entropy

Article

Machine Learning Classification of Event-Related Brain Potentials during a Visual Go/NoGo Task

Anna Bryniarska [1], José A. Ramos [2] and Mercedes Fernández [3,*]

[1] Department of Computer Science, Opole University of Technology, 45-758 Opole, Poland; a.bryniarska@po.edu.pl
[2] College of Computing and Engineering, Nova Southeastern University, Fort Lauderdale, FL 33314, USA; jr1284@nova.edu
[3] Department of Psychology and Neuroscience, Nova Southeastern University, Fort Lauderdale, FL 33314, USA
* Correspondence: mf934@nova.edu

Abstract: Machine learning (ML) methods are increasingly being applied to analyze biological signals. For example, ML methods have been successfully applied to the human electroencephalogram (EEG) to classify neural signals as pathological or non-pathological and to predict working memory performance in healthy and psychiatric patients. ML approaches can quickly process large volumes of data to reveal patterns that may be missed by humans. This study investigated the accuracy of ML methods at classifying the brain's electrical activity to cognitive events, i.e., event-related brain potentials (ERPs). ERPs are extracted from the ongoing EEG and represent electrical potentials in response to specific events. ERPs were evoked during a visual Go/NoGo task. The Go/NoGo task requires a button press on Go trials and response withholding on NoGo trials. NoGo trials elicit neural activity associated with inhibitory control processes. We compared the accuracy of six ML algorithms at classifying the ERPs associated with each trial type. The raw electrical signals were fed to all ML algorithms to build predictive models. The same raw data were then truncated in length and fitted to multiple dynamic state space models of order n_x using a continuous-time subspace-based system identification algorithm. The $4n_x$ numerator and denominator parameters of the transfer function of the state space model were then used as substitutes for the data. Dimensionality reduction simplifies classification, reduces noise, and may ultimately improve the predictive power of ML models. Our findings revealed that all ML methods correctly classified the electrical signal associated with each trial type with a high degree of accuracy, and accuracy remained high after parameterization was applied. We discuss the models and the usefulness of the parameterization.

Keywords: machine learning; binary classification; EEG signal; state space modeling; biological signal; event-related brain potentials

Citation: Bryniarska, A.; Ramos, J.A.; Fernández, M. Machine Learning Classification of Event-Related Brain Potentials during a Visual Go/NoGo Task. *Entropy* **2024**, *26*, 220. https://doi.org/10.3390/e26030220

Academic Editor: Karl Friston

Received: 4 December 2023
Revised: 16 February 2024
Accepted: 28 February 2024
Published: 29 February 2024

1. Introduction

There is an increasing interest in the application of machine learning (ML) methods to analyze biological signals, including signals from the human body [1,2] and electrical signals from the brain, i.e., the electroencephalogram (EEG) and event-related brain potentials (ERPs) [3–5]. ML methods have been successfully applied to EEG recordings to automate the detection of seizures and improve diagnostic accuracy [6] and to classify emotional states [7,8]. ML methods have also been successfully applied to ERPs to improve the diagnostic accuracy and prognosis of attention-deficit hyperactivity disorder (ADHD) [9]. ERPs are extracted from the ongoing EEG and represent the sum of electrical potentials that are time-locked to a cognitive event and are generated by populations of neurons that fire within milliseconds after the event. The temporal resolution of ERPs is unparalleled by other brain imaging procedures and they are considered the gold standard for observing neural activity over time. In [10], the authors present a comprehensive review of the major techniques used for EEG signal processing and feature extraction as they relate to decoding and classification of EEG signals. Other techniques that have been used to capitalize on the information in ERPs are averaging of the temporal waveforms (i.e., averaged ERPs), time–frequency representation, and phase dynamics. Indeed, in a recent study [11], the three techniques were applied in combination with a neural network-based ML model to better exploit the neural dynamic behavior in the ERP elicited during a visual oddball task. The oddball paradigm requires a response to a target stimulus that is presented infrequently (e.g., on 20% of trials) within a series of standard stimuli. The infrequent target stimulus elicits a P3 ERP component. The P3 ERP is a positive-going wave that occurs 250–500 ms after stimulus onset, with maximum amplitude over parietal electrode sites, and reflects updating of the memory trace [12]. Results showed that the three-feature model classified the averaged ERP signal to the rare target and the frequent standard stimulus with an accuracy level of 86.9%.

ML methods have also been applied to classify neural activity elicited during a Go/NoGo task. The Go/NoGo task is widely used in cognitive neuroscience to assess frontal-lobe inhibitory control processes associated with response inhibition and, more generally, with executive function (EF) [13]. Executive function refers to a set of abilities that work together to regulate thought and action. The Go/NoGo task requires a button press on Go trials and response withholding on NoGo trials. The underlying neural marker associated with frontal-lobe inhibitory control processes is the N2 ERP. The N2 ERP is a negative-going wave in the 200–350 ms post-stimulus time window, with maximum amplitude over frontal-central electrode sites. NoGo trials, which require greater inhibitory control, elicit greater N2 ERP amplitude than Go trials, which require less inhibition [14–16]. Indeed, studies of healthy adults reveal that the amplitude of the N2 ERP is larger in participants who accurately withhold a response on NoGo trials relative to those who do not withhold a response [17]. One study [18] applied ML methods to identify neural processes of response selection and response inhibition engaged during the Go and NoGo conditions. Results revealed an accuracy rate of 92%, estimated by 5-fold cross-validation. Another study investigated the influence of self-reported personality traits of impulsivity and compulsivity on performance based on the ERP. Regression tree analyses did not reveal a relationship between self-reported measures and behavior or the Go/NoGo ERPs [19].

While ML methods have made meaningful contributions to EEG classification, shortcomings related to EEG data make classification difficult for ML algorithms [20]. For example, ML algorithms have to deal with signals that are rich in noise. Additionally, most EEG studies involve a small number of study participants, usually between 10 and 20 [21], permitting only small data sets for the learning phase of the process. There are two situations that can degrade the performance of ML algorithms: (1) not having a sufficient number of study participants and (2) having a very large number of data points. The latter may lead to "the curse of dimensionality." For these reasons, it is sometimes difficult to make accurate classifications of the neural signal, and several techniques must be tested to determine which ones yield the best results. There are several techniques used to reduce the dimensionality of EEG data: Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA), and Independent Component Analysis (ICA) [22]. Discrete Wavelet Transform (DWT) is also often used for this purpose [23].

We propose a new approach, the use of a state space model as a dimensionality reduction step, followed by a PCA step to extract the minimum number of significant principal components (i.e., features) in an optimization approach, coupled with ML. To the knowledge of the authors, such an approach has not been considered in the literature related to ERP signals. Notably, state space analysis has been used [24] for estimating multivariate autoregressive (MVAR) models of cortical connectivity from noisy scalp recorded EEG signals for the purpose of modeling the spatial covariance structure of the noise in the EEG signal. That study differs from what we are proposing in that our goal was to substitute the data with parameters and test the accuracy of ML algorithms at classifying the ERP signals. The rest of the paper is organized as follows: in Section 2, we discuss the study methodology and EEG data collection process. In Section 3, we discuss the data reduction process and ML methodology. In Section 4, we present the state space methodology for EEG data, while in Section 5 we present the state space analysis. In Section 6, we introduce the system identification algorithm for impulse response data. In Section 7, we present our results. Finally, in Section 8 we draw conclusions and make recommendations for future work.

2. Study Methodology and EEG Recording

The ERP data reported in this article were collected as part of a larger study investigating neural and behavioral differences in a linguistically diverse student population [25]. Participants were recruited from the main campus of NSU and were invited to participate if they were right-handed, had normal hearing, normal or corrected-to-normal vision, intact color vision, met the language requirement, and did not report neurological or psychiatric conditions that affect cognition.

2.1. Participant Information

A total of 268 participants were tested. Data from seven participants were excluded from the analyses because these participants did not meet study criteria, and six participants did not yield usable data. Thus, ERP data from 255 study participants were used in the analyses. Participants were between 18 and 30 years of age (mean = 19.5, SD = 2.73) and the male to female ratio was 55/206.

2.2. Visual Go/NoGo Task

The stimuli for the Go/NoGo task were red and green circles, presented on a computer monitor against a black background, and subtending a visual angle of 2.9°. Each stimulus was presented for 80 ms. Each trial consisted of two stimuli separated by 1200 ms. For each trial, when a target circle was followed by another target circle (Go trials), participants pressed a response button to the second circle. When the target circle was followed by a nontarget circle (NoGo trials), participants withheld their response. Go and NoGo trials occurred with equal frequency (36% each trial type). Trials that started with a nontarget stimulus were not analyzed. The Go/NoGo task consisted of 200 trials, divided into four blocks of 50 trials, with an intertrial interval (ITI) of 1800 ms. To increase task difficulty, an auditory signal (300 ms at 1 kHz, 60 dB SPL tone burst) was sounded if the participant did not respond within 600 ms after the second target stimulus was presented. This time pressure was introduced after the first 100 trials. Participants focused on a fixation point, responded as quickly as possible to the second target in the pair on Go trials, and withheld responding on NoGo trials. The task began after participants read the instructions on the computer monitor and practiced the task. After the second block of trials, participants were trained on the task with the added time pressure (tone burst), after which the remaining two blocks of trials were presented. Participants were instructed to respond quickly to avoid the tone burst.

