argued that the difference between AR coefficients of neighboring AR-orders are negligible and the 4 unit estimation error is likely to be within the acceptable estimation range.

Yonghui et al. (2015) used combinations of autoregressive features extracted from phase space and Linear Discriminant Analysis (LDA) in an EEG classification study. Graze 2003 and Graze 2005 datasets are considered for evaluating the performance. AR-order is set on the basis of multiple runs of repeated 10- fold cross validations on which different AR-orders are evaluated and the best consistently performing of all evaluated AR-orders (on each subject) is later on utilized on the phase data. Although the procedure some what guaranties finding the best fitting single AR-order for the task, it is computationally expensive despite the fact that the estimation of AR features is a relatively low computation task. The results identified AR orders of 8 and 5 for EEG data of various participants.

Abo-Zahhad et al. (2015) introduced a human authentication approach on the basis of using the mixture of EEG and eye-blinks recordings. Combinations of AR features and time delineation of eye blinks are used as a fused feature set with an LDA classifier. The evaluation of the approach is done using EEG recordings of 31 subjects performing 3 tasks of relaxation, visual stimulation, and eye-blinking. AR coefficients are estimated using Burg algorithm. The modeling order is set to 50 based on investigator's intuition and later systematic sub-selection is conducted in order to cut out the none or less contributing higher order AR coefficients

Kayikcioglu et al. (2015) investigated implications of AR features on classifications of sleep and wake stages. Authors set the AR-order to 22 based on their intuition. Zhao et al. (2011) applied AIC for estimating optimal AR-order in an EEG classification study aiming to identify mental fatigue in drivers' signals. The study extracted MVAR features from EEG recordings of 10 drivers and used combinations of kernel-based principle component analysis (as feature decomposition) and SVM as classifier.

Ligeois et al. (2015) utilized the information in low rank structure of MVAR models in synthetic and real-world neuro-imaging datasets and used alternating direction method of multipliers (ADMM) to handle MVAR's model's spars plus low-rank graphical problem. First order AR is considered in the study (AR(1)). Camilleri et al. (2015a) investigated implication of AR coefficients in segmentation and classification of EEG-based BCI signals. The study considered a range of AR-orders between 2 to 10 and identified modeling order 6 as the optimal value for the study. Li et al. (2016) considered AR modeling in non-stationary EEG analysis of time-frequency domain. Time-varying AR modeling that utilizes radial base functions is used by authors and Particle Swarm Optimization (PSO) is employed to identify the optimal parameters of the RBF kernels while AR order estimation is performed using FPE method. Loukas et al. (2015) investigate the utility of AR modeling in graph-based signal filtering. Distributed AR with moving average is proposed in the study and the feasibility of the method is assessed on time varying signals. AR-order estimation is based on combination of using first and arbitrary order ARs.

(Karahan et al., 2015) studied multi-modal brain image fusion for parsing the brain structures that reflect human cognitive processes and brain structural and regional connectivity. A high-dimensional MVAR is employed to search for the influence fields. The chosen influence fields by MVAR represent spatial maps highlighting the degree of influence of one region on others. BIC is used for order estimation but the exact order utilized by the study is not specified. Liu et al. (2015) considered adaptive source imaging via processing spatio-temporal information of patch source. MVAR is utilized for describing the patch sources' spatio-temporal dynamics. AR-order estimation mechanism is not specified.

Hsu (2015) studied multi-feature classification of EEG-based BCI recordings of 2 motor imagery tasks. Several features including adaptive AR model, amplitude modulation, spectral power and asymmetry ratio and wavelet fuzzy approximate entropy are considered. A modified PSO is used for feature selection and SVM is employed for feature classification. Rubega et al. (2016) investigated EEG signal coherence in type 1 diabetes patients. MVAR models are used for computing information Partial Directed Coherence (iPDC) function and 3 and 2 sets of electrode clusters are considered on theta and alpha frequency bands respectively. AIC is employed for estimating the optimal ARorder. Shaw Routray (2015) presented the estimation of neural connectivity of EEG recordings during meditation. Time-varying MVAR models are used for investigating the connectivity estimate of time varying Granger Causality. AIC is employed for estimating the optimal modeling order of MVAR.

Wu et al. (2015) utilized AR in a study focused on spectral analysis of cortical EEG recordings of a rat-based epileptic seizure. AR is used for extraction of power spectra features. AR-order estimation mechanism is not specified. Rotondi et al. (2016) investigated the utility of AR models in EEG connectivity via Partial Directed Coherence (PDC) in a childhood absence epilepsy study. MVAR is used for measuring connectivity within frequency domain. AIC is utilized to estimate the optimal AR-order. Fang et al. (2015) studied implications of phase based feature classification on EEG-based BCI system. the study utilized AR coefficients for phase space reconstruction in time delay embedding. AR modeling order is estimated through a 10-fold cross validation over a modeling orders in the range of 5 to 8.

Table 1 provides an overview on the most recent uses of AR features in brain signal analysis studies with a focus on the order estimation methods and AR-orders utilized in the studies.

As evident from the set of recently published papers discussed in this section and the information reported in table 1, it is a common practice to either use conventional order estimation methods especially AIC or choose the ARorder based on the investigators' intuition from a range of AR modeling orders ($p \in [2, 30]$). To the best of our knowledge, the idea of identifying the optimal mixture of AR modeling orders rather than finding a single AR-order has never been studied by either EEG or BCI communities. In fact, (Fitzgibbon, 2007) is the only study which considered the contribution of more than one AR-order in its spectral pattern representation. The study included a range of well-known AR-orders that been suggested by the community and concatenated the resultant coefficients to generate its spectral feature vector.

3. Methods and Materials

Considering that the aim of this study is to assess the hypothesis of any set of AR features that only represent a single modeling order is likely to provide poorer representation of the underlying patterns in comparison to an adequate mixture of AR features with varying modeling orders, several approaches for model order estimation including fixed (modeling orders in the range of 2 to 30), Conventional estimation methods (AIC, BIC and FPE), and mixture of modeling order features (concatenation, Evolutionary fusion and ensemble mixture) are to be considered and assessed in this study. In order to address issues raised in (Krusienski et al., 2006) (higher model orders are expected to generate more accurate spectral estimation when working with SMR-BCI systems), (Engin et al., 2001) (order estimation gets influenced by sampling rate of recordings), and (Palaniappan Raveendran, 2001) (the modeling orders estimated by conventional methods are subject dependent) , 5 datasets from BCI competition III, containing Sensory Motor Rhythms (2, 3, and 4 classes) with several subjects are employed. All datasets are resampled to 250Hz (lowest sampling rate between datasets).

3.1. AR modeling order estimation methods suggested in literature (baseline)

3.1.1. Using well-known modeling orders and investigator's intuitions (Jansen et al., 1981; Krusienski et al., 2006)—In this study Matlab implementation of AR is utilized considering fixed modeling orders of 2, 4, 6, 8, 10, 16 and 30. Figure 1 illustrates the procedure used for measuring classification performance of autoregressive features using various modeling orders.

3.1.2. Conventional order estimation methods (Palaniappan, 2006a) (Palaniappan Raveendran, 2001)(Krusienski et al., 2006)—ACI, BIC, and FPE are three conventional approaches commonly employed in EEG studies for estimating modeling order of autoregressive method. The diagram flow for this category is similar to what is presented in Figure 1. First, a small portion of EEG data (10% of the samples randomly selected) are used for estimation of AR-order with conventional modeling order estimation methods. Afterwards, the AR coefficients are extracted and the classification performance is assessed within a repeated 10-fold cross validation scheme.

3.1.3. Concatenated modeling orders (Fitzgibbon, 2007)—Figure 2 illustrates the procedure used for measuring classification performance of autoregressive features using a concatenated feature vector. In this approach, first, seven separate feature vectors for each modeling order (AR modeling order(n), $n \in \{2, 4, 6, 8, 10, 16, 30\}$) are generated and later these feature vectors are concatenated together to shape the mixed vector. Using a repetitive 10-fold cross validation scheme and an Extreme Learning Machine (ELM) (Huang et al., 2011a, b; Liang et al., 2006; Huang et al., 2006; Zhu et al., 2005; Tang Han, 2009) as a classifier, the feasibility of concatenated feature vector is assessed.

3.2. Novelty and contribution of the current study

The main novelty of the current study lies within the idea of considering more than one modeling order for extraction of spectral features in a EEG single trial classification study. It is a common practice to employ conventional order estimation methods such as AIC, BIC and FPE or to use the investigators' intuition to set the AR's modeling order. Given the complex nature of EEG recordings that contains the brain activity of multiple sources that are captured over time from specific locations on the scalp, the main hypothesis of this study is that any set of AR features that only reflects a single modeling order is likely to only capture partial information from the existing spectral patterns in the signal. To the best of our knowledge no other study has targeted the optimal mixing of more than one AR-order before. Two mechanisms are proposed in this study to find the optimal mixture of AR-orders. These two mechanisms feature evolutionary-based and ensemble-based mixture algorithms.

3.2.1. Evolutionary based mixture of AR features with various modeling orders —Figure 3 illustrates the procedure considered for measuring classification performance of evolutionary-based fusion of autoregressive features with various modeling orders. In this approach, a repeated 10-fold cross validation scheme is utilized that generates 3 sets of training, validation and testing in each fold. Within each fold of the cross validation, an evolutionary technique is utilized to identify the best combinations of modeling orders. The classification performance on training and validation sets is used to guide the evolution. The final fittest solution is reassessed with the unseen testing set. Similar to previous approaches, ELM is utilized as the classifier.

Evolutionary based fusion utilizes following steps:

1. Initialization: generate a random population of binary strings with length of 7 (e.g., "0001101"). Each cell in the string represents an specific modeling order. In each member of the population, cells with value of 0 and 1 are to represent

inclusion and omittion of associated AR features originated from the associated modeling orders in the final mixture respectively.

2. Evaluation: Feature vectors of AR modeling orders representing activated cells in each member of the population are to be concatenated and the resulting mixed vector to be passed to the classifier for assessing the feasibility of the mixture.

3. Repeat following sub-steps until the stopping criteria is achieved

   a. update the population

   b. evaluate the population

4. Report the best performing member of the population

Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) are the two evolutionary methods considered in this study.

Genetic Algorithm (GA) is introduced by John Holland in 1975 and it is based on the biological principal of natural selection. Here the most successful members of the population for a given task survive and those not suited will perish. In GA, each member of the population is referred to as a chromosome, and each chromosome represents a possible solution for the problem at hand. The algorithm utilizes a fitness function in order to evaluate the feasibility of the chromosomes. GA updates its population and evolves chromosomes towards regions of the search space within which better possible solutions can be found using processes/operators such as selection, crossover, reproduction and mutation (Bonabeau et al., 1999; Grefenstette, 1990; Holland, 1992).

