Asynchronous Brain Computer Interfaces Using Echo State Networks

Eyal Zakkay
Department of Electrical and Computer Engineering
Ben-Gurion University
Beer-Sheva, Israel
eyalzk@gmail.com

Amjad Abu-Rmileh
Department of Cognitive and Brain Sciences
Ben-Gurion University
Beer-Sheva, Israel

Amir B. Geva
Department of Electrical and Computer Engineering
Ben-Gurion University
Beer-Sheva, Israel
geva@bgu.ac.il

Oren Shriki
Department of Cognitive and Brain Sciences
Ben-Gurion University
Beer-Sheva, Israel
shrikio@bgu.ac.il

Abstract— **Decoding brain activity for brain computer interfaces (BCI) is a challenging machine learning task. Many BCIs are designed in a synchronous setting, which allows to output a command from the brain to the computer at pre-determined, specific time points. The task becomes substantially more demanding when designing asynchronous BCIs, which allow control over the interface at any given time. Most BCIs use static classifiers, in which the readout is driven only by the current input. However, due to the complex nature of ongoing brain activity, dynamic classifiers, which capture the temporal dynamics of the input signal, are likely to be more suitable for the task. To examine this notion, we designed a classification framework based on Echo State Networks (ESNs), which have demonstrated strong performance in similar tasks in other domains. We evaluated the performance of this approach on pre-recorded EEG data of an asynchronous motor imagery (MI) BCI task, and compared our approach to a conventional method. ESN-based architectures exhibited superior performance in comparison. Our findings suggest that ESNs may prove useful in other BCI paradigms in virtue of their ability to reduce the detrimental effect of non-stationarity in brain signals.**

Keywords— Echo State Network, Brain-Computer Interface, Electroencephalography, Reservoir Computing, Motor-Imagery.

I. INTRODUCTION

A brain computer interface (BCI) allows a user to control technological applications using brain activity alone [1], as measured by a dedicated device (usually an electroencephalogram - EEG). As of today, many different methods were tested in order to design robust, practical BCIs. However, decoding brain activity reliably into computer commands is still a difficult machine-learning task. In addition, many practitioners in the BCI community agree that current approaches for BCI design are insufficient and do not address important questions regarding practical design [2]–[4].

One major question concerns whether the BCI operates at predefined times or continuously. Many BCIs work in a cued fashion, meaning that the user can establish control only at specific time points; these are called synchronous BCIs. Other types of BCIs are able to capture intent in arbitrary time steps, yet provide feedback at predefined time intervals. A more desirable type of BCI is an asynchronous one, in which the interface is active continuously, allowing the user to control it at any given time, and to stay idle when there is no specific intent.

This research was supported in part by the Helmsley Charitable Trust through the Agricultural, Biological and Cognitive Robotics Initiative and by the Marcus Endowment Fund, both at Ben-Gurion University of the Negev.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

Another important question is the type of classifier being used. Most BCIs use static classification methods [5]; these are classifiers in which the classification readout is driven solely by the current input (e.g. linear discriminant analysis, support vector machines, feedforward neural networks etc.). In some cases, representation of past inputs is included implicitly in the selected feature space, but the classifiers themselves cannot capture dynamic behavior. Brain activity signals have a dynamic, non-stationary nature, and conceptually represent mental states such as expectation, planning and execution as well as transitions between states. Therefore, the user's intent is likely to be dependent also on the recent past, and consequently, static classification methods might ignore valuable information. We propose that dynamic classification methods, which possess memory, and are able to capture the temporal dynamics of brain activity, can be useful for achieving a better decoding performance.

To address these questions, we considered a recurrent neural network (RNN) architecture called an Echo State Network (ESN) [6], which has the appealing ability of modeling complex dynamical systems at a relatively low computational cost.

We designed and implemented an asynchronous BCI, based on ESNs. Specifically, we focused on motor-imagery (MI) BCIs, in which the interface uses sensory-motor rhythms (SMRs) which are produced by imagining limb movements to establish control [7], [8].

Classic training protocols, which consist of visual cues such as pointing arrows or 2-D bars have been shown to be suboptimal due to lack of user motivation [2]. To achieve increased user engagement and consequently reduce noise and artifacts, we designed a game-like 3D environment [9]. Evaluation on test-data was performed as a simulation, where pre-recorded data were used as in a real-time scenario.

We tested our hypothesis to see if ESNs are able to successfully decode brain activity related to motor tasks, specifically discriminating between imagination of right-hand movement, left-hand movement and idle state (the times when the user does not imagine the movement of any limb). To assess the performance of the ESN BCI, we compared it to a common BCI design approach - classification using linear discriminant analysis (LDA) with band-power (BP) features [10].

Evaluation of asynchronous BCIs is not trivial. We first note that a pure asynchronous BCI, based solely on the free will of the user, cannot be evaluated objectively. Thus, the setting should simulate natural conditions but must include some instructions to the user at specific times. Classic evaluation methods, such as error rate measures or an MSE based measure, have considerable downsides. Instead, we employed evaluation methods based on sample-by-sample and event-by-event analyses using ROC (Receiver Operating Characteristic) measures (see Methods section, and [11] for details).