2.3. EEG Recording and Processing

The continuous EEG was recorded with a lycra cap fitted with 64 Ag/AgCl sintered electrodes (i.e., 62 scalp electrodes and 2 bipolar electrodes for vertical and horizontal eye movement recording) and amplified with a Neuvo amplifier (Compumedics U.S.A. Inc., Charlotte, NC, USA). The EEG was sampled at 500 Hz, which exceeds the Nyquist frequency [26]. Eye movement was recorded with electrodes placed above and below the left eye and on the outer canthus of each eye. Reference electrodes were placed on the right and left mastoid. Electrode impedance was maintained at <10 kΩ, and most were under <5 kΩ. After recording, the EEG data were processed offline with Curry 8 software (Compumedics U.S.A. Inc.). Offline, the EEG was re-referenced to the common average reference and filtered (high-pass filter set to 0.10 Hz, slope = 0.2; low-pass filter set to 30 Hz, slope = 6.0; 60 Hz notch filter, slope = 1.5). Eyeblinks exceeding ±75 µV were corrected using the covariance method [27]. The covariance analysis is performed between the eye artifact channel and each EEG channel. Linear transmission coefficients, similar to beta weights, are computed. Based on the weights, a proportion of the voltage is subtracted from each data point.

Stimulus locked trials (−140 to 800 ms) were then extracted from the ongoing EEG and baseline (−140 to 0 ms) corrected. The noise statistic was applied to automatically reject contaminated trials. Noise was computed over the baseline period and trials that exceeded the average noise level were automatically rejected. Only trials with correct responses were averaged together by trial type and exported for analysis. Thus, each participant generated two averaged ERP waves, one Go and one NoGo.

3. Data Reduction and Machine Learning Methodology

Our goal was to employ different ML algorithms to show which ones achieve the highest classification accuracy of the ERP signal as either corresponding to a Go or a NoGo trial. To achieve this, we divided the data into two sets, both having 510 subjects (255 for the Go trials and 255 for the NoGo trials) and 62 electrodes. One set of data contained 471 data points per electrode, i.e., the entire ERP signal, whereas the other set contained only 250 data points per electrode, representing the most significant portion of the ERP signal (see Figure 1). Due to the fact that the recorded data has a 3-dimensional (3D) structure, we applied a data unfolding procedure described in [28] (see Figure 2).

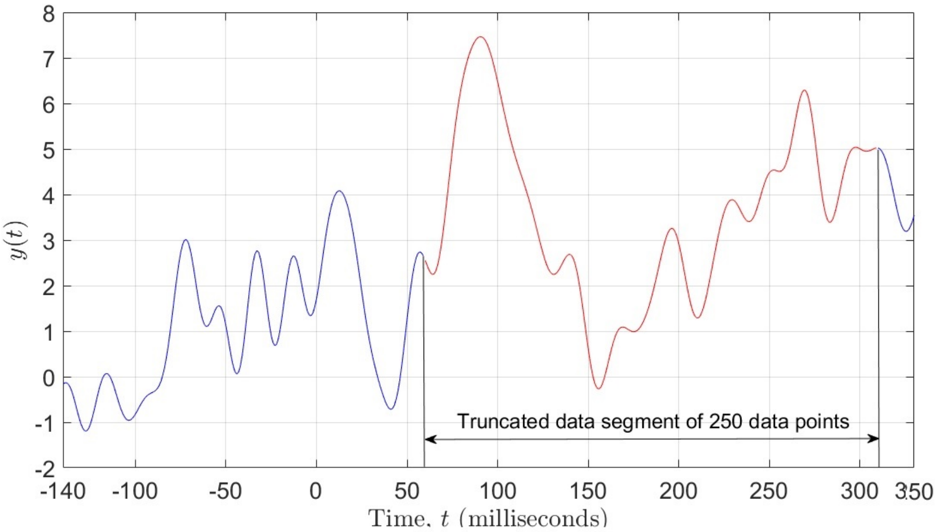

Figure 1. Figure showing 250 data points selected from the ERP signal.

Let X denote the data matrix

$$X = \left[\begin{array}{cccc|cccc|c|cccc} x_{11}^1 & x_{12}^1 & \cdots & x_{1T}^1 & x_{11}^2 & x_{12}^2 & \cdots & x_{1T}^2 & \cdots & x_{11}^N & x_{12}^N & \cdots & x_{1T}^N \\ x_{21}^1 & x_{22}^1 & \cdots & x_{2T}^1 & x_{21}^2 & x_{22}^2 & \cdots & x_{2T}^2 & \cdots & x_{21}^N & x_{22}^N & \cdots & x_{2T}^N \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{P1}^1 & x_{P2}^1 & \cdots & x_{PT}^1 & x_{P1}^2 & x_{P2}^2 & \cdots & x_{PT}^2 & \cdots & x_{P1}^N & x_{P2}^N & \cdots & x_{PT}^N \end{array}\right],$$

where $T = \{250, 471\}$ indicates the two different numbers of data points used in the study, $N = 62$ is the number of electrodes, and $P = 510$ is the number of subjects. For the data set containing 471 data points, X would have dimensions 510 × 29,202, which is a fairly large data set. On the other hand, with only 250 data points, X would have dimensions 510 × 15,500, which is a smaller data set, i.e., a 47% reduction. However, if we could fit dynamic models to the data set with 250 data points per electrode, then we could use the parameters of the models as a substitute for the data set. This could be a significant data reduction step, provided there is no loss of accuracy in modeling the data. One such type of dynamic model comes from the class of subspace-based state space system identification algorithms, collectively known as N4SID [29–32]. The idea would be to fit 62 state space models to the data containing 250 points per electrode, thus obtaining a set of 62 parameter triplets $\{A, B, C\}$, where $A \sim n_x \times n_x$, $B \sim n_x \times 1$, and $C \sim 1 \times n_x$, thus, totaling $n_x^2 + 2n_x$ parameters per electrode, where n_x is the system order. One could then convert the models to a transfer function form, which is a more parsimonious representation, resulting in $2n_x$ parameters per electrode, i.e., n_x numerator parameters and n_x denominator parameters. This could result in a data matrix of size $510 \times 124n_x$. Preliminary analyses carried out using data from the entire data set indicate that using an $n_x = 20$ results in models with great fidelity. That is, we would obtain a data matrix of size 510×2480, which is much smaller than $510 \times 29{,}202$, by a 91.5% reduction factor. However, the parameters of the transfer function model could result in being complex numbers, therefore, in the worst case scenario, one has to split the parameters into their real and imaginary parts, thus accounting for twice the number of parameters, i.e., 510×4960 or an 83% reduction. This approach alleviates the curse of dimensionality, which is quite common in machine learning. Comparison of the results would allow for the direct assessment of the effectiveness of dimensionality reduction to EEG analyses.

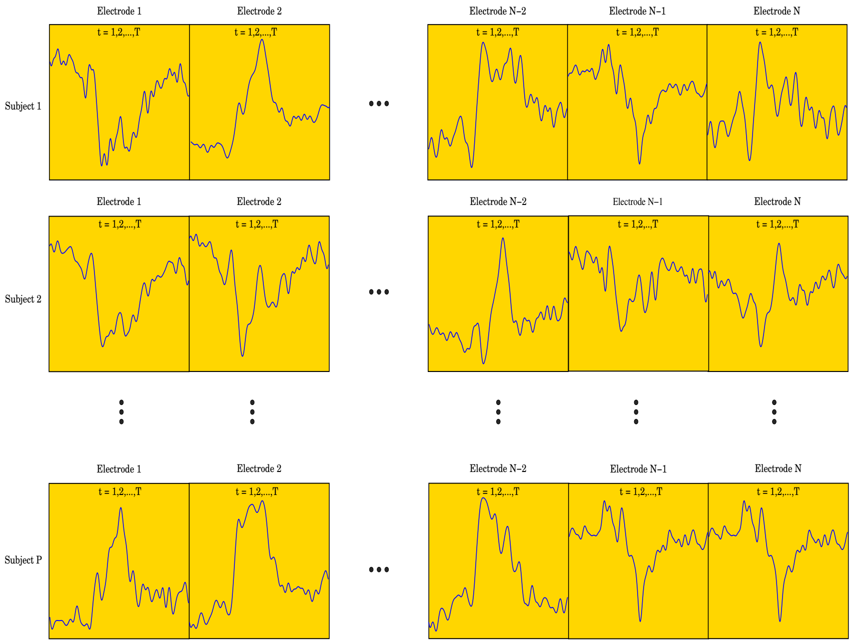

Figure 2. Figure showing the 3D structure of the data for the case when $T = 250$.

ML algorithms create a predictive model based on the provided data: classification labels, training data, and test data. This is called supervised learning. The available data are usually divided into training and test or validation data sets. The ML algorithms use the training data set to build a predictive model, which is then validated with the test data. Figure 3 shows the overall ML modeling process. One starts with the training data, along with a set of class labels, i.e., $\{0, 1\}$ for binary classification. This information is fed to the ML algorithm, which in turn uses a K-fold cross-validation procedure to obtain a predictive model. The test data, which are new to the model, are then used to predict its class labels. Such models can be employed for classification, much like the ones we use here for classifying the ERP signal into Go and NoGo trials (thus, a binary classification problem).