Particle Swarm Optimization (PSO) is an evolutionary algorithm inspired by animals' social behaviors (e.g. fish schooling and bird flocking). Introduced by Kennedy and Eberhart in 1995, this algorithm employs members of the population, referred to as particles, within processes such as generation, evaluation and reproduction/update in order to find a solution (Kennedy Eberhart, 1995). In PSO, each particle's position in the search space is represented by a position X. PSO evolve its solutions towards better regions of the search space by updating the particle's position in the search space using a velocity V. The best solution found by each particle is referred to as personal best (*PBest*) and the best performing particle found by the whole swarm is referred to as global best (*GBest*). In PSO, particles update their velocities using the following equations:

$$V_{i,j}(t)=wV_{i,j}(t-1)+C_{i,j}+S_{i,j}$$

$$C_{i,j}=c_1r_{1,j}\times(P\mathrm{Best}_{i,j}(t-1)-x_{i,j}(t-1)) \quad (2)$$

$$S_{i,j}=c_2r_{2,j}\times(G\mathrm{Best}_{i,j}(t-1)-x_{i,j}(t-1))$$

In equation 2, $V_{i,j}(t)$ represents the velocity in iteration t. i and j represent the particle's index and the dimension in the search space respectively. c_1 and c_2 represent the acceleration coefficients of cognitive ($C_{i,j}$) and social ($S_{i,j}$) components respectively. $r_{1,j}$ and $r_{2,j}$ are random values in the range of [0,1] while w is the inertia weight that controls the influence of the last velocity in the updated version.

The equation for updating the particles is as follows:

$$x_{i,j}(t)=x_{i,j}(t-1)+V_{i,j}(t) \quad (3)$$

$PBest_{i,j}$ and $GBest_{i,j}$ represent the best solution found by the particle and the best overall solution found by the swarm. A detailed discussion on variations of evolutionary methods and their performances can be found in (Ab Wahab et al., 2015).

3.2.2. Ensemble-based mixture of AR modeling orders—Figure 4 illustrates the procedure used for identifying the best mixture of AR modeling orders using an ensemble learning approach. First, the EEG data is divided to training, validation and testing sets using a repetitive 10-fold crossvalidation scheme. Later, seven ELM classifiers are utilized in order to construct the ensemble. Each ELM in the ensemble is only trained, validated and tested with AR features of an specific modeling order. The D operator in the ensemble diagram is representative of decision aggregation operator. Summation and weighted summation are the two decision aggregation operators considered in this study. In the wighted summation method, the performance achieved from validation set with each classifier in the ensemble is utilized as the weight.

4. Data and Preprocessing

In this study, 5 datasets containing motor imagery of two, three and four limbs are employed. These datasets are obtained from BCI competition III (Blanchard Blankertz, 2004; Schlogl et al., 2004). The detail information about these datasets is presented in table 2.

4.1. Restructuring the Data

Since the comparison of results between these datasets is one of the objectives of the study, they are restructured to a common framework containing:

- The data acquired during the time that subject was performing a cognitive task (denoted as '*task*' in datasets).
- The data acquired outside of the time specified for performing motor imagery tasks during the instructions, blank screen, inter-trial and so on (denoted as '*non-task*' in datasets). The task period in all datasets is labeled appropriately in a way to represent the performed motor imagery tasks. In this study, only the task period is used for the purpose of feature extraction and classification.
- The EEG epochs representing the '*task*' periods are divided to 0.5s non-overlapping intervals (sub-epochs) in order to increase the number of subsequent

training samples in addition to provide consistency between the datasets. The choice of 0.5s sub-epoch size as the basis of the study is made since this is the shortest epoch size that exists within the datasets utilized. In addition, in previous findings, using a sub-set of these datasets this window size was shown to be the shortest window size appropriate (Atyabi et al., 2012a, b).

- No electrode selection/reduction procedure is performed despite the inconsistency across datasets in terms of number of electrodes utilized. This decision is made on the basis of previous studies with a sub-set of these datasets in which evolutionary based electrode and feature selection methods revealed that in weaker subjects (classification performance wise) the brain regions associated with sub-set of electrodes that best captures the tasks performed are within frontal brain regions rather than areas associated with motor imagery tasks. This is indicative of inconsistency across subjects' performances and lack of appropriate control mechanisms preventing subjects from generating misleading and to some extent meaningless EEG data during the recording sessions (Atyabi et al., 2012c).

4.2. Preprocessing

To provide consistency across datasets the epoched task period data is resampled at 250Hz (the smallest sample rate existing across datasets used in the study). This step is utilized in order to cancel possible effect of sample rate on overall classification performance of AR features as suggested by Engin, Engin et al. (2001). In addition, in order to eliminate the possible influence of different electrode referencing mechanisms across datasets, Common Average Referencing (CAR) is performed on all datasets. The final preprocessing step includes demeaning the epoched task periods in order to eliminate drifts and offsets from datasets.

4.2.1. Re-referencing the data using Common Average Reference (CAR)—To provide consistency between datasets, CAR is performed by subtracting the average signal from all electrodes using the following equation. EEG data of each subject in each dataset is treated separately.

$$V_i^{'}=V_i-\frac{1}{n}\sum_{j=1}^{n}V_i \quad (4)$$

where $V_i^{'}$ is CAR of electrode i (i=1,…,n). V_i represents the electrode i's original signal before being commonly averaged.

4.2.2. Demeaning (D)—Demeaning is performed to remove the possible constant offset and drifts from the EEG data using following equation.

$$E_i^{'}=E_i-\frac{1}{k}\sum_{j=1}^{k}E_i \quad (5)$$

where $E_i^{'}$ is demeaned version of epoch i (i=1,…,k). Epoch is considered as the duration of time within which the subject performed the task. In here, the demeaning formulation is applied to 0.5s sub-epochs.

4.3. Performance Measures

Various performance measures derived from a contingency (confusion) matrix are used for evaluation of classifiers. It is noteworthy that commonly used single measures of performance derived from a confusion matrix (such as accuracy, precision, and recall) are influenced by bias in sample size and class distribution. *Bookmaker Informedness* introduced by Powers (2003) addresses the problem by taking into the account the difference between the correct/ incorrect informed decisions and uninformed (random) choices. The *Bookmaker Informedness* ($Sensitivity+Specificity-1$) provides a measure between −1 and + 1 with + 1 representing perfectly correct performance, −1 indicating perversely incorrect response and 0 representing chance level. Being normalized and unbiased makes *Bookmaker Informedness* suitable for assessing the performance of classifiers and for comparison purposes. By contrast, Accuracy and F-measure have non-zero, data prevalence and classifier bias-dependent chance levels and cannot be compared meaningfully. Kappa is a more general family of chance corrected measure but apart from Informedness these all have dependencies on the label bias of the classifiers and are also not comparable across classifiers and datasets (Powers, 2012).Bookmaker Informedness is considered as main performance metric in this study. In order to provide better understanding of the results accuracy is considered as an alternative performance metrics and its results are also reported in the study.

In all experiments within this study, 10 × 10 Cross-Validation (CV) is conducted resulting in three sets of training, validation and testing with 0.9, 0.05, and 0.05 ratios.

4.4. Pattern Classification

Despite the reasonable classification performance that can be obtained by Back-propagation multilayer feedforward neural networks, these learning algorithms are relatively slow, get stuck in local minima, and their activation functions need to be differentiable (Zhu et al., 2005).

The Extreme Learning Machine (ELM) is a special type of single-hidden-layer feedforward neural network that reduces the required training time in the network by substituting the required training/learning phase for connecting the input layer to the hidden layer by providing random connections (weights) (Huang et al., 2011a, b; Liang et al., 2006; Huang et al., 2006; Zhu et al., 2005; Tang Han, 2009). The fast learning capability of the ELM and its lower computational expense compared with alternative classifiers such as support vector machine and single-layer/multi-layer neural network makes it suitable for a EEG signal classification study. This is with the understanding of drawbacks such as i) its tendency towards over-fitting and ii) not being immune to outliers due to not considering heteroskedasticity (Deng, 2009).

In this study, Sigmoid ELM with 80 hidden nodes are considered. The number of hidden nodes in ELM is set based on previous studies with the BCI competition III datasets (Atyabi et al., 2012c, 2013). Ad-hoc experimentations (results not presented in here) with higher and lower number of hidden nodes within these datasets showed no significant difference, with the chosen setting achieving slightly better average performance.

5. Results

In this study several approaches for identifying suitable AR modeling order are considered. These approaches are as follows:

1. Commonly investigated fixed AR modeling orders of 2, 4, 6, 8, 10, 16 and 30
2. Conventional order estimation approaches (e.g., AIC, BIC and FPE)
3. Concatenation of a range of modeling orders explored in the study
4. Evolutionary-based fusion of AR features originating from various modeling orders
5. Ensemble-based mixture of AR features originating from various modeling orders

With respect to conventional AR-order estimation methods (AIC, BIC, and FPE), the details of their minimum values and their associated modeling orders are presented in Table 3. It is noteworthy that except with AA and IVB subjects in which the value estimated with FPE as AR-order is much lower than the other two approaches (AIC and BIC), no other considerable difference is observed within the estimated AR-orders between these conventional methods.

The results achieved from all AR-order estimation and mixture mechanisms considered in this study are presented in Figures 5, 6, and 7. The results are categorized based on the number of classes considered in the datasets rather than their originating dataset. The Figures represent average classification performance (e.g., Bookmaker Informedness and Accuracy) with a Sigmoid ELM with 80 nodes on the testing set (considering a 10×10 cross-validation scheme).

5.1. Two Motor imagery Class Problem (IVa and IVb datasets)

In this section, average classification performance achieved on a sub-group of subjects that only performed two motor imagery tasks are presented in Figure 5. These subjects include all participants in IVa (AA, AL, AV, AW and AY) and IVb (IVB) datasets. Fixed modeling orders are presented with their modeling order values while concatenated method is represented as *ALL* in the Figure. Following observations are made from the results reported in Figure 5.

- Fixed modeling order $p \in \{2, 4, 6, 8, 10, 16, 30\}$: the results across the subjects indicate the superiority of higher modeling orders (16 and 30) in comparison to the lower modeling orders. This is with the exception of subject AA in IVa dataset in which modeling order 10 performed slightly better than others.
- Conventional estimation methods (AIC, BIC, and FPE): these approaches failed to outperform the best performing fixed modeling orders across all subjects. These methods also failed to outperform classification performance achieved by the ALL feature vector as well. The exception is subject IVB in which conventional methods performed slightly better than ALL. It should be noted that performance differences between these methods and best performing fixed modeling orders and ALL feature vector are marginal and non-significant. In addition, no tangible performance difference was observed across AIC, BIC, and FPE's average classification performance with exception of AA and IVB subjects in which AIC and BIC performed slightly better than FPE.
- The ALL technique failed to outperform higher modeling orders (16 and 30) in weak (AA and AV), normal (AW and AY) and strong (IVB and AL) subjects.
- GA-based and PSO-based feature type fusion approaches performed as well or better than fixed modeling orders, the ALL feature vector, and the summation ensemble in all subjects.
- Weighted summation ensemble approach illustrated competitive results to the GA and PSO approaches in all subjects except with the weak subjects (AA and AV) in which this method outperformed all others.

5.2. Three Motor imagery Class Problem (IVc and V datasets)

In this section, average classification performance achieved on a sub-group of subjects that only performed three motor imagery tasks are considered and presented in Figure 6. These subjects include all participants in IVc (IVC) and V (V1, V2 and V3) datasets. Similar to Figure 5, fixed modeling orders are presented with their modeling order values and concatenation method is represented as ALL in the Figure. Following observations are made from the results illustrated in Figure 6.

