Prediction of Successful Memory Formation during Audiovisual advertising using EEG signals

Vangelis P. Oikonomou*,1, Kostas Geordiadis[1], Fotis P. Kalaganis[1], Spiros Nikolopoulos[1] and Ioannis Kompatsiaris[1]

***Abstract*—The prediction of memory performance using EEG signals is an active research area in Passive Brain Computer Interfaces community. In this type of prediction problem it is important to be able to use unlabeled data and to tackle the unbalanced nature of the data. In this work we propose two new Sparse Representation Classification schemes that are able to address the above properties of the data. The proposed classifiers have been tested in an EEG dataset related to neuromarketing. In the data analysis, we define a binary classification problem using EEG signals corresponding to the condition of remembering and forgetting. Furthermore, we compare the proposed classifiers with well-known classifiers. The obtained results show all classifiers perform above chance level, and, among them the proposed classifiers present the best performance in terms of Fscore and Kappa Value.**

***Index Terms*—Neuromarketing, Memory, EEG, Sparse Representation Classification, Semi-supervised Learning**

I. Introduction

Neuromarketing is an evolving emerging field combining consumer's behavior with neuroscience [1], hence its alternative name consumer's neuroscience, which gives rise to a more strict definition that defines neuromarketing as the application of neuroscience in the marketing domain. The overarching goal of neuromarketing is to understand customers' motivations, preferences, and decisions, which can help to inform creative advertising, product development, pricing, and other marketing areas [2], [3]. In order to achieve this goal it is necessary to acquire measurements of physiological and neural signals, which describes the participant's reaction due to the marketing stimuli. To this end, brain imaging technologies, which measure neural activity, and physiological tracking technologies, which measure eye movement and other proxies for that activity, are the most common methods of measurement [3]. Among the various brain imaging methods, electroencephalography (EEG) represents the most used method, mainly because it is the least invasive and cheaper method with high temporal resolution, compared to other brain imaging technologies such as functional Magnetic Resonance Imaging (fMRI) or Positron Emission Tomography (PET).

The field of EEG-based Neuromarketing presents a variety of approaches for measuring functionally relevant neural activity. For example, some research has focused on neural activity measurements during task performance (e.g buy/no-buy scenario, like /dislike) [4]–[6], which can link discrete neural signatures to behavioral outcomes recorded simultaneously. Another direction of research is related to 'task-free' neural recordings taken during passive stimulation conditions (e.g., naturalistic viewing). Task-free scenarios have increased in popularity because they provide a broader view on neural dynamics that exceed circumscribed, task scenarios [7], [8].

In neuromarketing community, the promotion of products and brands have been explored in many directions using audiovisual stimuli involving consumer's engagement [9], pleasantness [10] and memorization of commercials [11], [12]. It becomes evident that a core question in neuromarketing is if it is possible to observe neural signatures during memory encoding that can be used to evaluate an advertisement in term of memory performance. Assuming that such signatures are existent, they could be used to define important parts of the commercial in terms of memory. In the proposed work, we provide a methodology for the prediction of memory performance using EEG signals when the consumer (or subject or participant) is exposed to audiovisual marketing stimuli (i.e. video ads).

Memory formation (or encoding) seems to be a rather complicated cognitive and perceptual process if we take into account that various frequency bands are related to memory formation in various different ways. While most studies reported that memory formation is related with theta and gamma brain rhythms, alpha and beta bands seem to be equally important [13], [14]. These diverse effects are related to the engagement during memory encoding (i.e. items to be remembered) and how the memory is being tested (i.e. context of memory recall, the time between encoding and recall) [13]–[16]. A further complication in the literature arises from the fact that some studies found neural signatures in post-stimulus signals (i.e., after the stimulus had been shown), while other studies also have found significant neural signatures preceding stimulus onsets [13], [14]. Based on the above it is natural to expect that the classification of EEG trials related to the remembered (RMB) and forgotten (FRG) items would be a very difficult machine learning task as pointed out in

* Corresponding author.

[1]V. P. Oikonomou, K. Georgiadis, F.P. Kalaganis, S. Nikolopoulos and I. Kompatsiaris are with the Information Technologies Institute, Centre for Research and Technology Hellas, CERTH-ITI, 6th km Charilaou-Thermi Road, 57001 Thermi-Thessaloniki, Greece. {viknmu, kostas.georgiadis, fkalaganis, nikolopo, ikom}@iti.gr

This was supported by the NeuroMkt project, and co-financed by the European Regional Development Fund of the EU and Greek National Funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under RESEARCH CREATE INNOVATE (T2EDK-03661).

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00200

[13], [15], [17]. However, the results of [13], [15], [17] are encouraging since they provide above-chance classification rates indicating that neural signatures of memory formation can be used for predictive purposes.