As described above in Section 2, each of the 255 participants generated an averaged Go and a NoGo ERP signal. Thus, 510 ERP signals were used in the analyses. Sampled at 500 Hz for 940 ms, (−140 to 800 ms) including a 140 ms pre-stimulus baseline (−140 to 0 ms), each ERP signal consisted of 471 data points. The signal was collected from 62 electrodes placed over the scalp of the study participants. Thus, a matrix containing 29,202 data points

(62 electrodes × 471 data points) was obtained for each participant (see Figure 4). The processed signal was subjected to state space modeling in order to establish parameters that could replace the entire signal and reduce the dimensionality of the input data to the ML classifier (see Figure 5). For each electrode, 40 parameters were calculated according to the state space modeling methodology described in Section 6. Since each parameter is a complex number containing real and imaginary parts, hence 62 electrodes × 40 parameters × 2 (real and imaginary) equals 4960 data points after parameterization. For each data sample from a participant, a reduction in dimensionality by 83% was obtained. These data were then used to perform a PCA to assess the number of significant principal components as a function of accuracy of the ML classifier. Six different ML algorithms were analyzed in terms of accuracy versus the number of significant principal components.

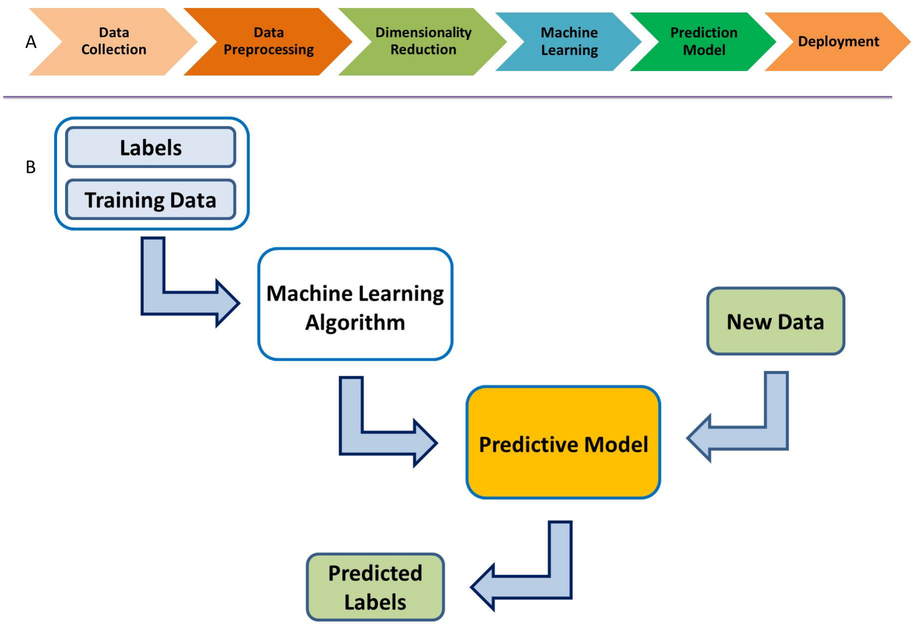

Figure 3. (**A**) Supervised machine learning process. (**B**) Predictive supervised machine learning.

The ML algorithms used in the research are k-nearest neighbors (KNN), Naive Bayes (NB), decision trees (DTs), linear discriminant analysis (LDA), support vector machines (SVM), and random forest (RF). KNN is a simple and powerful supervised machine learning algorithm that can be used for classification tasks. KNN is often used in cases where the data are nonlinear or do not fit well into traditional parametric models. The NB classifier is a probabilistic machine learning model based on Bayes' theorem with an assumption of independence among predictors. DTs are hierarchical structures used for classification tasks. They consist of decision nodes that split the data based on features, and leaf nodes, which represent the outcome. The algorithm selects the best feature to split the data at each node, aiming to maximize purity. Once constructed, the tree is used to predict outcomes for new data. Key features include interpretability and the ability to capture complex decision boundaries. LDA is a statistical model used for topic modeling. It assumes documents are composed of a mixture of topics, and each topic is characterized by a distribution of words. LDA aims to identify these topics in a collection of documents. The SVM method is a supervised learning method that analyzes given data and identifies patterns which are used for classification and regression analysis [33]. The SVM method is based on the concept of decision space, which is divided by building boundaries separating objects of different class affiliation. In binary classification there are two classes, and a boundary line is created to separate them. This method is widely used to analyze EEG signals of epileptic seizure activity [34], sleep recordings of patients [35], and in the recognition of emotional states [36]. RF builds multiple decision trees during training and outputs the mode for classification prediction of the individual trees. RF introduces randomness in the tree-building process by using a subset of features at each split and bootstrapping sample. This helps in reducing overfitting and improving generalization performance [37].

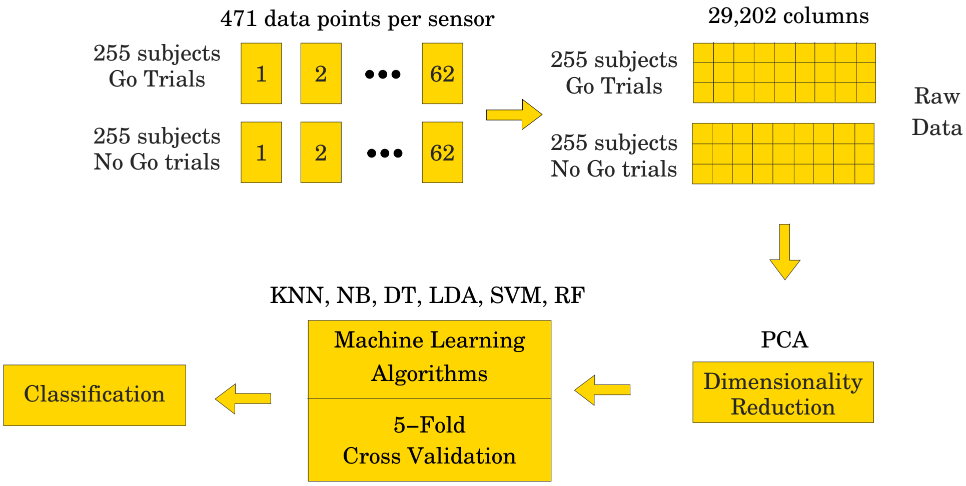

Figure 4. Figure showing the raw data unfolding process.

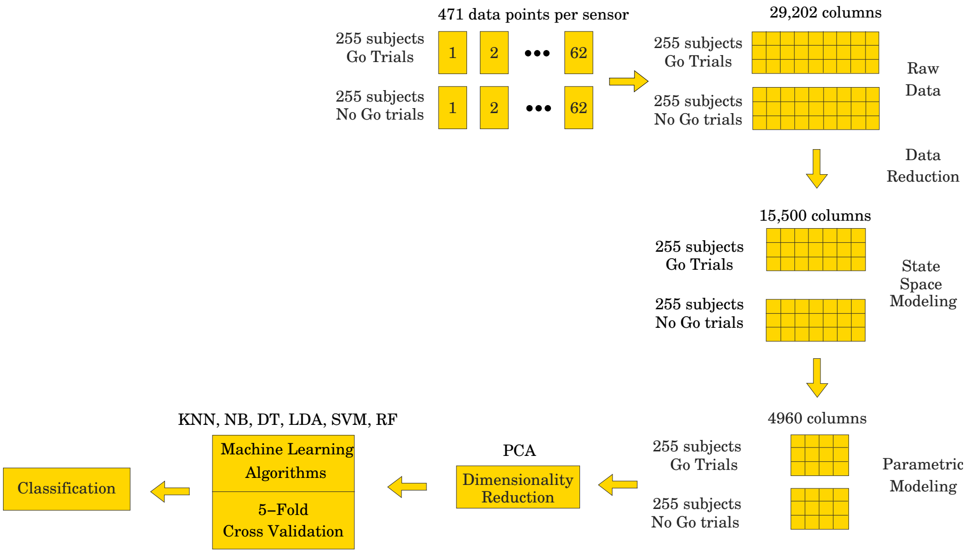

Figure 5. Figure showing the parametric data unfolding process.

PCA is a statistical technique used for dimensionality reduction and data visualization. PCA aims to transform the original data set into a new coordinate system where the variables (features) are uncorrelated, and the variance along each axis (principal components) is maximized. This transformation is achieved by identifying the principal components, which are linear combinations of the original variables [38]. Lastly, the 5-fold cross-validation method is used to determine the average classification results. The k-fold cross-validation process is shown in Figure 6, where $k = 5$. In each fold, different parts of the data set are taken as the test and training sets. This approach ensures that the outcomes remain unaffected by the selection of partitioning the data into training and test sets.