- Fixed modeling order $p \in \{2, 4, 6, 8, 10, 16, 30\}$: the results across the subjects indicate the superiority of modeling orders 8 and 10. This is contradictory to the previous findings in 2 class datasets in which other higher orders are competitive or even achieved better average classification performances. This pattern is observed across all subjects within dataset V which contained 3 very strong subjects while in dataset IVc which contained a weak subject the higher modeling orders (16 and 30) slightly performed better than all other lower modeling orders.
- Conventional estimation methods (AIC, BIC, and FPE) performed as well as higher modeling orders (16 and 30) in the weaker subject (e.g., IVC). The conventional approaches performed similarly to the highest modeling order in the remaining subjects in this category (V1, V2, and V3) as well but they failed

to outperform the best performing fixed modeling orders (8, and 10) in these subjects. In addition, considerable performance differences are observed between the ALL and the conventional approaches on V3 in which AIC, BIC, and FPE been less successful. These methods performed better than ALL on subject V1 and V2 however the difference of the mean classification performance is in order of 5% and likely to be non-significant.

- The ALL technique failed to outperform best performing modeling orders (e.g. 8 and 10) in subject V1 and V2. However, this approach performed better than all other fixed modeling orders in subject IVC and V3. This method outperformed all other techniques in subject V3.
- GA-based and PSO-based fusion approaches performed better than any fixed modeling order, the ALL feature vector, conventional estimation techniques, and the summation ensemble in subjects IVC and V1. GA performed poorly on subject V2 but PSO managed to outperform all other methods on this subject. GA outperformed PSO and all other approaches on subject V3 with the exception of ALL method which is the best performing technique in this subject.
- No obvious advantage is observed over summation and weighted summation ensemble approaches within either of subjects with exception of IVC in which weighted summation ensemble outperformed all other modeling methods.

5.3. Four Motor imagery Class Problem (IIIa dataset)

In this section, average classification performance achieved on a sub-group of subjects that performed four motor imagery tasks are considered. The results are presented in Figure 7. These subjects include all participants in IIIa (k3b, k4b and l1b) dataset. The method labeling utilized in this Figure is similar to Figures 5 and 6. Following observations are made from the results:

- Fixed modeling order $p \in \{2, 4, 6, 8, 10, 16, 30\}$: the results indicate the superiority of modeling orders of 16 and 30 with exception of subject L1B in which model order 30 is the best fixed modeling order. Unlike 2 and 3 class problems in which clear performance improvement is observed when modeling order is increased, in two weaker subjects within this category (K6b and L1b), lower modeling orders illustrated some advantages. In subject L1B, AR model orders of 4 and 6 performed better than higher modeling orders of 8, 10 and 16 and in subject K6B, AR model order of 4 performed closely to model orders of 8, 10 and 16.
- Conventional estimation methods (AIC, BIC, and FPE) are outperformed with AR modeling order 30 in L1B subject while they performed similar to modeling order 30 in K3B and K6B. These techniques performed better than all lower modeling orders across subjects. These methods performed inconsistently against the ALL technique. That is, AIC, BIC, and FPE outperformed ALL classification performance on K6B while ALL performed better on L1B. It should be noted that the performance differences between these methods and ALL is in order of 0.1% and considered to be neglect-able.

- The *ALL* technique failed to outperform best performing modeling order (e.g. 30) across all subjects. This is with understanding that the performance difference between *ALL* and the best performing fixed modeling order is dismissable across all subjects.
- GA and PSO-based fusion approaches performed better than or on pare with fixed modeling orders, the *ALL* technique, and the summation ensemble across all subjects and also outperformed the weighted summation ensemble-based mechanism on subject K3B.
- Weighted summation ensemble performed better than all other methods on subjects K6B and L1B (weaker subjects) while it is out-performed by PSO and GA on stronger subject (K3B).

Previous sections discussed the performances achieved with methods considered in the study on the basis of variations in tasks performed by subjects rather than considering cross n-class problem and strengths of subjects. In order to gain better understanding of the achievements, the findings are reassessed on the basis of 3 categories of datasets, number of motor imagery classes performed by subjects, and the classification strength of subjects. The last category is considered in order to also investigate the lack of performance consistency across AR modeling order estimation approaches in the presence of multiple subjects with varying EEG performance claimed by Palaniappan Raveendran (2001). Although bookmaker informedness is considered as the primary performance measurement metrics due to being normalized and unbiased, average accuracy results are also presented in following section. This is with the understanding that since the measured accuracy on 2, 3, and 4 class problems are not directly comparable with each other, first these values are rescaled to a common n-class problem (n=4 in here) and later the average value is calculated. That is, all 2-class and 3-class accuracy results are multiplied by 0.5 and 0.75 respectively prior to estimating the mean value.

5.4. Category 1: Datasets

Average classification performances of the approaches considered in the study are presented in table 4. The results are reported in the order of dataset complexity in terms of the number of motor imagery classes performed by subjects. The results indicate that weighted sum ensemble approach outperforms all other methods across all datasets. This performance is closely followed by PSO and GA.

Considering experiments with modeling orders in the range of 2 and 30 and the condition referred to as *All* (concatenation of all AR modeling orders), in most datasets, modeling order of 30 and *All* are the best performing approaches. This is with the exception of dataset V in which AR model order 8 is the best performing method.

5.5. Category 2: n Class problem

Table 5 represents the average performance across subjects and databases. The results are presented on the basis of the number of motor imagery classes performed by subjects in each dataset. The results indicate that weighted sum ensemble approach outperforms all other methods across all datasets. This performance is closely followed by PSO and GA. The results indicate slight advantage for PSO over GA across all n-class problems.

Considering experiments with modeling orders in the range of 2 and 30 and the condition referred to as *All*, across all datasets, modeling order 30 outperformed other approaches in 2 and 4 class problems. Modeling orders 8 and 10 performed better than other methods in the 3 class problem.

Conventional estimation methods performed as well as or close to the best fixed modeling orders across all n-class problems with FPE consistently reporting better average classification performance in comparison to AIC and BIC methods.

5.6. Category 3: Subject EEG pattern classification strength

In this category, average performance achieved by subjects are presented on the basis of subjects' classification strengths. The subjects' taxonomies considered in here are *strong*, *normal*, and *weak*. These categories are selected intuitively by considering all subjects who achieved average classification performance above 0.3 bookmaker informedness as *strong* (6 subjects across datasets), in the range of 0.2 and 0.3 bookmaker informedness as *normal* (2 subjects across datasets) and any subject having average bookmaker informedness lower than 0.2 as *weak* (5 subjects across datasets). It should be noted that the results presented in table 6 are averaged across subjects regardless of their originating datasets and the number of motor imagery classes performed by them.

Looking at the experiments with modeling orders in the range of 2 and 30 and the condition referred to as *All*, in all subject categories, *All* failed to outperform best performing fixed modeling orders. Condition with modeling order 30 performed favorably within *weak* and *normal* subject categories while modeling order 16 performed better in the *strong* category.

Similar to previous categories, FPE achieved better average performance in comparison with AIC, BIC, and *All*. FPE outperformed the best performing fixed modeling order in *weak* and *normal* category but it is outperformed by the best performing fixed modeling order in the *strong* category ($p = 16$).

Weighted summation ensemble method performed dominantly on *weak* subject category while it shared success with PSO-based fusion on *normal* category and been out-performed by this method on *strong* category. GA-based fusion is the second best performing method across all categories in this table.

A detailed statistical analysis of the results is conducted and presented in the supplement. The overall results indicated lack of significant differences between PSO-based fusion and weighted summation ensemble methods while statistical significances are observed between these methods and other approaches considered in this study.

6. Discussion

6.1. Fixed and Conventional modeling order estimation methods

The results achieved with variations of fixed modeling orders and conventional estimation methods indicate inconsistency in the achieved performance across subjects. Evidences of this are observed in subjects involved in 2-class (Figures 5), 3-class (Figure 6) and 4-class (Figure 7) problems. That is, although inconsistency in results across subjects of different datasets performing similar number of motor imagery tasks can be interpreted as the effect of utilizing different experimental setups, equipments or types of motor imagery tasks, however, no consistent performance is observed across subjects of the same dataset which confirms lack of consistency across subjects as reported by Palaniappan Raveendran (2001). That is, neither of the studied AR-orders been consistently the best performing order across subjects and the winning fixed AR-order varied across subjects. Such inconsistency is also observed across datasets, subject's EEG strength, and n-class problems as reported in tables 4, 5 and 6.

The results with conventional estimation methods confirms Palaniappan Raveendran (2001) findings in terms of lacking a clear and tangible advantage across such estimation methods. That is, in terms of classification performance, no significant difference is observed between AIC, BIC and FPE methods (neither across datasets nor n-class problems as presented in previous section).

The results partially confirmed Krusienski et al. (2006) findings in terms of favoring higher AR modeling orders. In the sense of modeling order estimation, this is with the exception of FPE that in two occasions (with subjects AA and IVB) choose to utilize lower AR-orders (19 and 4 respectively). In the sense of classification performance, although this is confirmed in all subjects in dataset IIIa (4-class problem) and most subjects in datasets IVa and IVb (2 class problems) and dataset IVc (3 class problem), a clear disadvantage is observed within all subjects in dataset V (3 class problem) in which modeling orders of 8 and 10 scroed a better classification performance (on average).

6.2. AR modeling order selection results with GA and PSO

Tables 7, 8 and 9 present the most commonly selected modeling orders by either GA or PSO methods on the datasets utilized in the study. The results are presented on the basis of average length of best performing population member of each evolutionary method across 100 folds in the 10 repetition of the 10-fold cross validation scheme utilized in the study. In order to gain better understanding of various modeling orders' contributions in the achieved performance, collections of breakdowns are offered based on subjects' strength and overall contributions within the n-class problems.

In datasets with 2-class problems (IVb and IVa presented in table 7), a variation in the chosen AR modeling orders is observed across both GA and PSO with regards to the subjects' strengths. For example, both GA and PSO neglected 30 in weak subjects while GA, unlike PSO, also ignored 2 for both weak subjects. Both approaches considered 8 in their final combined feature vectors for both weak subjects. Unlike weak subjects, GA and PSO utilized 30 in their final products for both normal subjects. GA and PSO reported a disagreement on the use of 8 on this category while GA also hesitated to included 16 in its solutions for neither of the normal subjects. It is noteworthy that PSO consistently included 6 and 10 in its solutions for all variations of subjects' strength within datasets with 2-class problem. In addition, it is noticeable that PSO considered a larger variations of AR modeling orders within its solutions. While GA considered more contributions, on average, from higher modeling orders (e.g., 10, 16 and 30) in its final solutions, PSO retained a fair balance between the contributions of higher and lower (e.g., 2, 4 and 6) modeling orders.