We found that ESNs obtain superior performance over the baseline method in the asynchronous MI task, while sticking to a practical experimental paradigm in terms of training time, electrode selection and training protocol. To further optimize performance, we also utilized filter-bank features, which differ slightly from BP features. Filter-bank features carry similar information compared to BP features, but are less smoothed over time and are therefore more suitable for dynamic classification.

Several previous studies employed ESNs in the context of BCI. In [12] ESNs were used to classify 2s segments of 4 mental states (for example counting backwards or thinking of a song). In [13], [14] and [15] MI data were classified using ESNs or long-short-term-memory (LSTM) RNNs. However, these works differ from ours by implementing synchronous BCIs, applying methods that involve demanding data and hardware requirements or reporting a narrower set of performance measures.

II. Methods

A. Signal Acquisition and Experimental Paradigm

Nine right handed, healthy subjects with no previous experience in BCI (6 males, 3 females) participated in the experiment. Each subject sat in a chair with both arms resting on the armrests. The experiment consisted of 7 recording sessions, each session comprised 30-40 trials of motor imagery. Each trial began with an idle state period in which the subject was instructed to 'do nothing', during the remainder of the trial the subject was instructed to imagine either a right-hand movement or a left-hand one. The instructions were presented in the form of a computer game in a 3D environment, where the appearance of spaceships on opposite sides of the screen indicated the requisite type of imagery (Fig. 1). The purpose of using this setup was to increase subject engagement and avoid loss of concentration. In sessions 2-7, trials varied in length with idle state periods ranging between 2.5-8 seconds and imagery periods ranging between 1.5-5 seconds. The variability of trial structure was meant to emulate a more realistic (and hence practical) setting for BCI inference. The first session comprised identical trials of 4s of idle state and 6s of imagery. The rationale for this was that our method of comparison does not utilize varying trial structure. Because the first session is dedicated to training and classifier design, we recorded this session in a way that will allow a fair comparison.

EEG signals were acquired using a g.USBamp system (Guger Technologies, Austria). Data were sampled at 256 Hz. An 8th order Butterworth bandpass filter of 0.1-60 Hz for artifact removal and a 48-52 Hz notch filter, were applied.

The data of interest are the C3 and C4 channels on the left and right hemispheres, respectively. However, to improve SNR and reduce artifacts, we implemented a Laplacian spatial filter using the 4 electrodes surrounding each of the two designated channels [1]. Thus, the total number of recording electrodes was 10.

B. Feature Extraction

Two feature extraction methods were employed, band-power features and filter-bank features. Filter-bank features are more appropriate in conjunction with RNNs. However, band-power features are more suitable when using the baseline method. Therefore, we evaluated performance using both options.

Band-power: band-power (BP) features are typically used in motor imagery BCI design. It is a measure for the amount of signal energy in a specified frequency band during a specified time window. BP is calculated by band-pass filtering the data in a desired frequency band, then squaring the values of all samples and finally averaging over a selected time window.

Band-Power is especially suitable for setups that consist of a small number of electrodes. It is also usually sufficient to use 1-2 bands for each channel, which is a relatively small number of features. However, BP feature suffer the disadvantage of being user-dependent, meaning different users could potentially have different reactive frequency bands. Although between-subject variability is not too large, even small differences could result in a significant drop in classification performance. Band-power features were extracted from the spatially filtered channels C3 and C4 using a 0.5s moving window. EEG spectrum was examined in order to select the most reactive frequency bands for each subject, based on comparison of imagery and idle-state spectrums using the calibration data. All selected frequency bands are within the range of 6-28 Hz, which is typical for motor imagery BCI.

Filter-Bank: These features are simply filtered versions of the input signal. Each spatially filtered EEG channel is band-pass filtered in different frequency bands (same bands as in the band-power method) using a 12th order Butterworth band-pass filter.

Following feature extraction, all features were down-sampled to 64 Hz. Using the raw EEG signals as features was also attempted. However, it yielded inferior results.

C. Echo State Networks

Echo state networks (ESNs) [6], [16] are a method of supervised learning based on a recurrent neural network (RNN). In an ESN, the connections between the network nodes are generated randomly and remain fixed throughout the learning process. Learning occurs when a linear classifier is trained to map the entire network state (also called a reservoir) into a desired output. This is accomplished by careful selection of a set of hyper-parameters which influence the reservoir's initialization. A good initialization will cause the recurrent network to act as an expansion of the input signal, where each unit gives a unique, rich, representation of the input signal and its recent past. It is likely that inputs which were not linearly separable in the original input space, will be separable in the ESN space. Furthermore, the use of linear classification results