In this work, we propose a new classifier for the discrimination of neural signatures of memory (i.e., FRG/RMB) during a neuromarketing experiment. The proposed classifier is based on the idea of sparse representations, called Sparse Representation Classification (SRC) [18]–[20]. The basic assumption is that brain activity patterns, belonging to the same memory encoding process, lie on the same linear subspace. The contributions of our paper are:

- We explore the sparsity of brain activity in neuromarketing scenarios and we propose a novel SRC-based classification algorithm with applications to memory formation. More specifically, we extend the algorithm provided in [19] to take into account pre-stimulus effects of memory during the classification procedure
- We performed a neuromarketing-related study which includes: a) a large number of participants (85 participants) b) task-free and naturalistic (and dynamic) stimuli (i.e. videos), and c) real life conditions for the experimental protocol. The above experiments resulted to an EEG dataset related to the memory encoding of audiovisual stimuli (i.e. ads).
- We carry out extensive data analysis experiments, and the results demonstrate that our proposed framework achieves superior performance in comparison with the existing state-of-the-art approaches on the same EEG-based neuromarketing dataset.

The paper is organized as follows. In Section II, we provide information about the EEG dataset that is used to predict memory performance. Then, in Section III a description of the overall approach and methodology is provided for the prediction of memory performance during a neuromarketing scenario using EEG signals. In Section IV, we present the results from our experiments and we provide a comparative analysis with well-known classifiers. Also, we provide a discussion related to our work. Finally, in Section V, some concluding remarks are drawn.

II. Materials - Dataset's Description

A. Participants (Demographics)

Eighty Five (85) healthy volunteers (28 males and 57 females) with an average age of 43 years (43 ± 13, ranging from 19 to 78 years) participated in our study. All the participants had a normal or corrected to normal vision, and they all signed a written consent before the experiment. The study was carried out in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of our Institution.

B. Description of Stimuli and Experiment's Procedure

Audiovisual advertising is a type of advertising that comes with high levels of potential and includes images, sounds and motion. The audiovisual message is the medium that offers advertisers the most effective and persuasive way to communicate with consumers. In our study, the audio-visual (dynamic) advertising message is a TV spot, in four different versions, which consumers see on their set-top boxes every day. The four different versions are related to the presence of a male and female voice in the spot, as well as the fact that the products are presented in a different order in each spot. Finally, besides the four products, the TV spot includes an introductory video which is the same across all versions of the TV spot. Hence, the TV spot can be divided into five segments: introductory video, ads for product 1, ads for product 2, ads for product 3 and ads for product 4. The TV spot had a duration of 30 seconds and it was presented to the participants during the viewing of a TV cartoon with a total duration of three minutes. In Fig. 1 we provide the timeline related to the watching of video. Once the video viewing experience was completed each participant completed a questionnaire which include demographic questions (e.g. age, etc.), profile questions (e.g. buying behavior, etc.), and, questions related to which presented products the participant remembers.

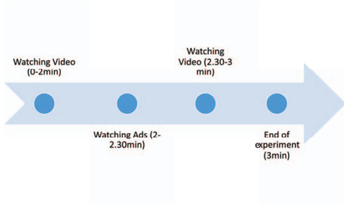

Fig. 1. Timeline of the experiment.

C. Data Collection and Preprocessing

The raw EEG data were recorded using Wearable Sensing's Dry Sensor Interface (DSI) with a sampling frequency of 300Hz, via 21 dry sensors, namely Fp1, Fp2, Fz, F3, F4, F7, F8, Cz, C3, C4, T7/T3, T8/T4, Pz, P3, P4, P7/T5, P8/T6, O1, O2, A1 and A2, that were placed at locations corresponding to the 10-20 International System. The Sensors A1 and A2 were the reference electrodes and were placed on the mastoids. Prior to the experimental procedure, impedance for all electrodes was set below 10KΩ and EEG signals were inspected to avoid any irregularities. We extracted EEG trials corresponding to a particular product for subsequent analysis, by manually identifying the starting time point and the ending time point of each commercial's product. For each participant five EEG trials are extracted, where four of them correspond to the products. For the labelling of EEG trials as FRG/RMB, we use the questionnaires. Finally, the raw EEG trials were subjected to a bandpass filter 0.5–45 Hz, followed by artifact removal using the Artifact Subspace Reconstruction (ASR) [21].

III. Proposed Framework

A. Feature extraction

The segmentation of the dataset results into five trials for each subject. From these five trials, four of them have labels as

FRG or RMB, while one trial is without label. The trial without label corresponds to the introductory video of the TV spot. More formally our dataset is composed from labeled trials, $\mathcal{D}_\ell = \{(\mathbf{X}_i, \ell_i)\}_{i=1}^N$, and unlabeled trials $\mathcal{D}_o = \{(\mathbf{X}_i)\}_{i=N+1}^{N+M}$, where $\mathbf{X}_i \in \Re^{N_{ch} \times N_t}$, N_{ch} denotes the number of channels, N_t denotes the number of samples, N denotes the number of labeled trials and M denotes the number of trials without labels. In our work we computed for each trial power's features per channel, hence, the feature vector $\mathbf{f}_i$, describing trial $\mathbf{X}_i$, is of size $N_{ch} \times 1$. These type of features describes the strength of brain activity. Note here that we assume that the multi-channel EEG signals are centralized since, in practice, the EEG trials are bandpass filtered.