From the results attained with 3-class problems (IVc and V presented in table 8), a noticeable difference in the chosen modeling orders is observed. That is, for example, in IVC subject, while GA only considered a handpick of modeling orders in its solutions, PSO included all modeling orders but 30. Both PSO and GA agreed on omitting 2, 4 and 30 when dealing with subject V1 and they also agreed on the use of 8 and 10 on subject V2 with PSO also including 4 in its solutions. It is noticeable that both approaches are in agreement on not utilizing 30 in either of solutions for subjects within 3-class problem while GA also ignored 4 in this category. Similar to 2-class problem, PSO considered a wider range of modeling orders in its solutions while GA preferred 8, 10 and 16 in most cases. In addition, PSO favored 8 in all its solutions within this category. Both GA and PSO are in agreement to utilize 8 when they are dealing with strong subjects in this category.

The results from 4-class problem (reported in table 9) once more highlighted the difference in the choices made by these two algorithms. That is, while PSO consistently utilized modeling orders of 2, 4, 6 and 8 in its solutions for both weak and strong subjects, GA preferred to use different arrangements. It is noticeable that when dealing with the strong subject (K3B), GA and PSO considered a similar range of AR-orders in the solutions with exception of GA neglecting 4 in addition to 10 while PSO only ignored the second and utilized all other AR-orders in its solution. Another major difference is observed with in subject K6B in which GA favored higher modeling orders (8-30) while PSO mostly focused on lower modeling orders (2-10).

7. Conclusion

Autoregressive (AR) is one of the commonly employed feature types in EEG single trial studies. Despite known advantageous of AR such as having low computation cost, better spectral resolution, smooth spectra and being applicable to short segments of data, identifying AR's model order is an important step that influences its performance in EEG analysis and classification. Low and high model orders are likely to represent poor signal representation and noise & inaccuracy respectively. The study hypothesized that an adequate mixture of AR features derived from various AR modeling orders is a better representative of the underlying signal compared with any fixed modeling order. This hypothesis was assessed using two mechanisms for identifying adequate mixture of AR modeling orders. These mechanisms included i) classifier mixture in the form of Ensemble Learning architecture and ii) Evolutionary based fusion of features originating from a range of modeling orders. The feasibility of these mechanisms were assessed against three sets of commonly employed approaches such as i) Fixed AR-orders that reported adequate results in similar studies (AR-orders of 2 ,4 ,6, 8, 10, 16 and 30 were used in here), ii) blind mixture of a set of well-known modeling orders and iii) conventional modeling order estimation algorithms (e.g., AIC, BIC and FPE). The feasibility of these approaches were assessed on the basis of their classification performance on five datasets from BCI competition III that contained 2, 3 and 4 motor imagery tasks. The results were investigated based on the averaged classification performance of the methods on each subject in each dataset. In addition, three categories were considered to provide better understanding of the achievements. These categories were i) the performance across datasets, ii) the performance across considered n-class problems, and iii) the performance across subjects' EEG pattern classification strengths.

Table 10 provides an overview of the best performing algorithms within all categories considered in this study. The results indicated superiority of Ensemble-based approach when weighted summation operator was utilized to aggregate the results from classifiers in the ensemble. In addition, the results highlighted feasibility of evolutionary-based fusion methods within all datasets. Considering the contributing subject strength in the sense of EEG classification performance, between all methods considered in the study, PSO-based fusion technique reported better performance under *strong* group while weighted summation ensemble method performed well under the *weak* and *normal* subjects. This is with the understanding that no statistical significant difference been observed between these two methods.

The consistent superior performances achieved by the PSO-fusion and the weighted ensemble techniques across all variations considered in the study (nclass problem, subject strength, and datasets) can be explained by their ability to include multiple representations of spectral patterns originating from different AR-orders which better captures the underlying pattern of the performed tasks by subjects while other methods (AIC,BIC,FPE and AR models of fixed orders) favored one representative pattern over the rest which likely resulted in distortion or ill-representation of the activities performed. In addition, owing to their learning capabilities, PSO-fusion and weighted ensemble methods prevented inclusion of AR-orders that poorly represented the subjects' intentions while *ALL* approach blindly concatenated all AR-orders which degraded its performance.

These findings approved the initial hypothesis and suggested weighted ensemble and PSO-based fusion as two candidates for identifying the adequate mixture of modeling orders.

7.1. Contributions of the study

Following contributions can be considered for the study:

- Two new mechanisms for automatically identifying the optimal mixture of AR features with varying modeling orders are proposed by the study.
- Assessment of the hypothesis is done on five publicly available datasets which captures important aspects such as i) replicate-ability of the study, ii) applicability of the proposed methodology to different EEG recording setups, protocols, and experimental paradigms and iii) robustness of the mechanisms proposed against a wide range of participating subjects in EEG and BCI studies.

- Comparison of the performance achieved with the mechanisms proposed against the three set of well-known AR-order identification and estimation approaches that are commonly utilized by BCI and EEG communities in similar studies. The inclusion of conventional order estimation methods such as FPE, AIC, and BIC in addition to covering a range of well-studied modeling orders by BCI community leaves minimum doubt on the superiority of the novel idea of utilizing optimal mixture of AR features with varying modeling orders.

To the best of our knowledge, the idea of identifying the optimal mixture of AR modeling orders rather than finding the optimal unique AR-order never been entertained by either EEG or BCI communities.

7.2. Limitations of the study

The limitations of the current findings are as follows:

- The current study only captures the application of AR features in single trial EEG classification. Albeit the idea of identifying optimal mixture of modeling orders is being proven promising, however, this can not be extended to other applications of AR in digital signal processing such as artifact rejection, general power spectral analysis, and its other possible applications.
- Within the context of real-time BCI systems, it should be noted that although PSO-based fusion method performed well under the assessed datasets, however, the computational intensiveness of PSO algorithm which lies within its inherent dependency to governing a high population of learning particles makes it less favorable. Under such circumstances, the Ensemble-based mixture mechanism seems to be more appropriate since it mainly relies on fast learning classification algorithms such as ELM.
- Authors advise caution against substituting the ELM algorithm with stronger learners such as SVM specially if a real-time BCI system is the targeting application. As mentioned in section 4.1, the choice of using ELM with only 80 hidden nodes is made due to its fast learning capability. Although using higher number of hidden nodes or replacing the algorithm with more efficient learners such as multi-layer perceptron, random forest, or SVM can improve the overall classification performance of the Ensemble, it would be with major negative impact on learning time of the Ensemble.

7.3. Future Work

Possible future directions of the current study are as follows:

- The main future direction of the current study is application of the proposed automatic mixture of AR modeling orders in the context of spectral analysis of EEG signals of patients suffering from Autism Spectrum Disorder (ASD). To the best of our knowledge, AR features are rarely (if ever) considered as biomarkers in ASD studies. The post-process nature of such study allows inclusion of evolutionary-based fusion methods in the mix which provides better opportunity in terms of identifying strong and generalizable mixture of AR modeling orders that best distinguish the two groups of ASD and TD (typically developing) participants.

- It is desirable to assess the main hypothesis of this study in wider spectral analysis applications in image segmentation (Sarkar et al., 2016), multispectral image classification (Li et al., 2011) and spectral clustering (nkaya, 2015),
- As stated in the limitations, the application of evolutionary-based fusion of AR-orders is limited to post-processing stages. This limitation been revoked in similar studies using transfer learning (Atyabi et al., 2013). However, the study been only conducted on FFT features. The extent within which the transfer learning can be applied to spectral features and the suitability of subject inclusion/exclusion criteria discussed in (Atyabi et al., 2013) are not clear yet and requires more investigation. As of interest is to identify an AR-order mixture that can cover a high range of BCI paradigms while supporting variations in recording protocols and participating subjects. This can be achieved by heterogeneous transfer learning which is rarely considered by EEG or Data Science communities due to its high level of theoretical complexities.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Ab Wahab N, Nefti-Meziani S, Atyabi A. A Comprehensive Review on Swarm Optimization Algorithms. PLoS ONE. 2015; 10(5):e0122827. [PubMed: 25992655]

Abo-Zahhad M, Ahmed SM, Abbas SN. A new multi-level approach to EEG based human authentication using eye blinking. Pattern Recognition Letters. 2015:1–10.

Atyabi A, Fitzgibbon S, Powers D. Multiplication of EEG samples through Replicating, Biasing, and Overlapping. Brain Informatics BI12, Springer Lecture Notes in Computer Science. 2012a:209–219.

Atyabi, A., Luerssen, M., Fitzgibbon, S., Powers, D. Biasing the Overlapping and Non-Overlapping Sub-Windows of EEG recording; International Joint Conference on Neural Networks (IJCNN), IEEE World Congress of Computational Intelligence (WCCI); 2012b. p. 741-746.

Atyabi A, Luerssen M, Fitzgibbon S, Powers D. The Impact of PSO based Dimension Reduction in EEG study. Brain Informatics, Lecture Notes in Computer Science, Springer. 2012c; 7670:220–231.

Atyabi A, Luerssen M, Powers D. PSO-Based Dimension Reduction of EEG Recordings: Implications for Subject Transfer in BCI. Neurocomputing, Elsevier. 2013; 119:319–331.

Blanchard G, Blankertz B. BCI Competition 2003 - Data set IIa: spatial patterns of self-controlled brain rhythm modulations. IEEE Trans Biomed Eng. 2004; 51(6):1062–1066. [PubMed: 15188879]

Bonabeau, E., Dorigo, M., Theraulaz, G. Swarm Intelligence: From Natural to Artificial Systems. Oxford University Press; 1999.

Camilleri, TA., Camilleri, KP., Fabri, SG. Segmentation and labelling of eeg for brain computer interfaces. In: Azzopardi, G., Petkov, N., editors. Computer Analysis of Images and Patterns: 16th International Conference, CAIP 2015, Valletta, Malta, September 2–4, 2015 Proceedings, Part I. Cham: Springer International Publishing; 2015a. p. 288-299.

Camilleri, TA., Camilleri, KP., Fabri, SG. Semi-supervised segmentation of EEG data in BCI systems; Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS; 2015b. p. 7845-7848.

Chun Shan Wong WKL. On a mixture autoregressive model. Journal of the Royal Statistical Society. Series B (Statistical Methodology). 2000; 62(1):95–115.

Deng, W. Regularized Extreme Learning Machine; IEEE Symposium on Computational Intelligence and Data Mining, CIDM ’09; 2009. p. 389-395.

Dirgenali F, Kara S. Recognition of early phase of atherosclerosis using principles component analysis and artificial neural networks from carotid artery doppler signals. Expert Systems with Applications. 2006; 31(3):643–651.

Engin, M., Unsal, S., Engin, EZ. Order Selection in autoregressive power spectrum estimation of Sleep EEG; ELECO Second International Conference on Electrical & Electronics Engineering; 2001.

Eriolu E, Gnay S. Bayesian model selection in {ARFIMA} models. Expert Systems with Applications. 2010; 37(12):8359–8364.

Fang Y, Chen M, Zheng X. Extracting features from phase space of {EEG} signals in brain computer interfaces. Neurocomputing. 2015; 151(Part 3):1477–1485.

Fitzgibbon, S. Ph.D. thesis. Flinders University of South Australia; 2007. A Machine Learning Approach to Brain-Computer Interfacing.

Grefenstette, J. GENESIS. Navy Center for Applied Research in Artificial Intelligence; 1990.

Holland J. Genetic algorithms. Scientific American Journal. 1992:66–72.