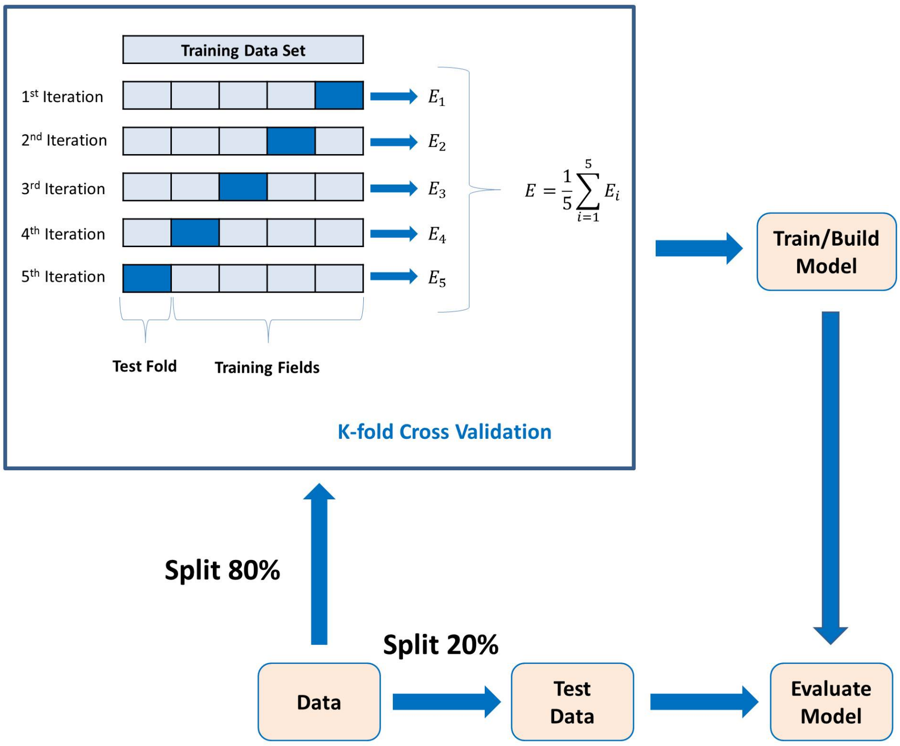

Figure 6. K-fold cross-validation process with model evaluation.

4. State Space Modeling of EEG Data

In the context of EEG measurements, an impulse response is a signal change that corresponds to a cerebral response to some stimuli. EEG data are therefore the result of an impulse response experiment. Thus, EEG data can be modeled as a continuous-time impulse response state space model of the form

$$\dot{x}_c(t) = A_c x_c(t) + B_c u(t) \tag{1}$$
$$y(t) = C_c x_c(t) + D_c u(t). \tag{2}$$

The matrices of parameters are given by

$$A_c = \left[\begin{array}{c|c|c|c} a_{11} & a_{12} & \cdots & a_{1,n_x} \\ \hline a_{21} & a_{22} & \cdots & a_{2,n_x} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline a_{n_x,1} & a_{n_x,1} & \cdots & a_{n_x,n_x} \end{array}\right] \tag{3}$$

$$B_c = \left[\begin{array}{c} b_1 \\ \hline b_2 \\ \hline \vdots \\ \hline b_{n_x} \end{array}\right] \tag{4}$$

$$C_c = \left[\begin{array}{c|c|c|c} c_1 & c_1 & \ldots & c_{n_x} \end{array}\right] \tag{5}$$

$$D_c = \left[\begin{array}{c} 0 \end{array}\right] \tag{6}$$

$$x_c(t) = \left[\begin{array}{c} x_1(t) \\ \hline x_2(t) \\ \hline \vdots \\ \hline x_{n_x}(t) \end{array}\right]. \tag{7}$$

Note that the feedback matrix $D = 0$. Then, in the transfer function form of (1)–(2), the order of the numerator polynomial is smaller than the order of the denominator polynomial. In traditional state space analysis, we have an n_x-order state space model with respective states, inputs, and outputs at time t, given by $x_c(t) \in \mathbb{R}^{n_x}$, $u(t) \in \mathbb{R}^{n_u}$, and $y(t) \in \mathbb{R}^{n_y}$, and $\{n_x, x_c(0), A_c, B_c, C_c, D_c\}$ are the unknown parameters of the system. Such a model is known as a multi-input, multi-output (MIMO) state space model. When the input and output dimensions are scalar values, the model is referred to as a single-input, single-output (SISO) state space model [39], which is the case of interest in this study.

The problem we address here is the following: Given a sequence of impulse response data $\{g(t)\}_{t=0}^{N-1}$, obtained from some experiment, determine the system order n_x, initial state vector $x(0)$, and parameters matrices $\{A_c, B_c, C_c, D_c\}$. We can only identify the parameters modulo an invertible similarity transformation matrix, $T \in \mathbb{R}^{n_x \times n_x}$. Therefore, the identified model is not unique. However, the input/output properties of the model are unique. That is, the Markov parameters

$$h(i) = \begin{cases} C_c A_c^{i-1} B_c, & i > 0 \\ D_c, & i = 0 \end{cases}, \tag{8}$$

the impulse response parameters

$$g(t) = \begin{cases} C_c e^{A_c t} B_c, & t \geq 0^+ \\ D_c, & t = 0^- \end{cases}, \tag{9}$$

and transfer function coefficients

$$H(s) = C_c(sI_{n_x} - A_c)^{-1} B_c \tag{10}$$

$$= \frac{\beta_{n_x} s^{n_x - 1} + \beta_{n_x - 1} s^{n_x - 3} + \cdots + \beta_2 s + \beta_1}{s^{n_x} + \alpha_{n_x} s^{n_x - 1} + \alpha_{n_x - 1} s^{n_x - 2} + \cdots + \alpha_2 s + \alpha_1} \tag{11}$$

are unique, where I_{n_x} is an $n_x \times n_x$ identity matrix and s is the Laplace variable. The parameters $\{\alpha_1, \alpha_2, \ldots, \alpha_{n_x}, \beta_1, \beta_2, \ldots, \beta_{n_x}\}$ are the parameters of an observable canonical state space model of the form

$$A_{oc} = \left[\begin{array}{c|c|c|c|c} 0 & 0 & \cdots & 0 & -\alpha_1 \\ \hline 1 & 0 & \cdots & 0 & -\alpha_2 \\ \hline 0 & 1 & \cdots & 0 & -\alpha_3 \\ \hline \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline 0 & 0 & \cdots & 0 & -\alpha_{n_x-1} \\ \hline 0 & 0 & \cdots & 1 & -\alpha_{n_x} \end{array}\right]$$

$$B_{oc} = \left[\begin{array}{c} \beta_1 \\ \hline \beta_2 \\ \hline \vdots \\ \hline \beta_{n_x} \end{array}\right]$$

$$C_{oc} = \left[\begin{array}{c|c|c|c} 0 & 0 & \ldots & 1 \end{array}\right]$$

$$D_{oc} = \left[\begin{array}{c} 0 \end{array}\right].$$

Therefore, the minimum number of parameters needed to represent the state space system (1)–(2) is $2n_x$, if the initial states are ignored. There is a similarity transformation matrix, $T \in \mathbb{R}^{n_x \times n_x}$, such that $A_{oc} = TA_cT^{-1}$, $B_{oc} = TB_c$, and $C_{oc} = C_cT^{-1}$. Note that $y(t) = g(t)$ when $u(t) = \delta(t)$, the Dirac delta function. To identify the continuous-time model, we use the impulse response coefficients and apply Kung's realization algorithm [29] to determine $\{n_x, x_c(0), A_c, B_c, C_c\}$ directly from the data. In Section 5, we will use this approach.

5. Identification of $\{n_x, A_c, B_c, C_c\}$ via the Impulse Response Coefficients

One can identify the continuous-time model (1)–(2) using the measured impulse response coefficients and Equation (9). This leads to a Hankel matrix decomposition of the form

$$G = \begin{bmatrix} g(0) & g(1) & g(2) & \cdots & g(j-1) \\ g(1) & g(2) & g(3) & \cdots & g(j) \\ g(2) & g(3) & g(4) & \cdots & g(j+1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g(i-1) & g(i) & g(i+1) & \cdots & g(N-1) \end{bmatrix}.$$

Note that the Hankel matrix is characterized by having constant antidiagonals. The matrix G needs to be factored into the product of the observability ($\mathcal{O}_c$) and controllability ($\mathcal{C}_c$) matrices, two rank n_x matrices. Such matrix decomposition is possible via the singular value decomposition (SVD) [29–32]. That is,

$$\begin{aligned} G &= \begin{bmatrix} C_cB_c & C_ce^{A_c\Delta T}B_c & C_ce^{2A_c\Delta T}B_c & \cdots & C_ce^{(j-1)A_c\Delta T}B_c \\ C_ce^{A_c\Delta T}B_c & C_ce^{2A_c\Delta T}B_c & C_ce^{3A_c\Delta T}B_c & \cdots & C_ce^{jA_c\Delta T}B_c \\ C_ce^{2A_c\Delta T}B_c & C_ce^{3A_c\Delta T}B_c & C_ce^{4A_c\Delta T}B_c & \cdots & C_ce^{(j+1)A_c\Delta T}B_c \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_ce^{(i-1)A_c\Delta T}B_c & C_ce^{iA_c\Delta T}B_c & C_ce^{(i+1)A_c\Delta T}B_c & \cdots & C_ce^{(N-1)A_c\Delta T}B_c \end{bmatrix} \\ &= \begin{bmatrix} C_c \\ C_ce^{A_c\Delta T} \\ C_ce^{2A_c\Delta T} \\ \vdots \\ C_ce^{(i-1)A_c\Delta T} \end{bmatrix} \cdot \begin{bmatrix} B_c & e^{A_c\Delta T}B_c & e^{2A_c\Delta T}B_c & \cdots & e^{(j-1)A_c\Delta T}B_c \end{bmatrix} \\ &= \begin{bmatrix} C_c \\ C_c(e^{A_c\Delta T}) \\ C_c(e^{A_c\Delta T})^2 \\ \vdots \\ C_c(e^{A_c\Delta T})^{i-1} \end{bmatrix} \cdot \begin{bmatrix} B_c & (e^{A_c\Delta T})B_c & (e^{A_c\Delta T})^2B_c & \cdots & (e^{A_c\Delta T})^{j-1}B_c \end{bmatrix} \\ &= \begin{bmatrix} U_1 \mid U_2 \end{bmatrix} \left[\begin{array}{c|c} S_1 & 0_{n_x\times(j-n_x)} \\ \hline 0_{(i-n_x)\times n_x} & 0_{(i-n_x)\times(j-n_x)} \end{array}\right] \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \\ &= U_1S_1V_1^T, \end{aligned}$$