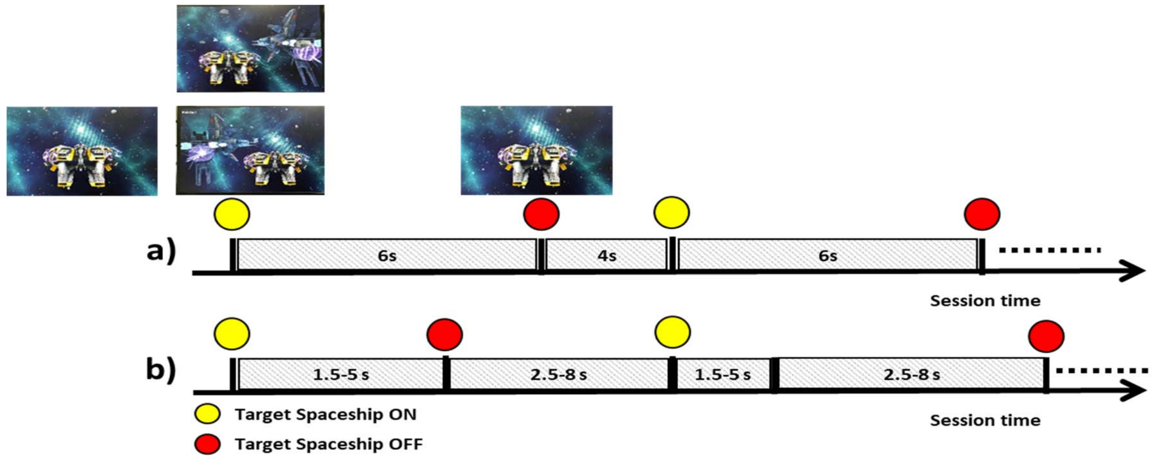

Figure 1: Trial structure. (a) During 1st session (training data). (b) During sessions 2-7 trails have varying lengths of idle\imagery periods.

in low computational demands, as opposed to common RNN architectures.

In an ESN, a reservoir is generated with random connections. The update equation of the units in the reservoir is:

$$\tilde{\boldsymbol{x}}(n+1) = \tanh\left(\mathrm{W}^{in}\left[1;\boldsymbol{u}(n)\right] + \mathrm{W}\boldsymbol{x}(n)\right) \quad (1)$$

$$\boldsymbol{x}(n+1) = (1-\alpha)\boldsymbol{x}(n) + \alpha\tilde{\boldsymbol{x}}(n+1) \quad (2)$$

Where $\boldsymbol{x}(n)$ is a vector of reservoir network activations at time step n, $\boldsymbol{u}(n)$ is the input, W^{in} and W are the input and recurrent weight matrices respectively. $\alpha \in (0,1]$ is the leaking rate. The output of the ESN, $y(n)$, is given by:

$$\boldsymbol{y}(n) = W^{out}\left[1;\boldsymbol{u}(n);\boldsymbol{x}(n)\right] \quad (3)$$

where W^{out} is the output weight matrix, and is in fact the only set of parameters being learned (most commonly using ridge regression [16]).

When designing an ESN, it is crucial to determine the global hyperparameters which govern the distribution of the generated networks and the reservoir's internal dynamics. The most impactful hyperparameters being: (1) The spectral radius of the recurrent weight matrix, (2) the scaling of the input, (3) the leaking rate - α and (4) the size of the network.

Following [11], we separated the task of the 3-class classification (right hand, left hand, idle) to two sub-tasks: classification of right-hand imagery vs. left-hand\idle, and classification of left-hand imagery vs. right-hand\idle (a multiclass approach of using a single classifier was also attempted, but yielded inferior results). For each sub-task an ensemble of 50 ESNs was generated and trained. Using ensembles with ESNs is recommended in the literature [17], as it yields a more stable classifier, which is less sensitive to outliers or poor reservoir initializations. In our case, using ensembles resulted in better classification, and the improvement plateaued at 50. The added computational complexity is linear in the number of reservoirs.

Inputs were centered and normalized to unity standard deviation. Reservoirs were initialized sparsely using a uniform distribution. Training was performed using weighted ridge-regression, in which each sample was weighted in order to avoid a bias to a specific class. The weights were determined according to the estimation of the a-priori distribution of the classes. Hyperparameters were optimized using a grid-search, for each individual subject, on a held-out set. In each search, 288 parameter combinations were evaluated, selected uniformly from the following ranges: Spectral radius – [0.75, 1.2], input scaling – [0.4, 1.5], leaking rate – [0.1, 0.9], reservoir size - [500, 1000]. For inference, the outputs of the 50-reservoir ensemble were averaged. The output activation function was the Softmax function, squeezing the outputs to the range [0,1]. For classification, a threshold was set, where an output that exceeded the threshold was classified as detection.

D. Baseline Method for comparison

Linear discriminant analysis (LDA) classifiers were designed for each subject in the following manner: BP features were extracted as explained above and subsequently log transformed. An LDA classifier was then trained to discriminate between right and left motor imagery. This resulted in a synchronous classifier that differentiates between right and left-hand imagery. The performance (successful classification rate, trial-by-trial) of this synchronous classifier was evaluated in order to assess the linear separability of the subject's data, before moving forward to a more complex, asynchronous, three-class problem. In order to use this classifier in an asynchronous environment, we used a thresholding mechanism on the prediction signal - each sample was classified using the designed LDA classifier, resulting in a 1-D prediction signal. The values of this signal represented the distance of the predicted sample from the separating LDA hyperplane. Values that exceeded a certain positive threshold were classified as 'right imagery', values that exceeded a certain negative threshold were classified as 'left imagery', and values between the two thresholds were classified as 'idle state'. This expansion of a synchronous LDA classifier to an asynchronous one follows the same method used in [11]. For reproducibility of the results, both feature extraction and LDA classifier training were performed using the g.BSanalyze software package by g.tec (Guger Technologies, Austria).