Hsu W-Y. Assembling a multi-feature eeg classifier for leftright motor imagery data using wavelet-based fuzzy approximate entropy for improved accuracy. International Journal of Neural Systems. 2015; 25(08):1550037. [PubMed: 26584583]

Huang G-B, Zhou H, Ding X, Zhang R. Extreme Learning Machine for Regression and Multiclass Classification. IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics. 2011a: 1–17.

Huang G-B, Zhou H, Ding X, Zhang R. Extreme learning machines: a survey. Int. J. Mach. Learn & Cyber. 2011b; 2:107–122.

Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machines: Theory and applications. Neurocomputing. 2006; 70:489–501.

Inoue, K., Sugioka, K., Ishii, K., Pfurtscheller, G., Kumamaru, K. Pattern recognition of EEG signals during right and left motor imagery. SICE Annual conference; Japan. 2003. p. 2432-2437.

Jansen B, Bourne J, Ward J. Autoregressive estimation of short segment spectra for computerized EEG analysis. IEEE Trans Biomed Eng. 1981:630–638. [PubMed: 7319522]

Karahan E, Rojas-Lpez PA, Bringas-Vega ML, Valds-Hernndez PA, Valdes-Sosa PA. Tensor analysis and fusion of multimodal brain images. Proceedings of the IEEE. 2015; 103(9):1531–1559.

Kayikcioglu T, Maleki M, Eroglu K. Fast and accurate pls-based classification of eeg sleep using single channel data. Expert Systems with Applications. 2015; 42(21):7825–7830.

Kennedy J, Eberhart R. Ieee international conference on neural networks. Particle swarm optimization. 1995; 4:1942–1948.

Kotkatvuori-rnberg J. Measuring actual daily volatility from high frequency intraday returns of the s&p futures and index observations. Expert Systems with Applications. 2016; 43:213–222.

Krusienski, D., McFarland, D., Wolpaw, J. An Evaluation of Autoregressive Spectral Estimation Model Order for Brain-Computer Interface Applications; Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA; 2006. p. 1323-1326.

Kus R, Ginter JS, Blinowska KJ. Propagation of EEG activity during finger movement and its imagination. ACTA Neurobiologi Experiment. 2006; 66:195–206.

Li X, Nie P, Qiu Z-J, He Y. Using wavelet transform and multiclass least square support vector machine in multi-spectral imaging classification of chinese famous tea. Expert Systems with Applications. 2011; 38(9):11149–11159.

Li Y, Liu Q, Tan S-R, Chan RH. High-resolution time-frequency analysis of eeg signals using multiscale radial basis functions. Neurocomputing. 2016; 195:96–103. Learning for Medical Imaging.

Liang N-Y, Saratchandran P, Huang G-B, Sundararajan N. Classification of Mental tasks from EEG signals using Extreme Learning Machines. Int. J. Neural Systems. 2006; 16(1):29–38.

Liu K, Yu ZL, Wu W, Gu Z, Li Y. Straps: A fully data-driven spatio-temporally regularized algorithm for m/eeg patch source imaging. International Journal of Neural Systems. 2015; 25(04):1550016. [PubMed: 25903226]

Ligeois, R., Mishra, B., Zorzi, M., Sepulchre, R. Sparse plus low-rank autoregressive identification in neuroimaging time series; 2015 54th ieee conference on decision and control (cdc); 2015. p. 3965-3970.

Loukas A, Simonetto A, Leus G. Distributed autoregressive moving average graph filters. IEEE Signal Processing Letters. 2015; 22(11):1931–1935.

Nagata, K., Mihara, M., Yamagutchi, T., Taniguchi, M., Inoue, K., Pfurtscheller, G., Kumamaru, K. Pattern recognition of EEG signals during motor imagery; International Joint conference SICE-ICASE; 2006. p. 5169-5173.

Ning T, Bronzino J. Autoregressive and bispectral analysis techniques: EEG applications. IEEE Eng. in Med. and Bio. Mag. 1990:47–50. [PubMed: 18238318]

Palaniappan R. Towards optimal model order selection for autoregressive spectral analysis of mental tasks using genetic algorithm. International Journal of computer science and network security IJCSNS. 2006a; 6(1):153–162.

Palaniappan R. Towards Optimal Model Order Selection for Autoregressive Spectral Analysis of Mental Tasks Using Genetic Algorithm. IJCSNS International Journal of Computer Science and Network Security. 2006b; 6(1A):153–162.

Palaniappan, R., Raveendran, P. Autoregressive order selection criteria: a case study for EEG signals; CD Proceedings of International Student Conference on Research and Development; 2001.

Peiyang L, Xurui W, Fali L, Rui Z, Teng M, Yueheng P, Xu L, Yin T, Daqing G, Tiejun L, Dezhong Y, Peng X. Autoregressive model in the Lp norm space for EEG analysis. Journal of Neuroscience Methods. 2015; 240:170–178. [PubMed: 25448380]

Pippa, E., Mporas, I., Megalooikonomou, V. Automatic Estimation of the Optimal AR Order for Epilepsy Analysis Using EEG Signals; IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE); 2015. p. 1-4.

Powers, D. The Problem of Kappa; Proceedings of the 13th Conference of the European Chapter of the Association for computational Linguistics; 2012. p. 345-355.

Powers, DMW. Recall and Precision versus the Bookmaker; International Conference on Cognitive Science (ICSC); 2003. p. 529-534.

Rotondi F, Franceschetti S, Avanzini G, Panzica F. Altered eeg resting-state effective connectivity in drug-nave childhood absence epilepsy. Clinical Neurophysiology. 2016; 127(2):1130–1137. [PubMed: 26437574]

Rubega M, Sparacino G, Sejling AS, Juhl CB, Cobelli C. Hypoglycemia-induced decrease of eeg coherence in patients with type 1 diabetes. Diabetes Technology & Therapeutics. 2016; 18(3):178–184. [PubMed: 26745007]

Sarkar S, Das S, Chaudhuri SS. Hyper-spectral image segmentation using rnyi entropy based multi-level thresholding aided with differential evolution. Expert Systems with Applications. 2016; 50:120–129.

Schlogl, A., Filz, O., Ramoser, H., Pfurtscheller, G. Technical Report Technical Report. 2004. GDF - A general data format for biosignals.

Shahabi H, Moghimi S. Toward automatic detection of brain responses to emotional music through analysis of EEG effective connectivity. Computers in Human Behavior. 2016; 58:231–239.

Shaw, L., Routray, A. Efficacy of adaptive directed transfer function for neural connectivity estimation of eeg signal during meditation; Signal processing and integrated networks (spin), 2nd international conference on; 2015. p. 198-202.

Tang X, Han M. Partial lanczos Extreme learning machines for single-output regression problems. Neurocomputing. 2009; 72:3066–3076.

Tsoi, AC., So, DSC., Sergejew, A. Classification of Electroencephalogram using artificial neural networks. Advances in Neural Information Processing Systems 6, 7th NIPS Conference; Denver, Colorado, USA. 1993. p. 1151-1158.

Vaz F, Guedes de Oliveira P, Principe J. A Study on the Best Order for Autoregressive EEG Modelling. Int. J. Bio-Medical Computing. 1987; 20:41–50.

Vedavathi K, Rao KS, Devi KN. Unsupervised learning algorithm for time series using bivariate ar(1) model. Expert Systems with Applications. 2014; 41(7):3402–3408.

Vijayan, AE., Sen, D., Sudheer, A. EEG-Based Emotion Recognition Using Statistical Measures and Auto-Regressive Modeling; Computational Intelligence & Communication Technology (CICT), IEEE International Conference on; 2015. p. 587-591.

Wang C-C, Kang Y, Shen P-C, Chang Y-P, Chung Y-L. Applications of fault diagnosis in rotating machinery by using time series analysis with neural network. Expert Systems with Applications. 2010; 37(2):1696–1702.

Wolpaw JR, McFarland DJ, Vaughan TM. The classification of Transient time-varying EEG signals via Wavelet Packets decomposition. Brain–Computer Interface Research at the Wadsworth Centre, IEEE Transactions on Rehabilitation Engineering. 2000; 8(2):222–226.

Wu Y, Jiao B, Wu Z, Zhen J, Jia Q, Zhang H, Guan B, Wang S. Autoregressive spectral analysis of cortical electroencephalographic signals in a rat model of post-traumatic epilepsy. Neurological Research. 2015; 37(11):959–966. [PubMed: 26215653]

Xu BG, Song AG. Pattern recognition of motor imagery EEG using wavelet transform. Journal of Biomedical Science and Engineering. 2008; 1:64–67.

Yonghui F, Minyou C, Xufei Z. Extracting features from phasespace of EEG signals in brain computer interfaces. Neurocomputing. 2015; 151:1477–1485.

Zhao C, Zheng C, Zhao M, Tu Y, Liu J. Multivariate autoregressive models and kernel learning algorithms for classifying driving mental fatigue based on electroencephalographic. Expert Systems with Applications. 2011; 38(3):1859–1865.

Zhu Q-Y, Qin A, Suganthan P, Huang G-B. Evolutionary Extreme learning machines. Pattern Recognition. 2005; 38:1759–1763.

nkaya T. A parameter-free similarity graph for spectral clustering. Expert Systems with Applications. 2015; 42(24):9489–9498.

Highlights

- Two methods for mixing AR features for EEG signal classification are proposed
- Evolutionary and ensemble learning methods are considered
- The results are assessed against a set of conventional order estimation methods
- The feasibilities are investigated using several BCI competition datasets
- Adequacy of Ensemble-based mixture and EA-based fusion methods are shown

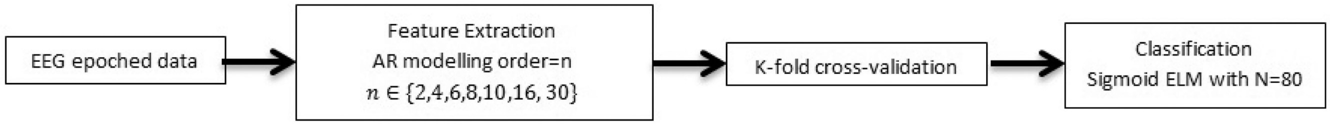

Figure 1.

Diagram flow of assessing classification performance of AR features with various modeling orders.