where $U = \begin{bmatrix} U_1 \mid U_2 \end{bmatrix}$ and $V = \begin{bmatrix} V_1 \mid V_2 \end{bmatrix}$ are orthogonal matrices, and

$$S_1 = \begin{bmatrix} s_1 & & & \\ & s_2 & & \\ & & \ddots & \\ & & & s_{n_x} \end{bmatrix}$$

is a diagonal matrix of the n_x most significant singular values of the continuous-time system (1)–(2), thus n_x is the best estimate of the system order. From the above matrix decomposition we can compute the observability and controllability matrices, $\mathcal{O}_c$ and $\mathcal{C}_c$, respectively, from

$$\begin{aligned} \mathcal{O}_c &= U_1S^{\frac{1}{2}} \\ \mathcal{C}_c &= S^{\frac{1}{2}}V_1^T. \end{aligned}$$

Furthermore, we need to define two shifted observability matrices $\mathcal{O}_c^1$ and $\mathcal{O}_c^2$ as

$$\mathcal{O}_c^1 = \begin{bmatrix} C_c \\ C_c(e^{A_c\Delta T}) \\ C_c(e^{A_c\Delta T})^2 \\ \vdots \\ C_c(e^{A_c\Delta T})^{i-2}, \end{bmatrix}$$

$$\mathcal{O}_c^2 = \begin{bmatrix} C_c(e^{A_c\Delta T}) \\ C_c(e^{A_c\Delta T})^2 \\ \vdots \\ C_c(e^{A_c\Delta T})^{i-1} \end{bmatrix}.$$

Likewise, we define the matrix exponential $e^{A\Delta T}$ as

$$e^{A\Delta T} = \left(\mathcal{O}_c^1\right)^{\dagger} \mathcal{O}_c^2,$$

where ΔT is the sampling interval.

We can now identify the parameters $\{n_x, A_c, B_c, C_c\}$ from

$$\begin{aligned} A_c &= \frac{\log_e\left(\left(\mathcal{O}_c^1\right)^{\dagger}\mathcal{O}_c^2\right)}{\Delta T} \\ B_c &= \mathcal{C}_c(:, 1:n_u) \\ C_c &= \mathcal{O}_c(1:n_y, :) \\ n_x &= \mathrm{rank}\{G\}, \end{aligned}$$

where $\log_e(M)$ is the base e matrix logarithm of the matrix M [31].

6. System Identification of an EEG Signal

Here, we have taken EEG data from a single electrode and conducted a system identification exercise on the data. Figure 7 shows the singular values versus system order plot, showing a significant cut-off at around $n_x = 17$. Also evident is the noise floor of singular values for $\sigma_x > 22$. The fitting error was calculated as

$$f = \left(\frac{1}{N}\right) \sum_{t=0}^{N-1} (g(t) - g_{fitted}(t))^2,$$

where $g(t)$ is the observed impulse response (observed EEG data), $g_{fitted}(t)$ is the fitted impulse response (fitted EEG data), and $N = 250$ is the number of observations. The fitting error was $f = 3.3271 \times 10^{-7}$. Clearly, a state space model with $n_x = 20$ performed very well, as can be seen in Figure 8.

It is clear that the singular value plot cuts off between $17 < n_x < 21$. Several tests showed that not all electrodes had the same system order properties as the example above. Therefore, we set the system order to $n_x = 20$ for all the models. We selected electrode 19 as an example and fitted a state space model to it. The system order was between 17 and 21. We chose $n = 20$ and the mean squared error (MSE) was in the order of 10^{-7}. Not all electrodes showed a constant system order throughout. However, the average system order was about 20. For each electrode, we conducted an optimization of system order versus MSE. We decided to take $n = 20$ as an average system order and either truncate or zero pad the transfer function parameters accordingly. This was a result of the decaying behavior of the parameters in the transfer function as the system order increased. So, we started with a minimum system order of $n = 17$ and calculated the MSE. We then increased the system order to $n = 18$ and calculated the new MSE. If the new MSE improved, we kept increasing the system order by one, thus trying to bound the MSE to a minimum. In essence, we obtained the optimal system order for each electrode. Since we computed the transfer function parameters, we either truncated the parameters to $n = 20$ or zero padded them in cases where $n < 20$. The singular value plot versus system order is a common tool used in state space modeling for determining the system order [29,30].

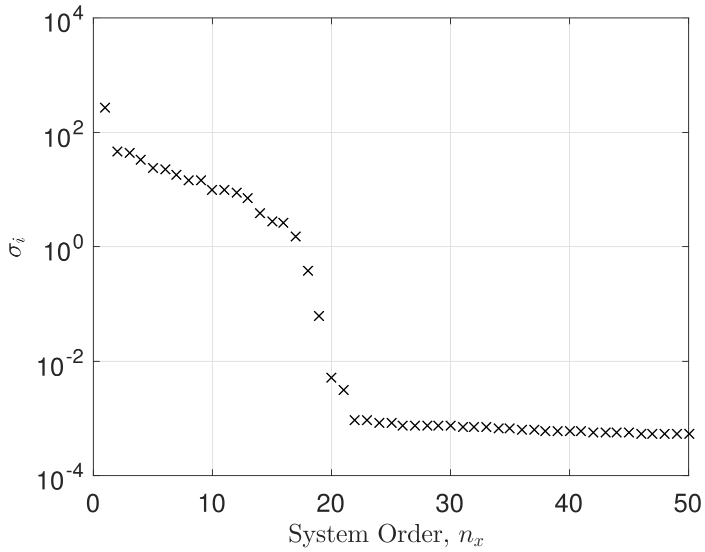

Figure 7. Singular value plot.

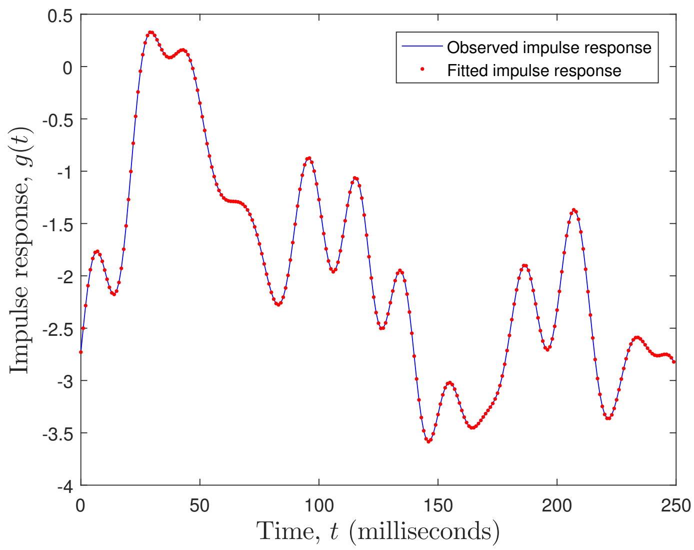

Figure 8. Observed versus fitted EEG plot.

7. Results

First, all six ML models (KNN, NB, DT, LDA, SVM, RF) were tested on the full data set before applying state space modeling for dimensionality reduction. Matrices as large as 510 × 29,202 data points were taken into account for each study participant. For each ML algorithm, accuracy results are presented as a function of the number of principal components required to achieve the given accuracy. The PCA method was used to calculate the score matrix, and a given subset of principal components were used in a 5-fold cross-validation analysis for each ML method. The results are presented in Table 1. The best results for each ML model are marked in blue font for ease of readability.

Table 1. Machine learning algorithm performance versus number of principal components for the entire ERP signal, i.e., X ∼ 510 × 29,202.