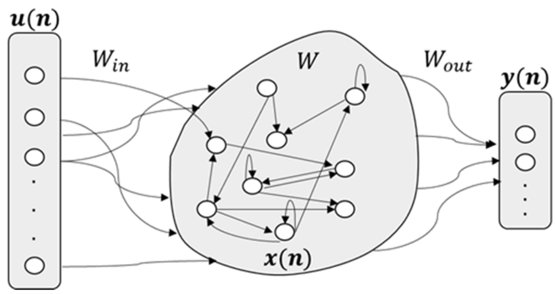

Figure 2: An Echo State Network. Input propagates through the cyclic connections of the reservoir, which is then mapped into the output units. W_{out} is the only trainable part of the ESN.

E. Evaluation method

Two types of analyses were used; (a) A multi-session analysis, in which classifiers were trained using the first recording session and evaluated using sessions 2-7, and (b) a single-session analysis using only the first session. Specifically, we used the first 50% of this session for training and the other 50% for evaluation. This analysis was meant to isolate the effect of non-stationarity, which is common in motor-imagery recordings without feedback over multiple recording sessions [9].

Following [11], we evaluated performance using both sample-by-sample analysis and event-by-event analysis.

1) Sample-by-sample analysis

For each of the two binary classifiers, each sample of the signals was classified. An ROC curve was generated as a measure of true positive rate (TPR) vs. false alarm rate (FAR). The measure of performance was the area under the curve (AUC) of the ROC curve. The sample-by-sample measure reflects the performance of the classifiers and generates smooth and informative ROC curves. However, when using a BCI, measuring performance on all samples is sub-optimal. For example, in a short motor imagery segment, a classifier that detects the user's intent even once during the imagery segment will be considered a good classifier (with respect to its practicality), even if not all samples of the segment were classified as imagery. Sample-by-sample measures will not reflect this, therefore we also used an analysis that relates to discrete events.

2) Event-by-event analysis

For each binary classifier, we defined periods of 'events' and 'non-events'. For example, when classifying between right-hand imagery to idle\left-hand imagery – right-hand imagery periods were considered as 'events' whereas other periods were considered as 'non-events'. Our goal was to measure correct detection of events (true positives-TP) vs. wrong detections during non-events (false alarms-FA). To this end, an event with at least one detection was classified as a TP, and a non-event with at least one detection was classified as a FA. Since non-events in our paradigm comprise a significant portion of the recording session (around 80%), we treated each non-event as a series of shorter non-events, with length corresponding to the average segment length of events in each recording session. Overlooking the imbalance in interval lengths would have resulted in an unfair comparison between detection capabilities for events and non-events, as longer intervals are more susceptible to be labeled as an event. We note that our approach slightly differs from the approach taken in [11], where non-events are not segmented (for this reason we omitted the 'TF-difference' measure, which is an alternative way to deal with long resting periods in the data).

Our approach allowed us to perform a valid ROC analysis, since the number of events and non-events was fixed, rather than determined by the number of detections. Again, we measured performance by the AUC. In this analysis, the ROC represents correct classifications of each 'event', which is a more suitable measure for a practical BCI. However, the ROC was generated under much more strict conditions, resulting in performance levels that seem 'poorer'. It is important to remember that this analysis evaluated the exact same classifiers as in the sample-by-sample analysis, where only the evaluation criterion was changed. For further motivation regarding using sample-by-sample and event-by-event analysis, see [11].

To reduce the potential effect of outliers on the results, the detector indicated the presence of an event only after the classification signal was over the detection threshold for at least 1s (for illustration see Fig. 6). This is similar to the 'dwell-time' concept introduced in [11].

F. User Reaction Time Consideration

When recording, the user was instructed to start or stop the imagination of hands at specific times. These timing signals were later used to construct the teacher signal for our supervised learning methods. However, there is an unknown delay between the time in which the user was instructed to imagine and the time the user actually did so. This is due to human reaction time and concentration level fluctuations during the task [18]. In order to compensate for the false label signal, we ignore the first 0.5s of each trial, and of each idle state period – both in training and in evaluation.

III. RESULTS

Experiments were performed with 9 subjects (6 males). For each subject, data were evaluated in two different ways. First, the classifiers were trained using the 1st recording session (40 trials) and results were evaluated on sessions 2-7. Acknowledging that long sessions of non-feedback EEG recordings are tedious and cause loss of concentration, we omitted recording sessions that produced chance level performance in all compared methods (AUC of 0.5±0.02), and ones in which the users claimed they were not concentrating or did not follow instructions. Two subjects (103,109) were completely excluded from the analysis due to poor SNR (signal to noise ratio) in the 1st session which precluded the design of any type of classifier.