Author Manuscript
Author Manuscript
Author Manuscript

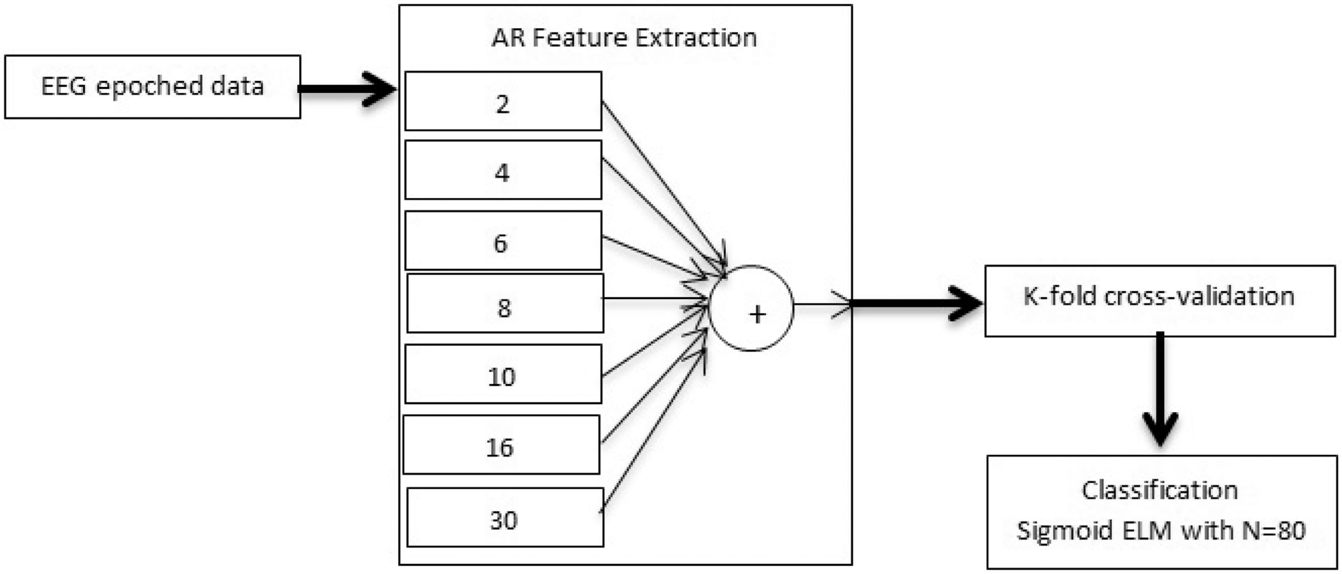

Figure 2.
Diagram flow of assessing classification performance of AR features with concatenation method. The '+' sign represent simple vector concatenation.

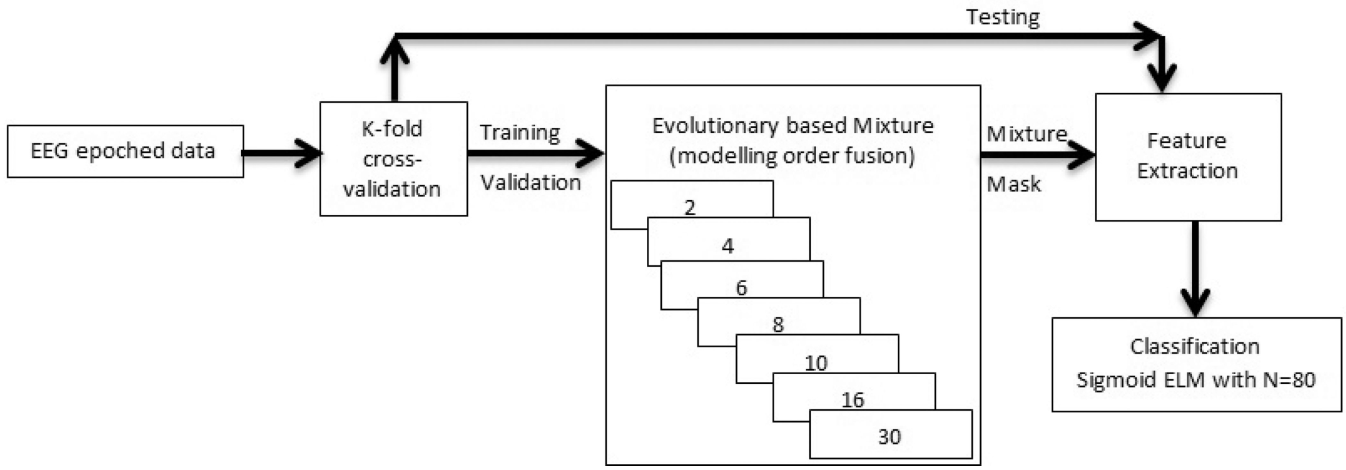

Figure 3.
Diagram flow of assessing classification performance of evolutionary-based mixture of AR features.

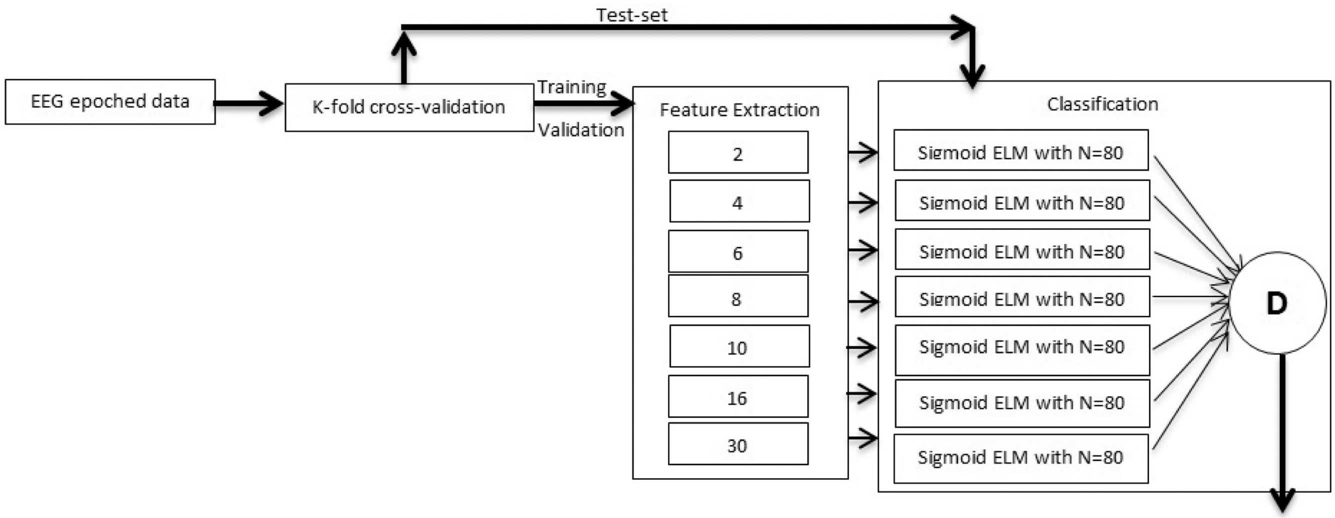

Figure 4.
Diagram flow of assessing classification performance of Ensemble-based mixture of AR modeling orders. 'D' operator represent decision aggregation operator.

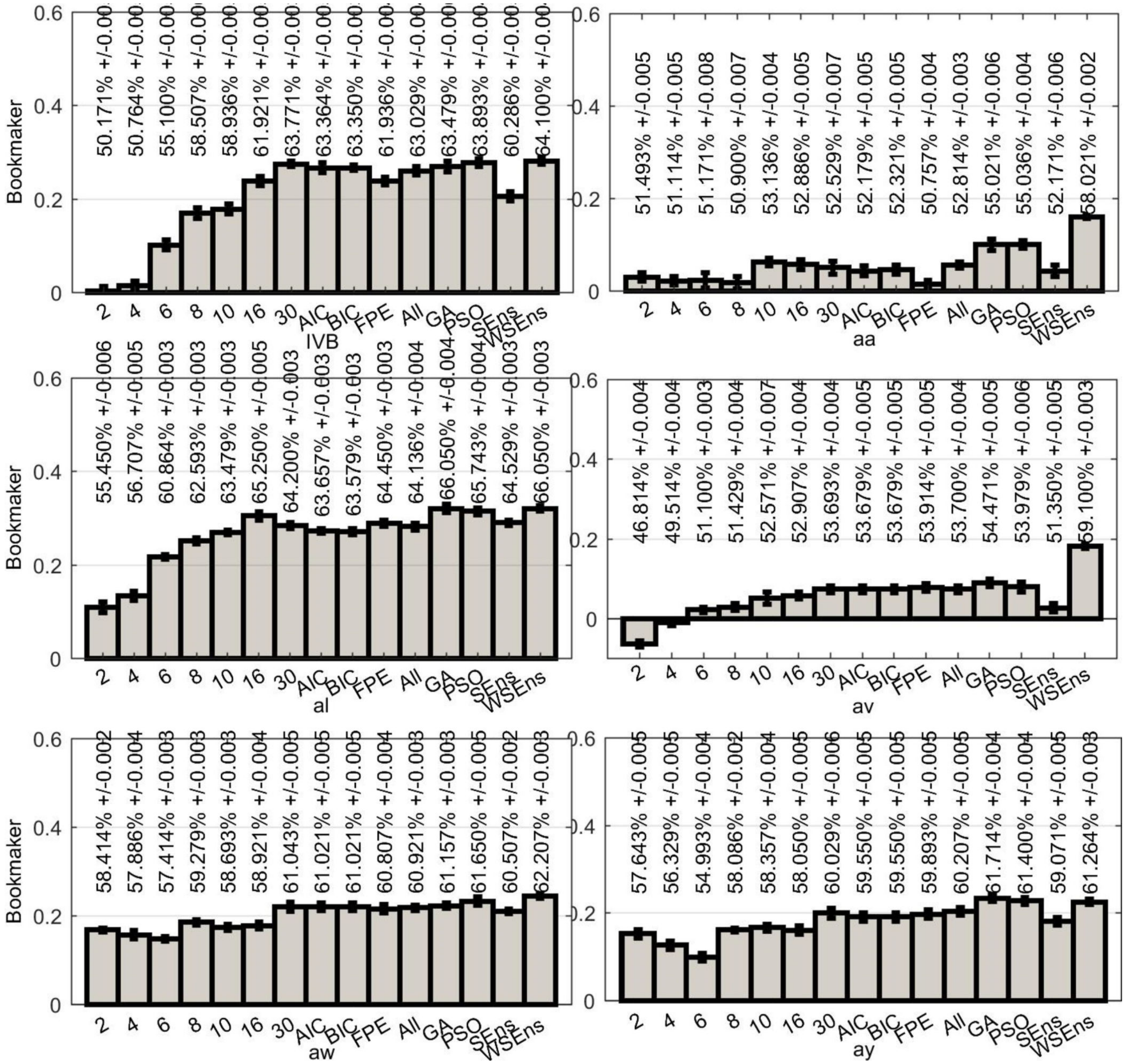

Figure 5.

Average bookmaker performance achieved with subjects performing two motor imagery tasks. Error bars are representing standard errors. Average accuracies and their associated standard errors are presented as additional texts on the bars. SEns and WSEns represent summation and weighted summation ensembles.