PC (n)	KNN	NB	DT	LDA	SVM	RF
1	0.9627	0.8902	1.0000	0.6098	0.9725	1.0000
2	0.6647	0.8549	1.0000	0.6863	0.6392	1.0000
3	0.5627	0.8510	1.0000	0.6608	0.6294	1.0000
4	0.6078	0.8588	1.0000	0.6725	0.6196	1.0000
5	0.6412	0.8529	1.0000	0.7098	0.6588	1.0000
6	0.6118	0.8314	1.0000	0.7157	0.6529	1.0000
7	0.5980	0.8176	1.0000	0.7157	0.6863	1.0000
8	0.6196	0.8647	1.0000	0.7784	0.7137	1.0000
9	0.6157	0.8431	1.0000	0.7843	0.7000	1.0000
10	0.6255	0.8667	1.0000	0.8059	0.7216	1.0000
11	0.6196	0.8667	1.0000	0.8157	0.7588	1.0000
12	0.6157	0.8706	1.0000	0.8157	0.7431	1.0000
12	0.6333	0.8941	1.0000	0.8510	0.7882	1.0000
14	0.6078	0.8529	1.0000	0.8490	0.7765	1.0000
15	0.6196	0.8745	1.0000	0.8588	0.7961	1.0000
16	0.6549	0.9039	1.0000	0.9020	0.8333	1.0000
17	0.6176	0.9000	1.0000	0.8980	0.8431	1.0000
18	0.6471	0.9137	1.0000	0.8961	0.8333	1.0000
19	0.6196	0.9000	1.0000	0.9078	0.8451	1.0000
20	0.6176	0.9235	1.0000	0.9392	0.8549	1.0000
21	0.6078	0.8980	1.0000	0.9392	0.8627	1.0000
22	0.6275	0.9039	1.0000	0.9392	0.8627	1.0000
23	0.6176	0.8941	1.0000	0.9392	0.8745	1.0000
24	0.6176	0.8941	1.0000	0.9373	0.8706	1.0000
25	0.6137	0.8882	1.0000	0.9392	0.8725	1.0000
26	0.6373	0.8745	1.0000	0.9431	0.8765	1.0000
27	0.6294	0.8706	1.0000	0.9373	0.8686	1.0000
28	0.6059	0.8863	1.0000	0.9412	0.8941	1.0000
29	0.6412	0.8569	1.0000	0.9431	0.8686	1.0000
30	0.6294	0.8922	1.0000	0.9392	0.8588	1.0000
31	0.6078	0.8745	1.0000	0.9431	0.8745	1.0000
32	0.6059	0.8627	1.0000	0.9412	0.8882	1.0000
33	0.6020	0.8725	1.0000	0.9353	0.8882	1.0000
34	0.5941	0.8647	1.0000	0.9353	0.8706	1.0000
35	0.6098	0.8627	1.0000	0.9275	0.8824	1.0000
36	0.6039	0.8490	1.0000	0.9333	0.8863	1.0000

Table 2 shows the best results for each of the ML algorithms from Table 1 with calculated metrics [40]:

$$ERR = \frac{FP+FN}{FP+FN+TP+TN} \tag{12}$$

$$ACC = \frac{TP+TN}{FP+FN+TP+TN} = 1-ERR \tag{13}$$

$$SPE = \frac{TN}{FP+TN} \tag{14}$$

$$SEN = \frac{TP}{FN+TP} \tag{15}$$

$$PRE = \frac{TP}{FP+TP} \tag{16}$$

$$F1 = 2\times\frac{PRE\times SEN}{PRE+SEN}, \tag{17}$$

where ERR denotes the error, ACC denotes accuracy, SPE denotes specificity, SEN denotes sensitivity, PRE denotes precision, and a measure of model performance is the F1 statistic. Accuracy is a widely used metric for evaluating classification models, representing the proportion of correctly classified samples among the total samples assessed. Precision, on the other hand, calculates the ratio of accurately predicted positive cases to the sum of all positively predicted cases, where TP represents the true positives and FP represents the false positives, thus precision reveals the accuracy of positive predictions. Sensitivity, also known as recall or true positive rate, determines the ratio of TP to the sum of false negatives (FNs) and TPs, thus it highlights the model's capability in correctly identifying actual positive cases. Specificity can be described as the model's ability to predict a true negative (TN) of each category available. In the literature, it is also known as the true negative rate. The F1 metric combines both precision and recall to provide a single score that balances the trade-off between them. Thus, the F1 statistic uses the average measures of sensitivity and precision to calculate the F-score statistic. It is calculated as the harmonic mean of precision and recall. It is particularly useful when there is an imbalance between the classes in the data set. The metrics {ACC, ERR, PRE, SEN, SPE, F1} were used as measures of fidelity toward judging the performances of the different models. Note that all metrics are scalars in the range [0,1], with higher values indicating a better model performance, except for the error metric, ERR, in which a lower value indicates a better model performance since it is $1 - ACC$. See Figure 9 for the confusion matrix as a function of TP, TN, FP, and FN.

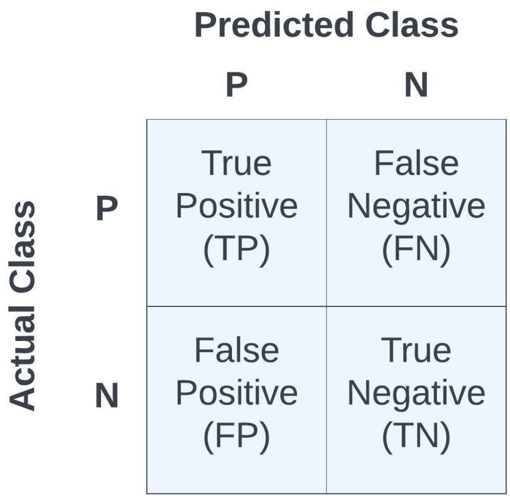

Figure 9. Summary of confusion matrix terminologyfor binary classification.

Table 2. Summary of machine learning algorithm performances for the entire ERP signal, i.e., $X \sim 510 \times 29{,}202$.

Metrics	KNN	NB	DT	LDA	SVM	RF
ACC	0.9627	0.9235	1.0000	0.9431	0.9725	1.0000
ERR	0.0373	0.0765	0.0000	0.0608	0.0275	0.0000
PRE	0.9629	0.9241	1.0000	0.9392	0.9726	1.0000
SEN	0.9627	0.9235	1.0000	0.9392	0.9725	1.0000
SPE	0.9627	0.9235	1.0000	0.9392	0.9725	1.0000
F1	0.9627	0.9235	1.0000	0.9392	0.9725	1.0000

After using the state space modeling procedure on the raw data matrix X of size $510 \times 15{,}500$, thus resulting in a reduced data matrix of size 510×4960, it was then fed to the same six ML algorithms {KNN, NB, DT, LDA, SVM, RF} versus PCA and using 5-fold cross-validation. Once again, accuracy results are presented as a function of the number of principal components required to achieve the given accuracy. The results are presented in Table 3.

Table 3. Machine learning algorithm performance versus number of principal components for the parametric data case, i.e., 510×4960 data points.

PC(n)	KNN	NB	DT	LDA	SVM	RF
1	0.9706	0.9333	1.0000	0.4902	0.9765	1.0000
2	0.7647	0.7843	1.0000	0.5784	0.7902	1.0000
3	0.5824	0.7549	1.0000	0.5941	0.6059	1.0000
4	0.5961	0.7431	1.0000	0.6020	0.6255	1.0000
5	0.5824	0.7549	1.0000	0.6098	0.6373	1.0000
6	0.5706	0.7529	1.0000	0.6235	0.6431	1.0000
7	0.6020	0.7451	1.0000	0.6255	0.6392	1.0000
8	0.5745	0.7471	1.0000	0.6176	0.6373	1.0000
9	0.5882	0.7412	1.0000	0.6196	0.6294	1.0000
10	0.5804	0.7078	1.0000	0.6020	0.6373	1.0000
11	0.5686	0.7196	1.0000	0.6353	0.6588	1.0000
12	0.5784	0.7059	1.0000	0.6333	0.6373	1.0000
12	0.5608	0.7000	1.0000	0.6294	0.6490	1.0000
14	0.5922	0.6784	1.0000	0.6196	0.6490	1.0000
15	0.6059	0.7059	1.0000	0.6235	0.6392	1.0000
16	0.5922	0.6804	1.0000	0.6176	0.6529	1.0000
17	0.5765	0.6667	1.0000	0.6275	0.6490	1.0000
18	0.5529	0.6804	1.0000	0.6255	0.6294	1.0000
19	0.5725	0.6529	1.0000	0.6235	0.6549	1.0000
20	0.5843	0.6745	1.0000	0.6255	0.6608	1.0000
21	0.5549	0.6588	1.0000	0.6333	0.6373	1.0000
22	0.5784	0.6196	1.0000	0.6353	0.6490	1.0000

Table 3. *Cont.*

PC(n)	KNN	NB	DT	LDA	SVM	RF
23	0.5882	0.6294	1.0000	0.6353	0.6431	1.0000
24	0.6039	0.6549	1.0000	0.6569	0.6529	1.0000
25	0.5863	0.6451	1.0000	0.6608	0.6667	1.0000
26	0.5608	0.6431	1.0000	0.6667	0.6510	1.0000
27	0.5784	0.6333	1.0000	0.6647	0.6608	1.0000
28	0.5686	0.6471	1.0000	0.6725	0.6588	0.9980
29	0.5471	0.6412	1.0000	0.6706	0.6647	1.0000
30	0.5824	0.6255	1.0000	0.6647	0.6569	1.0000
31	0.5686	0.6275	1.0000	0.6647	0.6765	1.0000
32	0.5961	0.6294	1.0000	0.6745	0.6765	1.0000
33	0.5941	0.6353	1.0000	0.7078	0.6627	0.9961
34	0.6333	0.6451	1.0000	0.7157	0.6824	1.0000
35	0.5784	0.6471	1.0000	0.7235	0.6804	0.9980
36	0.6098	0.6451	1.0000	0.7216	0.6510	1.0000

Shown in Table 4 are the best results from Table 3 with calculated metrics: {ACC, ERR, PRE, SEN, SPE, F1}.