A. Sample-by-Sample Analysis

Table 1 and Fig. 3 depict multi-session analysis results for all subjects in terms of the AUC of the ROC curve (AUC=1 corresponds to ideal performance). The first column of Table 1 shows the performance (correct classification rate) of a synchronous LDA classifier, demonstrating the linear separability of the left and right-hand classes in the given data-set. The remaining results are mean AUC values over all valid sessions. As evident, the ESN outperforms the LDA classifier. Furthermore, an ESN with a filter bank is typically better than

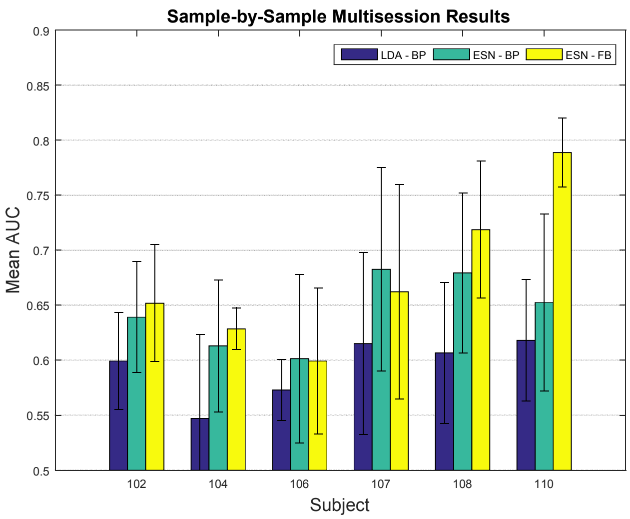

Figure 3: Sample-by-sample multisession analysis. Each group of bars indicates the AUC (mean±SD) value of a user for each compared method.

an ESN with band-power features. The mean AUC difference between the ESN-FB classifier and the LDA classifier is 0.08. For comparison, Table 3 presents AUC values for the single session analysis. The superiority of the ESN classifiers for a range of false-positive rates is demonstrated in Fig. 5 (upper row), which depicts examples of ROC curves for the 3 compared methods (dashed line represents chance level, AUC values are indicated in the figure).

B. *Event-by-Event Analysis*

In the event-based analysis, there was a relatively small number of events per-session (as opposed to the sample-based analysis which included many thousands of samples). Furthermore, the manner in which detection was determined reduced sensitivity to the thresholds used. As a result, the corresponding ROC curves were far less smooth compared to traditional ROC curves. Fig. 5 (bottom row) depicts example ROC curves of the three methods in the same manner as in the sample-by-sample analysis. Table 2 and Fig. 4 present the mean AUC values of the event-based, multisession analysis and Table 4 shows the AUC values of the single session analysis. As could be expected, the ESN classifiers are superior by the event-by-event metric as well. Fig. 6 presents the prediction signal in comparison with the true label signal, event detections are marked in red over the prediction signal. The figure also depicts an example of the dwell mechanism as explained in the previous section.

IV. Discussion

In this work, we suggest a new framework for asynchronous motor imagery BCI design using ESN classifiers. To examine if ESNs are a suitable tool for the decoding of brain activity, we tested this method on a classic asynchronous MI task, classifying a continuous EEG signal as belonging to one of three mental states (left and right-hand imagery, and the idle state). ESN classifiers have shown promising results and were superior to the baseline classifiers across all subjects under multisession analysis, and in the vast majority of cases using single session analysis.

It is interesting to note that when evaluating the performance over multiple sessions, the advantages of using ESNs are highlighted. ESNs are sensitive to the dynamics of the input signals, which possibly allows the inherent feature expansion to be more robust to changes in signal statistics. Since a practical BCI needs to function reliably over long periods of time without frequent calibrations, this is a very desirable property.

The range of obtained AUC values does not imply a very successful classification in standard terms. However, for a perfectly functioning BCI we do not need a perfect classifier in terms of AUC. For instance, in an imagination period we do not need all samples to be classified correctly, but only enough to declare a detection. In addition, since the true label signals rely on the instructions given to the subject, and not on the true imagination itself (which is unknown) - classification measures can be poorer due to noisy labels, both in training and in evaluation. Erroneous labeling can result from lack of concentration, slow reaction time to instructions and user confusion. Furthermore, current results are based on single-hour experiments of non-experienced BCI users. With further training most users are expected to exhibit much better results [9].

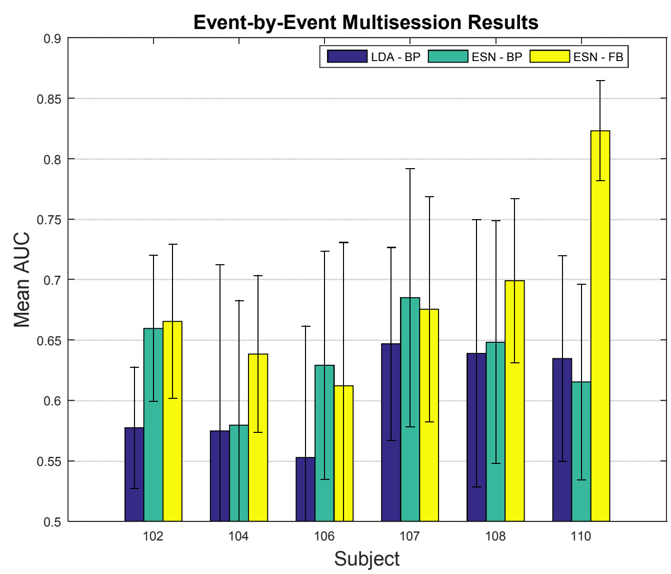

Figure 4: Event-by-event multisession analysis. Each group of bars indicates the AUC (mean±SD) value of a user for each compared method.