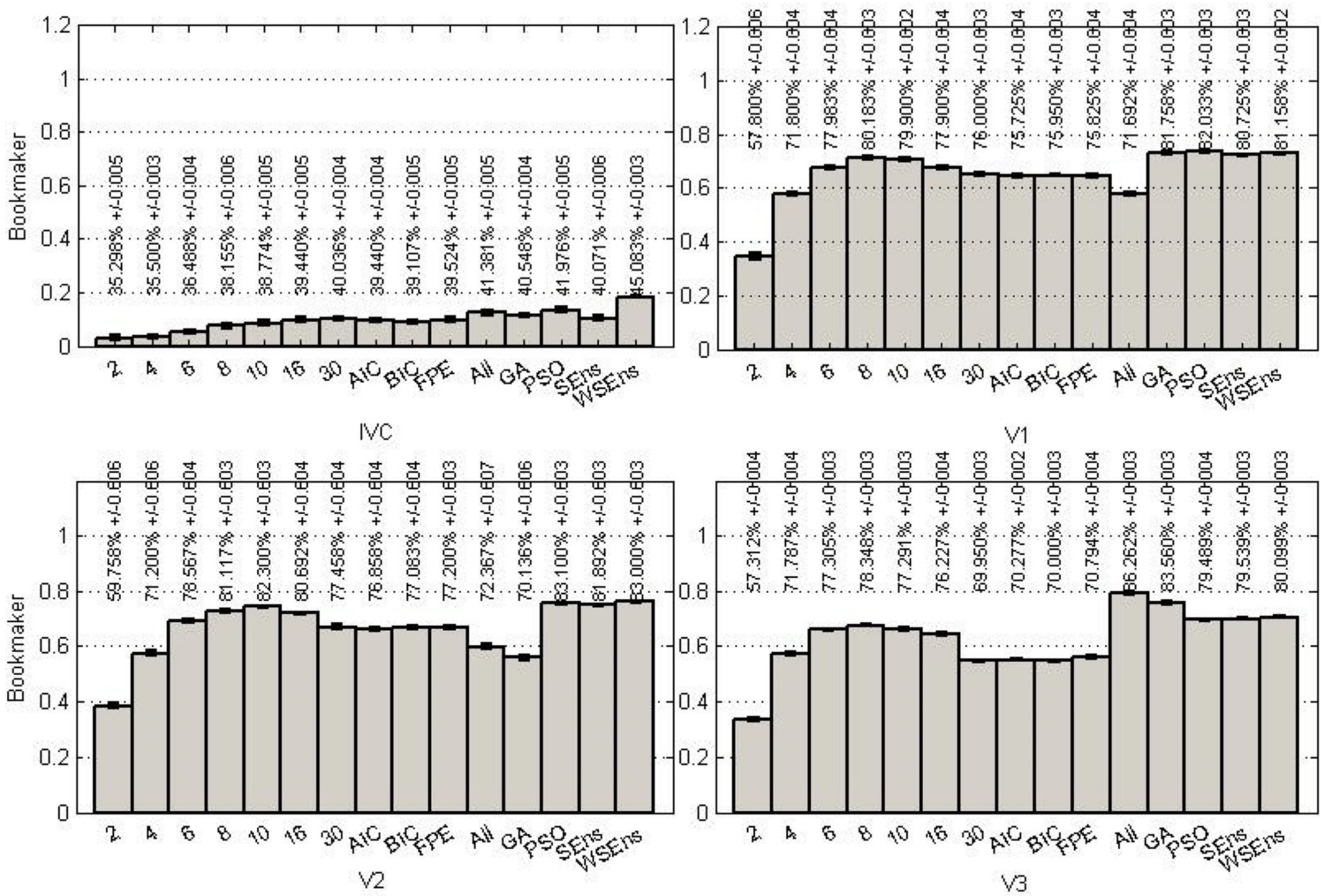

Figure 6.
Average bookmaker performance achieved with subjects performing three motor imagery tasks. Error bars are representing standard errors. Average accuracies and their associated standard errors are presented as additional texts on the bars. SEns and WSEns represent summation and weighted summation ensembles.

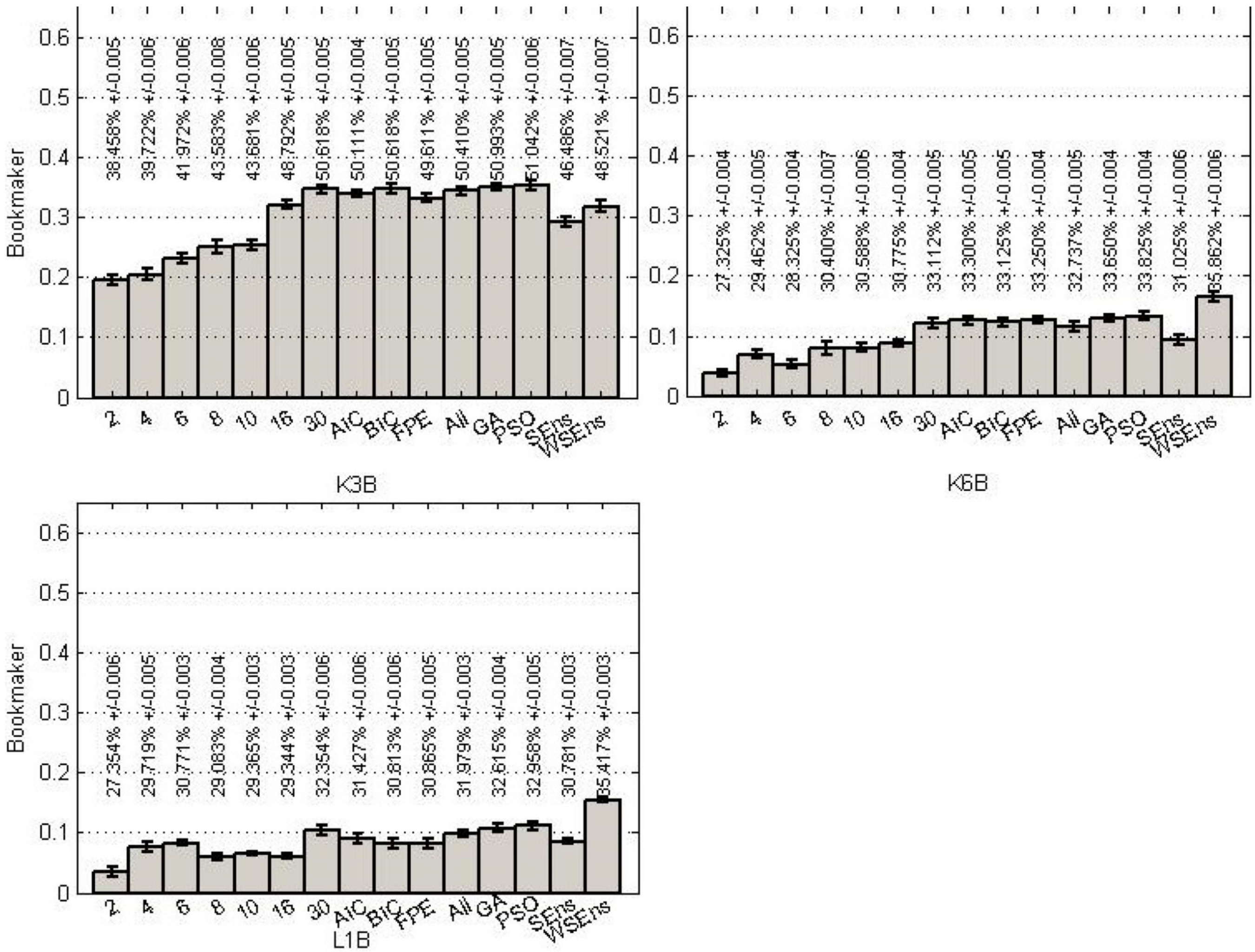

Figure 7.

Average bookmaker performance achieved with subjects performing four motor imagery tasks. Error bars are representing standard errors. Average accuracies and their associated standard errors are presented as additional texts on the bars. SEns and WSEns represent summation and weighted summation ensembles.

Table 1

Current trends in autoregressive-based studies

Reference	Scope of study	AR-order and the order estimation method
Peiyang et al. (2015)	Power spectrum analysis of the resting state EEG signal in the presence of artifacts and outliers	Lp norm ($p \leq 1$)
Camilleri et al. (2015b)	Segmentation of EEG data in a BCI study	Semi-supervised model learning
Shahabi Moghimi (2016)	Connectivity patterns of EEG electrodes under emotional responses to a set of Iranian and classical music	p=6, AIC
Pippa et al. (2015)	Epileptic seizure detection	Combinations of statistical features and regression analysis resulting in upto 4 unita estimation error
Yonghui et al. (2015)	EEG Phase analysis and classification	p=8 for Graze 2003 dataset and subject O3 from Graze 2005, p=5 for subjects S4 and X11 from Graze 2005. Repeated 10-fold cross validation
Abo-Zahhad et al. (2015)	Human authentication based on the mixture of EEG and eye-blinks recordings	p= 50, investigator's intuition
Kayikcioglu et al. (2015)	Classifications of sleep and wake stages	p=22, investigator's intuition.
Zhao et al. (2011)	EEG classification	p=3, AIC
(Ligeois et al., 2015)	Neuro-imaging	p=1, investigator's intuition.
(Camilleri et al., 2015a)	Segmentation and classification in EEG-based BCI	p=6 after evaluating a range of 2–10
(Li et al., 2016)	Non-stationary EEG analysis in time-frequency domain	p=15 for seizure EEG signal and p=9 for healthy participants, FPE
(Loukas et al., 2015)	Graph based signal filtering	Combinations of p=1 and p= arbitrary value are used as orders in time and graph domains respectively.
(Karahan et al., 2015)	Multi-modal brain image fusion for parsing the brain structures that reflect human cognitive processes and brain structural and regional connectivity	p=not specified, BIC
(Liu et al., 2015)	Adaptive source imaging via processing spatio-temporal information of patch source	Recursive penalized least squares procedure is used for model estimation
(Rubega et al., 2016)	Investigation of EEG coherence in type 1 diabetes patients	p= 9 and 10, AIC
(Shaw Routray, 2015)	estimation of neural connectivity of EEG recordings during meditation	p=10, AIC
(Wu et al., 2015)	Spectral analysis of cortical EEG recordings in a rat-based epileptic seizure study	not specified.
(Rotondi et al., 2016)	Investigation of EEG connectivity via partial directed coherence (PDC) childhood absence epilepsy	p=not specified, AIC
(Fang et al., 2015)	Implications of phase based feature classification on EEG-based BCI system	$p \in [5,8]$, 10-fold cross validation

Table 2

Details of datasets utilized in the study. All datasets are down sampled to 250Hz in preprocessing stage.

Dataset	Number of EEG Channels	Task Trials	Sample Rate	Task Duration(s)	Number of Participants	Performed Tasks
BCI Competition III, Dataset V	32	1903	512Hz	0.5s	3 subjects V1, V2, V3	right-hand, left-hand, word
BCI Competition III, Dataset IVb	118	210	1000Hz	3.5s	1 subject IVB	right-foot, left-hand
BCI Competition III, Dataset IVc	118	420	1000Hz	1s	1 subject IVC	right-foot, left-hand, relax
BCI Competition III, Dataset IVa	118	280	1000Hz	3.5s	5 subjects AA, AL, AV, AW, AY	right-foot, left-hand
BCI Competition III, Dataset IIIa	60	240/360	250Hz	3s	3 subjects K3B, K6B, L1B	foot, tongue, right-hand, left-hand

Table 3

Minimum value and its associated estimation of AR modeling order with AIC, BIC, and FPE approaches for modeling order(n) with $1 \leq n \leq 30$.