Table 4. Summary of machine learning algorithm performances parametric data set, i.e., 510 $\times$ 4960 data points.

Metrics	KNN	NB	DT	LDA	SVM	RF
ACC	0.9706	0.9333	1.0000	0.7235	0.9765	1.0000
ERR	0.0294	0.0667	0.0000	0.2843	0.0235	0.0000
PRE	0.9706	0.9340	1.0000	0.7288	0.9772	1.0000
SEN	0.9706	0.9333	1.0000	0.7157	0.9765	1.0000
SPE	0.9706	0.9333	1.0000	0.7157	0.9765	1.0000
F1	0.9706	0.9333	1.0000	0.7116	0.9765	1.0000

We varied the neighboring parameter k and number of principal components as a function of accuracy for the KNN models. The results are shown in Figure 10 for the full data set and in Figure 11 for the reduced data set. As can be seen, only one neighbor and one principal component were required for accuracies of 96% and 97%, respectively. Based on the overall results, it can be concluded that ML algorithms showed similarly high accuracy despite a much smaller number of input data after parameterization. Only in the case of the LDA model, can a reduction in the effectiveness of the model be observed. In the case of the remaining ML methods, there is not even a slight change in the results. This means that the use of state space modeling does not affect the accuracy of ML models and additionally allows for obtaining similar results to the case of using the full data set. It should be emphasized that state space modeling reduced the dimensionality by 83%.

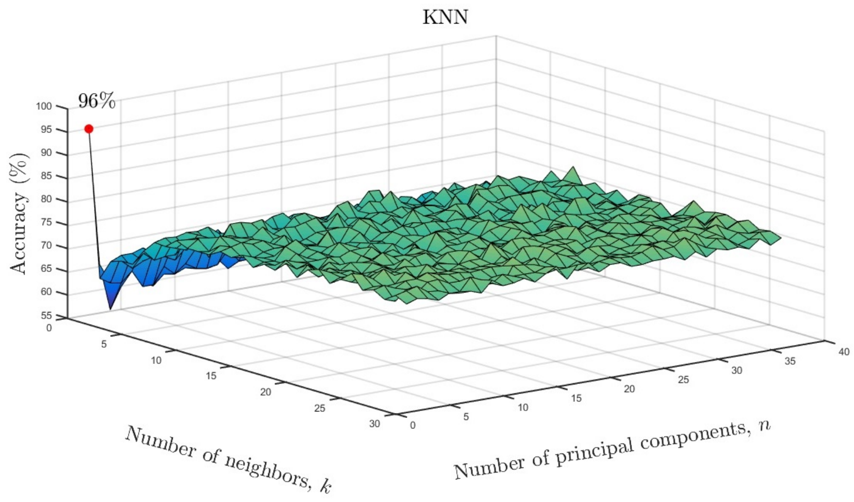

Figure 10. KNN accuracy as a function of number of neighbors and number of significant principal components for the full data set.

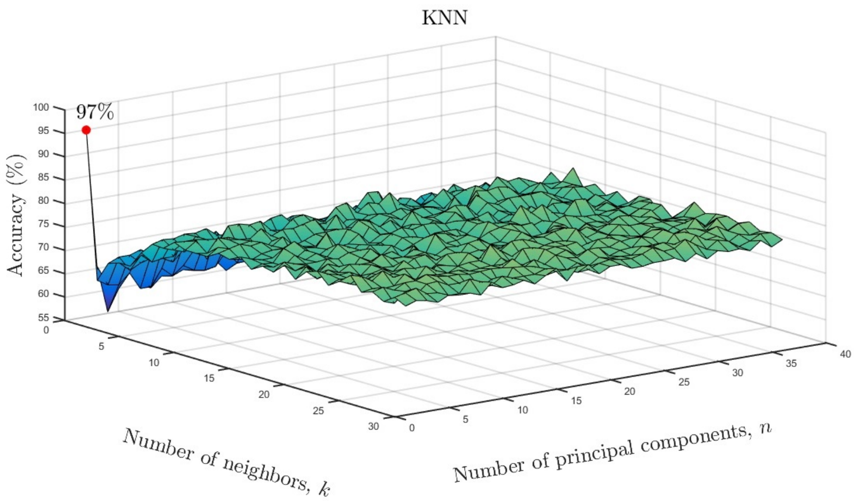

Figure 11. KNN accuracy as a function of number of neighbors and number of significant principal components for the reduced data set.

8. Conclusions

We applied six different ML algorithms to analyze and classify EEG signals collected from 62 scalp electrodes, and we used state space modeling to reduce dimensionality before applying these algorithms. Our findings revealed that the algorithms yielded high accuracy rates comparable to those obtained without application of the state space modeling. The obtained results are important because the use of state space modeling for this purpose has not been previously described in the literature and may spark new ideas for the development of ML algorithms.

It is worth noting that, when working with large data sets, dimensionality reduction is essential for signal classification, noise reduction, and may ultimately improve the predictive power of ML models. Furthermore, it is important to weigh the trade-offs between size of the data matrices and the number of parameters, where a parsimonious model (i.e., a model with a minimum number of parameters) is always preferred.

The ML methods employed in this study successfully classified, with a high degree of accuracy, Go and NoGo trials in a task in which Go and NoGo trials were equiprobable, which made it more difficult to distinguish between the two trial types. Go trials are usually presented more often than NoGo trials, e.g., 80%/20%, respectively, which primes the Go response. Once primed, greater control is required to stop or inhibit the Go response during NoGo trials. We presented an equal number of Go and NoGo trials because when an unequal number is presented, it cannot be determined whether the neural response on NoGo trials is due to response inhibition or to the relative novelty of the less frequent NoGo stimulus [15,16]. Thus, to avoid the influence of stimulus probability, we presented Go and NoGo trials with equal frequency. Research shows that when Go and NoGo trials occur with equal frequency, the neural response to the Go and NoGo trials is more similar, which increases the difficulty of distinguishing between trial types [14,41]. Our findings suggest that ML algorithms may be useful to classify neural electrical responses that may otherwise be difficult to distinguish. For instance, in early or pre-clinical cases associated with deficient inhibition, such as ADHD and Parkinson's Disease, ML algorithms may assist with early detection and diagnosis since research reveals smaller NoGo N2 ERP amplitude in patients compared to controls [42,43]. In pre-clinical cases, ML algorithms may detect small changes in the N2 ERP signal that may be missed by visual inspection alone.

Compared to existing methods, the use of state space modeling on preprocessed data used in ML algorithms makes it possible to reduce the sizes of the input data. This allows ML algorithms to run faster and to use a larger number of input variables to classify data, even with a small number of samples. Reducing dimensionality also significantly affects the running time of ML algorithms. This approach is important because a smaller number of input parameters has a positive impact on the interpretability of the results and the operation of ML algorithms that are susceptible to overfitting. Given the successful application of state space modeling to ERP signals in the current study, future studies may want to explore this data reduction approach in other biological signals.

Author Contributions: Conceptualization, J.A.R. and M.F.; Methodology, A.B., J.A.R. and M.F.; Software, A.B. and J.A.R.; Validation, A.B. and J.A.R.; Formal analysis, A.B., J.A.R. and M.F.; Investinal draft, A.B., J.A.R. and M.F.; Writing – review & editing, Anngation, A.B. and J.A.R.; Data curation, M.F.; Writing, A.B., J.A.R. and M.F.; Visualization, A.B. and J.A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This project is based upon work funded by the National Science Foundation (No. BCS–1632377) awarded to Mercedes Fernández.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Nova Southeastern University (IRB approval # 2016-226-NSU, on 10 June 2016).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The dataset used in this study will be made available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zemouri, R.; Zerhouni, N.; Racoceanu, D. Deep learning in the biomedical applications: Recent and future status. *Appl. Sci.* **2019**, *9*, 1526. [CrossRef]
2. Li, Y.; Huang, C.; Ding, L.; Li, Z.; Pan, Y.; Gao, X. Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. *Methods* **2019**, *166*, 4–21. [CrossRef] [PubMed]
3. Subasi, A.; Ercelebi, E. Classification of EEG signals using neural network and logistic regression. *Comput. Methods Programs Biomed.* **2005**, *78*, 87–99. [CrossRef]
4. Guo, L.; Rivero, D.; Seoane, J.A.; Pazos, A. Classification of EEG signals using relative wavelet energy and artificial neural networks. In Proceedings of the First ACM/SIGEVO Summit on Genetic and Evolutionary Computation, Shanghai China, 12–14 June 2009; ACM: New York, NY, USA, 2009; pp. 177–184.
5. Aldayel, M.; Ykhlef, M.; Al-Nafjan, A. Recognition of consumer preference by analysis and classification EEG signals. *Front. Hum. Neurosci.* **2021**, *14*, 604639. [CrossRef]
6. Abbasi, B.; Goldenholz, D.M. Machine learning applications in epilepsy. *Epilepsia* **2019**, *60*, 2037–2047. [CrossRef] [PubMed]
7. Jang, K.I.; Kim, S.; Chae, J.H.; Lee, C. Machine learning-based classification using electroencephalographic multi-paradigms between drug-naïve patients with depression and healthy controls. *J. Affect. Disord.* **2023**, *338*, 270–277. [CrossRef]