We have shown that ESN classifiers outperform linear classifiers with band-power features. However, there are some more advanced methods for feature extraction, such as common spatial patterns (CSP) [19]. While CSP methods (which are also usually followed by an LDA classifier [19]) might result in more informative features, they usually require recording from a large number of electrodes to be efficient [20], [21]. Since we aim for a practical BCI, we used only 2 spatially filtered electrodes above the left and right hemispheres (C3 and C4 respectively), using a Laplacian spatial filter (practically recording from 10 electrodes, and potentially needing only 2-4 if the SNR is good enough). It is also important to mention that the focus of this research was to evaluate the proposed classifier (i.e. ESNs) for BCI design, whereas parameters such as feature extraction methods were outside the scope for this work. For these reasons, we decided to compare to a common, well researched [10], [21]–[23], relatively robust method such as the LDA with BP features.

Now that we have learned that ESNs are suitable for BCI design, in future research we can further optimize ESN architecture using different feature spaces, pre-processing techniques etc. Another future plan is incorporating methods to cope with inter-subject variability [24], [25].

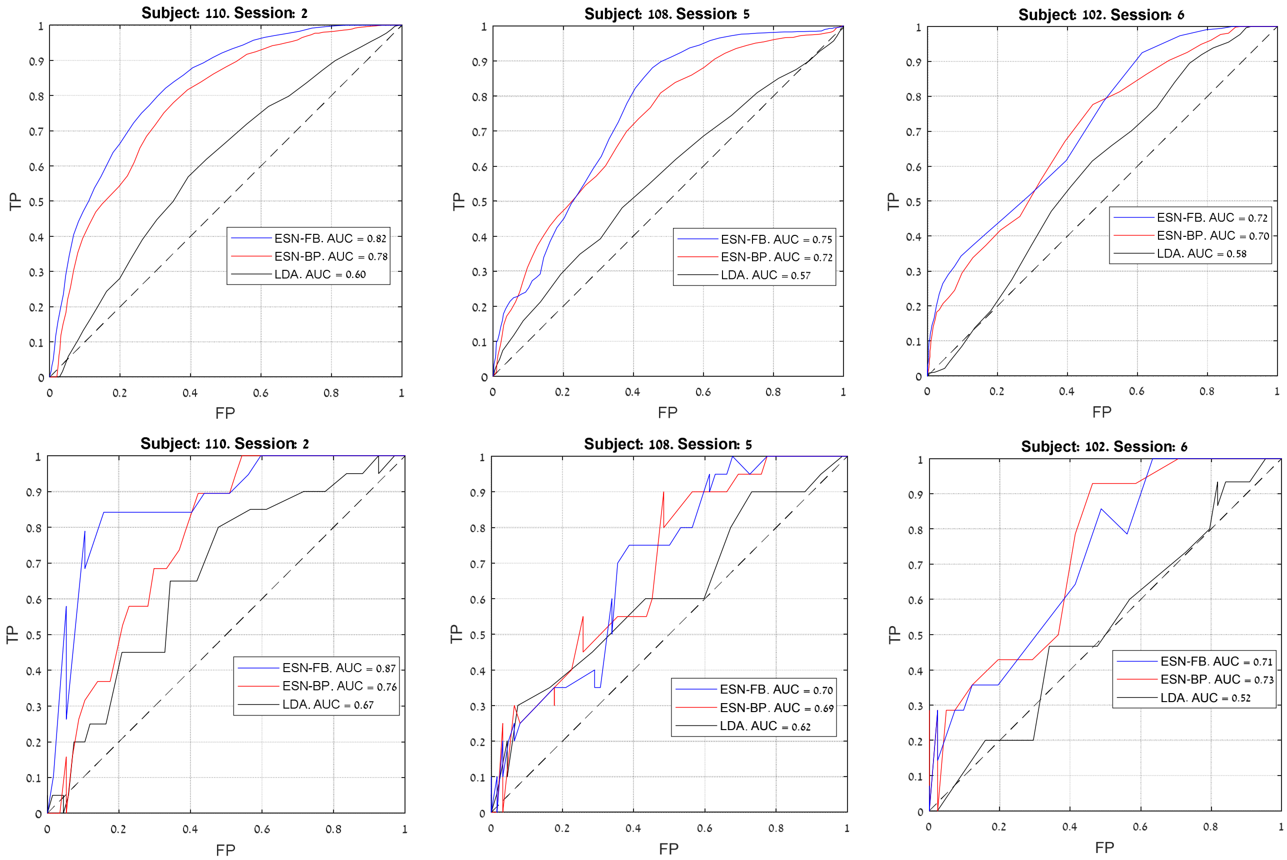

Figure 5: Example ROC curves. Top row: sample-by-sample analysis. Bottom row: event-by-event analysis. Each example shows results of one binary classifier (right vs left\idle or left vs right\idle). The corresponding classifier of each example exhibits similar behavior.