Approach	AR modeling order values												
	V			IVb	IVc	IVa					IIIa		
	V1	V2	V3	IVB	IVC	AA	AL	AV	AW	AY	K3B	K6B	L1B
FPE min est. value* 1.0e+004	30 1.4761	30 2.8989	28 0.8991	19 0.4661	30 0.7586	4 0.0113	23 0.0031	24 0.0020	28 0.0051	25 0.2665	21 0.0006	27 0.0004	27 0.0003
AIC min est. value* 1.0e+008	29 0.9488	28 1.0079	30 0.9131	29 2.4560	29 1.4899	30 1.5698	28 1.3274	26 1.2229	29 1.4262	21 2.2161	30 0.5067	30 0.3045	28 0.2785
BIC min est. value* 1.0e+008	29 0.9488	28 1.0079	30 0.9131	29 2.4560	29 1.4899	30 1.5698	28 1.3274	26 1.2229	29 1.4262	21 2.2161	30 0.5067	30 0.3045	28 0.2785

Table 4

Average classification performance(bookmaker informedness and accuracy) across subjects on datasets considered in the study. The best performing approach is identified using a **bold font**.

AR modeling order value	IVb	IVa	IVc	V	IIIa
2	0.0034286 & 0.25%	0.079257 & 0.27%	0.030013 & 0.26%	0.35617 & 0.44%	0.088289 & 0.31%
4	0.015286 & 0.25%	0.0862 & 0.27%	0.034677 & 0.27%	0.57616 & 0.54%	0.11641 & 0.33%
6	0.102 & 0.28%	0.10217 & 0.28%	0.05104 & 0.27%	0.67599 & 0.58%	0.12204 & 0.34%
8	0.17014 & 0.29%	0.12914 & 0.28%	0.076395 & 0.29%	0.70534 & 0.6%	0.12963 & 0.34%
10	0.17871 & 0.29%	0.14494 & 0.29%	0.08402 & 0.29%	0.70373 & 0.6%	0.13303 & 0.35%
16	0.23843 & 0.31%	0.15206 & 0.29%	0.096711 & 0.3%	0.6801 & 0.59%	0.15592 & 0.36%
30	0.27543 & 0.32%	0.16597 & 0.29%	0.10347 & 0.3%	0.62298 & 0.56%	0.19042 & 0.39%
All {2,4,6,8,10,16,30}	0.26729 & 0.32%	0.16034 & 0.29%	0.095546 & 0.3%	0.61963 & 0.56%	0.18423 & 0.38%
AIC	0.267 & 0.32%	0.1606 & 0.29%	0.089821 & 0.29%	0.62042 & 0.56%	0.18382 & 0.38%
BIC	0.23871 & 0.31%	0.15929 & 0.29%	0.097783 & 0.3%	0.62482 & 0.56%	0.17991 & 0.38%
FPE	0.26057 & 0.32%	0.16711 & 0.29%	0.12381 & 0.31%	0.65736 & 0.58%	0.18587 & 0.38%
GA-Fusion	0.26957 & 0.32%	0.19366 & 0.3%	0.11411 & 0.3%	0.68377 & 0.59%	0.19592 & 0.39%
PSO-Fusion	0.27786 & 0.32%	0.19123 & 0.3%	0.13423 & 0.31%	**0.73005 & 0.61%**	0.19858 & 0.39%
S-Ensemble	0.20571 & 0.3%	0.15051 & 0.29%	0.10254 & 0.3%	0.72399 & 0.61%	0.15685 & 0.36%
WS-Ensemble	**0.282 & 0.32%**	**0.22657 & 0.31%**	**0.18303 & 0.34%**	**0.73267 & 0.61%**	**0.21234 & 0.4%**

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Average classification performance(bookmaker informedness and accuracy) across subjects in the order of n-class problems. The best performing approach is identified using a **bold font**.

AR modelling order value	2 Class Problem (IVb & IVa)	3 Class Problem (IVc & V)	4 Class Problem (IIIa)
2	0.041343 & 0.26%	0.19309 & 0.35%	0.088289 & 0.31%
4	0.050743 & 0.26%	0.30542 & 0.4%	0.11641 & 0.33%
6	0.10209 & 0.28%	0.36352 & 0.43%	0.12204 & 0.34%
8	0.14964 & 0.29%	0.39087 & 0.44%	0.12963 & 0.34%
10	0.16183 & 0.29%	0.39387 & 0.44%	0.13303 & 0.35%
16	0.19524 & 0.3%	0.38841 & 0.44%	0.15592 & 0.36%
30	0.2207 & 0.31%	0.36323 & 0.43%	0.19042 & 0.39%
All	0.21381 & 0.3%	0.35759 & 0.43%	0.18423 & 0.38%
AIC	0.2138 & 0.3%	0.35512 & 0.43%	0.18382 & 0.38%
BIC	0.199 & 0.3%	0.3613 & 0.43%	0.17591 & 0.38%
FPE	0.21384 & 0.3%	0.39058 & 0.44%	0.18587 & 0.38%
GA-Fusion	0.23161 & 0.31%	0.39894 & 0.45%	0.19592 & 0.39%
PSO-Fusion	0.23454 & 0.31%	0.43214 & 0.46%	0.19858 & 0.39%
S-Ensemble	0.17811 & 0.29%	0.41327 & 0.45%	0.15685 & 0.36%
WS-Ensemble	**0.25429 & 0.31%**	**0.45785 & 0.47%**	**0.21234 & 0.4%**

Table 6

Average classification performance (bookmaker informedness and accuracy) across subjects in the order of subject task performance strength. The best performing approach is identified using a **bold font**.

AR modeling order value	**Weak Subjects mean of bookmaker <0.2 AA, AV, IVC, K6B, L1B**	**Normal Subject mean of bookmaker ∈ [0.2, 0.3) AW, AY**	**Strong Subject mean of book-maker ≥0.3 IVB, AL, V1, V2, V3, K3B**
2	0.013319 & 0.26%	0.16057 & 0.29%	0.22923 & 0.37%
4	0.038444 & 0.27%	0.14214 & 0.29%	0.34703 & 0.42%
6	0.046299 & 0.28%	0.12407 & 0.28%	0.42973 & 0.46%
8	0.052346 & 0.28%	0.17364 & 0.29%	0.46469 & 0.47%
10	0.068795 & 0.28%	0.1705 & 0.29%	0.46879 & 0.47%
16	0.071936 & 0.29%	0.16971 & 0.29%	0.48407 & 0.48%
30	0.090505 & 0.3%	0.21071 & 0.3%	0.4625 & 0.47%
All	0.085475 & 0.29%	0.20571 & 0.3%	0.45622 & 0.47%
AIC	0.082858 & 0.29%	0.20571 & 0.3%	0.4578 & 0.47%
BIC	0.079915 & 0.29%	0.207 & 0.3%	0.45559 & 0.47%
FPE	0.093548 & 0.3%	0.21129 & 0.3%	0.47655 & 0.48%
GA-Fusion	0.10828 & 0.3%	0.22871 & 0.31%	0.49871 & 0.49%
PSO-Fusion	0.11164 & 0.31%	**0.2305 & 0.31%**	**0.52249 & 0.5%**
S-Ensemble	0.070363 & 0.29%	0.19579 & 0.3%	0.49333 & 0.48%
WS-Ensemble	**0.16902 & 0.33%**	**0.23471 & 0.31%**	0.51973 & 0.49%

Table 7

The chosen AR modeling orders with evolutionary based fusion techniques on datasets IVb and IVa with 2-Class problems.

Approach & Subject	AR modeling order values						
	2	4	6	8	10	16	30
GA-IVB	✓		✓		✓	✓	✓
GA-AA	✓		✓	✓			
GA-AL	✓			✓	✓	✓	
GA-AV	✓			✓	✓	✓	
GA-AW	✓			✓	✓		✓
GA-AY		✓	✓	✓			✓
PSO-IVB		✓	✓	✓	✓	✓	✓
PSO-AA	✓	✓	✓	✓	✓		
PSO-AL			✓	✓	✓		✓
PSO-AV		✓	✓	✓	✓	✓	
PSO-AW	✓	✓	✓		✓		✓
PSO-AY	✓		✓		✓	✓	✓
GA & 2 weak subjects (AA and AV)	2	0	1	2	1	1	0
GA & 2 normal subjects (AW and AY)	1	1	1	2	1	0	2
GA & 2 strong subjects (IVB and AL)	2	0	1	1	2	2	1
PSO & 2 weak subjects (AA and AV)	1	2	2	2	2	1	0
PSO & 2 normal subjects (AW and AY)	2	1	2	0	2	1	2
PSO & 2 strong subjects (IVB and AL)	0	1	2	2	2	1	2
GA Overall	5	1	3	5	4	3	3
PSO Overall	3	4	6	4	6	3	4

Table 8

The chosen AR modeling orders with evolutionary based fusion techniques on datasets IVc and V with 3-Class problems.

Approach & Subject	AR modeling order values						
	2	4	6	8	10	16	30
GA-IV	✓				✓	✓	
GA-V1			✓	✓	✓	✓	
GA-V2				✓	✓		
GA-V3			✓	✓		✓	
PSO-IVC	✓	✓	✓	✓	✓	✓	
PSO-V1			✓	✓	✓	✓	
PSO-V2		✓		✓	✓		
PSO-V3	✓	✓	✓	✓			
GA & 1 weak subject (IVC)	1	0	0	0	1	1	0
GA & 3 strong subjects (V1, V2 and V3)	0	0	2	3	2	2	0
PSO & 1 weak subject (IVC)	1	1	1	1	1	1	0
PSO & 3 strong subjects (V1, V2 and V3)	1	2	2	3	2	1	0
GA Overall	1	0	2	3	3	3	0
PSO Overall	2	3	3	4	3	2	0

Table 9

The chosen AR modeling orders with evolutionary based fusion techniques on dataset IIIa with 4-Class problems.

Approach & Subject	AR modeling order values						
	2	**4**	**6**	**8**	**10**	**16**	**30**
GA-K3B	✓		✓	✓		✓	✓
GA-K6B				✓	✓	✓	✓
GA-L1B	✓	✓	✓		✓		
PSO-K3B	✓	✓	✓	✓		✓	✓
PSO-K6B	✓	✓	✓	✓	✓		
PSO-L1B	✓	✓	✓	✓	✓		✓
GA & 2 weak subjects (K6B and L1B)	1	1	1	1	2	1	1
GA & 1 strong subject (K3B)	1	0	1	1	0	1	1
PSO & 2 weak subjects (K6B and L1B)	2	2	2	2	2	0	1
PSO & 1 strong subject (K3B)	2	1	1	1	0	1	1
GA Overall	2	1	2	2	2	2	2
PSO Overall	3	3	3	3	2	1	2

Table 10

Overview of the best performing algorithms in the study.

Category	Best performing	2^{nd} best performing	3^{rd} best performing
n-class problem			
2 Class Problem	WS-Ensemble	PSO	GA
3 Class Problem	WS-Ensemble	PSO	GA
4 Class Problem	WS-Ensemble	PSO	GA
Datasets			
IIIa	WS-Ensemble	PSO	GA
IVa	WS-Ensemble	GA	PSO
IVb	WS-Ensemble	PSO	AR-order 30
IVc	WS-Ensemble	PSO	FPE
V	WS-Ensemble	PSO	S-Ensemble
Subjects' Classification Strengths			
Week mean bookmaker < 0.2	WS-Ensemble	PSO	GA
Normal mean bookmaker ∈ [0.2, 0.3)	WS-Ensemble	PSO	GA
Strong mean bookmaker ≥ 0.3	PSO	WS-Ensemble	GA