8. Houssein, E.H.; Hammad, A.; Ali, A.A. Human emotion recognition from EEG-based brain–computer interface using machine learning: A comprehensive review. *Neural Comput. Appl.* **2022**, *34*, 12527–12557. [CrossRef]
9. Hámori, G.; File, B.; Fiath, R.; Paszthy, B.; Réthelyi, J.M.; Ulbert, I.; Bunford, N. Adolescent ADHD and electrophysiological reward responsiveness: A machine learning approach to evaluate classification accuracy and prognosis. *Psychiatry Res.* **2023**, *323*, 115139. [CrossRef]
10. Saeidi, M.; Waldemar, K.; Farahani, F.V. Neural Decoding of EEG Signals with Machine Learning: A Systematic Review. *Brain Sci.* **2021**, *11*, 1525. [CrossRef]
11. Ouyang, G.; Zhou, C. Exploiting Information in Event-Related Brain Potentials from Average Temporal Waveform, Time–Frequency Representation, and Phase Dynamics. *Bioengineering* **2023**, *9*, 1054. [CrossRef]
12. Donchin, E.; Coles, M.G.H. Is the P300 component a manifestation of context updating? *Behav. Brain Sci.* **1988**, *11*, 357–374. [CrossRef]
13. Criaud, M.; Boulinguez, P. Have we been asking the right questions when assessing response inhibition in go/no-go tasks with fMRI? A meta-analysis and critical review. *Neurosci. Biobehav. Rev.* **2013**, *37*, 11–23. [CrossRef]
14. Nieuwenhuis, S.; Yeung, N.; van den Wildenberg, W.; Ridderinkhof, K.R. Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cogn. Affect. Behav. Neurosci.* **2003**, *3*, 17–26. [CrossRef]
15. Fernandez, M.; Tartar, J.; Padron, D.; Acosta, J. Neurophysiological marker of inhibition distinguishes language groups on a non-linguistic executive function test. *Brain Cogn.* **2013**, *83*, 330–336. [CrossRef] [PubMed]
16. Fernandez, M.; Acosta, J.; Douglass, K.; Doshi, N.; Tartar, J. Speaking Two Languages Enhances an Auditory but Not a Visual Neural Marker of Cognitive Inhibition. *AIMS Neurosci.* **2014**, *1*, 145–157. [CrossRef]
17. Falkenstein, M.; Hoormann, J.; Hohnsbein, J. ERP components in Go/Nogo tasks and their relation to inhibition. *Acta Psychol.* **1999**, *101*, 267–291. [CrossRef]
18. DeLaRosa, B.L.; Spence, J.S.; Motes, M.A.; To, W.; Vanneste, S.; Kraut, M.A.; Hart, J., Jr. Identification of selection and inhibition components in a Go/NoGo task from EEG spectra using a machine learning classifier. *Brain Behav.* **2020**, *10*, e01902. [CrossRef] [PubMed]
19. Dück, K.; Overmeyer, R.; Mohr, H.; Endrass, T. Are electrophysiological correlates of response inhibition linked to impulsivity and compulsivity? A machine-learning analysis of a Go/Nogo task. *Psychophysiology* **2023**, *60*, e14310. [CrossRef] [PubMed]
20. Singh, A.K.; Krishnan, S. Trends in EEG signal feature extraction applications. *Front. Artif. Intell.* **2023**, *5*, 1072801. [CrossRef]
21. Melnik, A.; Legkov, P.; Izdebski, K.; Kärcher, S.M.; Hairston, W.D.; Ferris, D.P.; König, P. Systems, Subjects, Sessions: To What Extent Do These Factors Influence EEG Data? *Front. Hum. Neurosci.* **2017**, *11*, 150. [CrossRef]
22. Rabcan, J.; Levashenko, V.; Zaitseva, E.; Kvassay, M. Review of methods for EEG signal classification and development of new fuzzy classification-based approach. *IEEE Access* **2020**, *8*, 189720–189734. [CrossRef]
23. Zubair, M.; Belykh, M.V.; Naik, M.U.K.; Gouher, M.F.M.; Vishwakarma, S.; Ahamed, S.R.; Kongara, R. Detection of epileptic seizures from EEG signals by combining dimensionality reduction algorithms with machine learning models. *IEEE Sens. J.* **2021**, *21*, 16861–16869. [CrossRef]
24. Cheung, B.L.; Riedner, B.; Tononi, G.; Van Veen, B.D. Estimation of cortical connectivity from EEG using state-space models. *IEEE Trans. Biomed. Eng.* **2010**, *57*, 2122–2134. [CrossRef]
25. Fernandez, M.; Banks, J.B.; Gestido, S.; Morales, M. Bilingualism and the executive function trade-off: A latent variable examination of behavioral and event-related brain potentials. *J. Exp. Psychol. Learn. Mem. Cogn.* **2023**, *49*, 1119–1144. [CrossRef] [PubMed]
26. Jing, H.; Takigawa, M. Low sampling rate induces high correlation dimension on electroencephalograms from healthy subjects. *Psychiatry Clin. Neurosci.* **2000**, *54*, 407–412. [CrossRef] [PubMed]
27. Semlitsch, H.V.; Anderer, P.; Schuster, P.; Presslich, O. A Solution for Reliable and Valid Reduction of Ocular Artifacts, Applied to the P300 ERP. *Psychophysiology* **1986**, *23*, 695–703. [CrossRef] [PubMed]
28. Leon-Medina, J.X. Desarrollo de un Sistema de Clasificación de Sustancias Basado en un Arreglo de Sensores Tipo Lengua Electrónica. Ph.D. Thesis, Universidad Nacional de Colombia, Facultad de Ingeniería, Departamento de Ingeniería Mecánica y Mecatrónica, Bogotá, Colombia, 2021.
29. Kung, S. A New Identification and Model Reduction Algorithm via Singular Value Decomposition. In Proceedings of the 12th Asilomar Conference on Circuits, Systems and Computers, Pacific Grove, CA, USA, 6–8 November 1978; pp. 705–714.
30. Mercère, G.; Prot, O.; Ramos, J.A. Identification of parameterized gray-box state-space systems: From a black-box linear time-invariant representation to a structured one. *IEEE Trans. Autom. Control* **2014**, *59*, 2873–2885. [CrossRef]
31. Sinha, N.K. Identification of continuous-time systems from samples of input-output data: An introduction. *Sadhana* **2000**, *25*, 75–83. [CrossRef]
32. Van Overschee, P.; De Moor, B. *Subspace Identification for Linear Systems: Theory—Implementation—Applications*; Springer Science & Business Media: Berlin, Germany, 2012.
33. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
34. Omidvar, M.; Zahedi, A.; Bakhshi, H. EEG signal processing for epilepsy seizure detection using 5-level Db4 discrete wavelet transform, GA-based feature selection and ANN/SVM classifiers. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*, 10395–10403. [CrossRef]

35. Upadhyay, P.K.; Nagpal, C. Wavelet based performance analysis of SVM and RBF kernel for classifying stress conditions of sleep EEG. *Sci. Technol.* **2020**, *23*, 292–310.
36. George, F.P.; Shaikat, I.M.; Ferdawoos, P.S.; Parvez, M.Z.; Uddin, J. Recognition of emotional states using EEG signals based on time-frequency analysis and SVM classifier. *Int. J. Electr. Comput. Eng.* **2019**, *9*, 2088–8708. [CrossRef]
37. Bonaccorso, G. *Machine Learning Algorithms*; Packt Publishing Ltd.: Birmingham, UK, 2017.
38. Kurita, T. Principal component analysis (PCA). In *Computer Vision: A Reference Guide*; Springer: Cham, Switzerland, 2019; pp. 1–4.
39. Ljung, L. *System Identification: Theory for the User*; Prentice Hall: Upper Saddle River, NJ, USA, 1987.
40. Ballabio, D.; Grisoni, R.; Todeschini, R. Multivariate comparison of classification performance measures. *Chemom. Intell. Lab. Syst.* **2018**, *174*, 33–44. [CrossRef]
41. Lavric, A.; Pizzagalli, D.A.; Forstmeier, S. When 'go' and 'nogo' are equally frequent: ERP components and cortical tomography. *Eur. J. Neurosci.* **2004**, *20*, 2483–2488. [CrossRef]
42. Smith, J.L.; Johnstone, S.J.; Barry, R.J. Inhibitory processing during the Go/NoGo task: An ERP analysis of children with attention-deficit/hyperactivity disorder. *Clin. Neurophysiol.* **2004**, *115*, 1320–1331. [CrossRef] [PubMed]
43. Wu, H.M.; Hsiao, F.J.; Chen, R.S.; Shan, D.E.; Hsu, W.Y.; Chiang, M.C.; Lin, Y.Y. Attenuated NoGo-related beta desynchronisation and synchronisation in Parkinson's disease revealed by magnetoencephalographic recording. *Sci. Rep.* **2019**, *9*, 7235. [CrossRef] [PubMed]