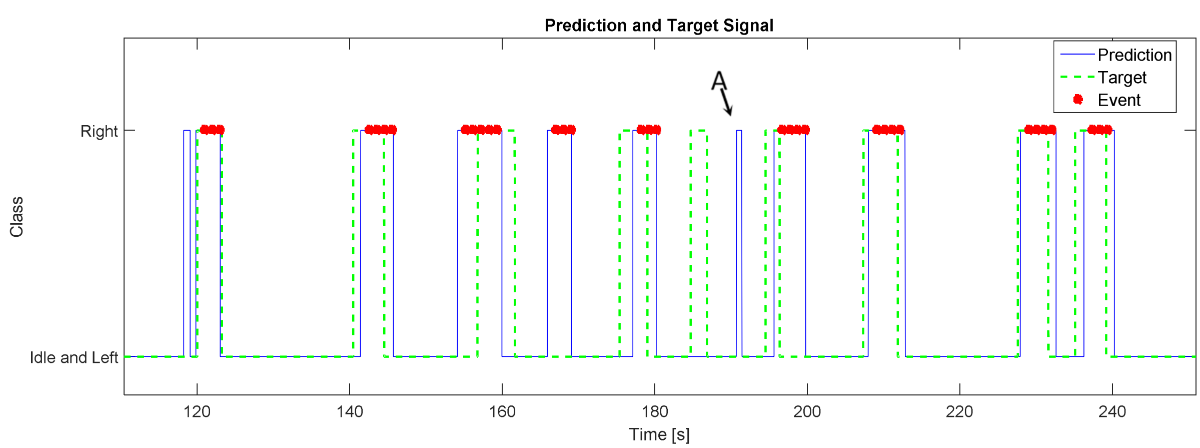

Figure 6: In this example, the green dashed line is the expected output of the test set, the blue line is the ESN output. Red markers are event detections. Point A illustrates how waiting a 1s period before declaring a detection can prevent false detection.

V. CONCLUSIONS

We have demonstrated that echo state network classifiers are better at asynchronous MI BCI tasks, compared to a common approach. Claiming superiority in the BCI field is fraught with difficulties, due to the lack of standard benchmark tests as well as variability among participants, recording conditions, sensors, user instructions and more, across different studies. Due to these reasons and more, there is no 'state of the art' result we can compare to ours. Nevertheless, we believe that this work justifies further exploration of ESN classifiers for BCI. ESNs gain access to valuable information in the data that is otherwise ignored, at a small computational cost. All this in an engaging, asynchronous and practical scenario.

*Table 1: AUC (mean±SD) for sample-by-sample performance evaluation, multisession analysis. The synchronous classification rate of the LDA classifier is presented to demonstrate the linear separability of the data in a binary classification task (left\right). Best classifier in each row is marked in bold face. *Subject 101 is not presented because his performance in sessions 2-7 was at chance level using all 3 methods.*

Subject	LDA classification rate	LDA	ESN Band-power	ESN Filter Bank
110	0.85	0.62±0.05	0.65±0.08	**0.79±0.03**
108	0.83	0.61±0.06	0.68±0.07	**0.72±0.06**
107	0.88	0.62±0.08	**0.68±0.09**	0.66±0.09
102	0.83	0.60±0.04	0.64±0.05	**0.65±0.05**
104	0.75	0.55±0.07	0.61±0.06	**0.63±0.01**
106	0.73	0.57±0.02	**0.62±0.04**	0.61±0.05

*Table 2: AUC (mean±SD) for event-by-event performance evaluation, multisession analysis. Best classifier in each row is marked in bold face. *Subject 101 is not presented because sessions 2-7 were disqualified.*

Subject	LDA classification rate	LDA	ESN Band-power	ESN Filter Bank
110	0.85	0.63±0.08	0.62±0.08	**0.82±0.04**
108	0.83	0.64±0.11	0.65±0.10	**0.70±0.06**
107	0.88	0.65±0.08	**0.69±0.10**	0.68±0.09
102	0.83	0.58±0.05	0.66±0.06	**0.67±0.06**
104	0.75	0.57±0.13	0.58±0.10	**0.64±0.06**
106	0.73	0.55±0.10	**0.63±0.09**	0.61±0.11

Table 3: AUC for sample-by-sample, single session analysis (50% of the 1st session was used for training and 50% for evaluation).

Subject	LDA	ESN Band-power	ESN Filter Bank
110	0.65	0.75	**0.76**
108	0.68	**0.79**	0.75
107	0.71	**0.79**	0.77
102	0.61	0.59	**0.62**
104	0.59	**0.68**	0.65
106	0.53	**0.62**	0.56
101	0.66	**0.73**	0.71

Table 4: AUC for event-by-event, single session analysis (50% of the 1st session was used for training and 50% for evaluation).

Subject	LDA	ESN Band-power	ESN Filter Bank
110	0.74	0.69	**0.83**
108	**0.79**	0.77	0.70
107	0.73	**0.79**	0.76
102	0.59	0.65	**0.68**
104	0.65	0.66	**0.70**
106	0.60	**0.69**	0.62
101	0.76	0.72	**0.77**

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, Jun. 2002, doi: 10.1016/S1388-2457(02)00057-3.

[2] F. Lotte, F. Larrue, and C. Mühl, "Flaws in current human training protocols for spontaneous Brain-Computer Interfaces: lessons learned from instructional design," *Front. Hum. Neurosci.*, vol. 7, 2013.

[3] C. Jeunet, E. Jahanpour, and F. Lotte, “Why standard brain-computer interface (BCI) training protocols should be changed: an experimental study,” *J. Neural Eng.*, vol. 13, no. 3, p. 036024, 2016.
[4] B. Z. Allison and C. Neuper, “Could anyone use a BCI?,” in *Brain-computer interfaces*, Springer, 2010, pp. 35–54.
[5] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, “A review of classification algorithms for EEG-based brain–computer interfaces,” *J. Neural Eng.*, vol. 4, no. 2, p. R1, 2007, doi: 10.1088/1741-2560/4/2/R01.
[6] H. Jaeger, “The ‘echo state’ approach to analysing and training recurrent neural networks-with an erratum note,” *Bonn Ger. Ger. Natl. Res. Cent. Inf. Technol. GMD Tech. Rep.*, vol. 148, no. 34, p. 13, 2001.
[7] G. Pfurtscheller and F. L. Da Silva, “Event-related EEG/MEG synchronization and desynchronization: basic principles,” *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, 1999.
[8] H. Yuan and B. He, “Brain–computer interfaces using sensorimotor rhythms: current state and future perspectives,” *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1425–1435, 2014.
[9] A. Abu-Rmileh, E. Zakkay, L. Shmuelof, and O. Shriki, “Co-adaptive Training Improves Efficacy of a Multi-Day EEG-Based Motor Imagery BCI Training,” *Front. Hum. Neurosci.*, vol. 13, 2019, doi: 10.3389/fnhum.2019.00362.
[10] R. Scherer, G. R. Muller, C. Neuper, B. Graimann, and G. Pfurtscheller, “An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate,” *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 979–984, 2004.
[11] G. Townsend, B. Graimann, and G. Pfurtscheller, “Continuous EEG classification during motor imagery-simulation of an asynchronous BCI,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 12, no. 2, pp. 258–265, 2004.
[12] E. M. Forney, C. W. Anderson, W. J. Gavin, P. L. Davies, M. C. Roll, and B. K. Taylor, “Echo State Networks for Modeling and Classification of EEG Signals in Mental-Task Brain-Computer Interfaces,” *Tech. Rep. Colo. State Univ. Fort Collins Colo. USA*, 2015.
[13] P.-J. Kindermans, P. Buteneers, D. Verstraeten, and B. Schrauwen, “An uncued brain-computer interface using reservoir computing,” in *Workshop: Machine Learning for Assistive Technologies (NIPS-2010)*, 2010.
[14] M. Li, M. Zhang, X. Luo, and J. Yang, “Combined long short-term memory based network employing wavelet coefficients for MI-EEG recognition,” in *Mechatronics and Automation (ICMA), 2016 IEEE International Conference on*, 2016, pp. 1971–1976.
[15] Z. H. Khan, N. Hussain, and M. I. Tiwana, “Classification of EEG signals for wrist and grip movements using echo state network,” 2017.
[16] M. Lukoševičius, “A practical guide to applying echo state networks,” in *Neural networks: Tricks of the trade*, Springer, 2012, pp. 659–686.
[17] H. Jaeger, M. Lukoševičius, D. Popovici, and U. Siewert, “Optimization and applications of echo state networks with leaky-integrator neurons,” *Neural Netw.*, vol. 20, no. 3, pp. 335–352, 2007.
[18] G. Pfurtscheller, C. Neuper, H. Ramoser, and J. Müller-Gerking, “Visually guided motor imagery activates sensorimotor areas in humans,” *Neurosci. Lett.*, vol. 269, no. 3, pp. 153–156, 1999.
[19] F. Lotte, “Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain–computer interfaces,” *Proc. IEEE*, vol. 103, no. 6, pp. 871–890, 2015.
[20] J. Farquhar, J. Hill, T. N. Lal, and B. Schölkopf, “Regularised CSP for sensor selection in BCI,” 2006.
[21] G. Pfurtscheller *et al.*, “Current trends in Graz brain-computer interface (BCI) research,” *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 216–219, 2000.
[22] C. Vidaurre and B. Blankertz, “Towards a cure for BCI illiteracy,” *Brain Topogr.*, vol. 23, no. 2, pp. 194–198, 2010.
[23] R. Boostani and M. H. Moradi, “A new approach in the BCI research based on fractal dimension as feature and Adaboost as classifier,” *J. Neural Eng.*, vol. 1, no. 4, p. 212, 2004.
[24] B. Blankertz *et al.*, “Neurophysiological predictor of SMR-based BCI performance,” *NeuroImage*, vol. 51, no. 4, pp. 1303–1309, Jul. 2010, doi: 10.1016/j.neuroimage.2010.03.022.
[25] O. Alkoby, A. Abu-Rmileh, O. Shriki, and D. Todder, “Can We Predict Who Will Respond to Neurofeedback? A Review of the Inefficacy Problem and Existing Predictors for Successful EEG Neurofeedback Learning,” *Neuroscience*, vol. 378, pp. 155–164, 15 2018, doi: 10.1016/j.neuroscience.2016.12.050.