B. Proposed Semi-supervised SRC-based classification framework

SRC-based classification frameworks use the training samples directly as the basis to construct the overcomplete dictionary. The idea behind this approach is that a test sample can be accurately represented by a linear combination of training examples from the same class. Hence, in terms of EEG-related studies, the idea is that brain's features of a test example can be represented well by a sparse linear combination of brain's features from the same class. In this subsection, we provide a short introduction to the basic SRC scheme, and then, we describe the proposed semi-supervised SRC scheme.

Given the labeled dataset $\mathcal{D} = \{(\mathbf{f}_i, \ell_i)\}_{i=1}^N$, where $\mathbf{f}_i$ are the feature vectors and ℓ_i the corresponding labels, we can collect all the features vectors in a matrix, $\mathbf{X} \in \Re^{N_{ch} \times N}$. The basic idea behind SRC is that the label of the test vector $\mathbf{y} \in \Re^{N_{ch}}$ is unknown; however, we can represent it as a linear combination of the training samples from all classes, where their labels are known:

$$\mathbf{y} = \mathbf{X}\mathbf{w} \tag{1}$$

where $\mathbf{X} \in \Re^{N_{ch} \times N}$ is a matrix containing all the training vectors from all classes, N is the number of training vectors, and $\mathbf{w} \in \Re^N$ is the coefficient vector. Furthermore, in the presence of noise, the model is provided by:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{e} \tag{2}$$

where $\mathbf{e} \in \Re^{N_{ch}}$ is the noise term with bound energy $\|\mathbf{e}\|_2 \leq \epsilon$. In this case the coefficients $\mathbf{w}$ are found by solving the following minimization problem:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \{\|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \rho\|\mathbf{w}\|_1\}. \tag{3}$$

Now that we have seen how a test vector can be described as a linear combination of training vectors, we will discuss how we could use this linear combination to provide a classification rule. In order to provide the classification rule, we use the residuals of linear combination. More specifically, if $\delta_c(\cdot) : \Re^N \to \Re^N$ is the function that selects the coefficients associated with the class c, then the residuals for each class is: $r_c(\mathbf{y}) = \|\mathbf{y} - \mathbf{X}\delta_c(\hat{\mathbf{w}})\|_2$, $c = 1, \cdots, C$. The class for the given test signal is found by using the minimum of the residuals $class(\mathbf{y}) = \arg\min_c\{r_c(\mathbf{y})\}$. We can see that the algorithm contains two basic steps. The first step is related to the minimization problem, while the second step is related to the classification rule.

Based on the above algorithm various extensions have been proposed related to the optimization problem of Step 1, as well as to the calculation of residuals in Step 2. In the proposed work we adopt the algorithm provided in [19]. This algorithm exploits the manifold's structure of the data be utilizing a specialized prior which has properties rising from graph theory, while at the same time has a tendency for sparsity. Besides the solver that someone adopts to find the weights of linear combination, one important aspect is the matrix $\mathbf{X}$ which contains the training samples (or the dictionary matrix). In our approach we extend this matrix by including EEG samples that do not possess any labels and preceding the presentation of four ads. These EEG samples are extracted from the time period corresponding to the introduction of the TV spot. With this extension, the weights can be divided into two groups, weights that are related to training samples with labels and weights that are related to training samples without labels. The main effect of this extension is that the estimation procedure for the weights of the labeled training samples take into account information related to the pre-stimulus status of the brain. Overall this affects the calculation of residuals, and hence the classification performance.

A more formal description about the new extended matrix $\mathbf{X}_e$ is presented next. Given a dataset $\mathcal{D}$, $\mathcal{D} = \mathcal{D}_\ell \cup \mathcal{D}_o$, that constitutes from training samples with labels, $\mathcal{D}_\ell = \{(\mathbf{f}_i, \ell_i)\}_{i=1}^N$, and without labels, $\mathcal{D}_o = \{(\mathbf{f}_i)\}_{i=N+1}^{N+M}$ where $\mathbf{f}_i$ are feature vectors of size $N_{ch} \times 1$ and ℓ_i the corresponding labels (if exist), we collect all features vectors in a matrix (extended version), $\mathbf{X}_e \in \Re^{N_{ch} \times (N+M)}$. Observe here that the sub-matrix $\mathbf{X}_o \in \Re^{N_{ch} \times (M)}$, corresponding to the columns $N+1$ to $N+M$ of $\mathbf{X}_e$, contains features vector that do not have labels. The adoption of the above extended matrix changes significantly the properties of the basic SRC algorithm since it gives to it the ability to treat training samples which do not possess any label's information, hence it can be considered as a semi-supervised extension of SRC algorithm presented in [19]. Furthermore, in the case where we need to address the problem of unbalanced classes we can weight each training sample. This end up to the use of diagonal matrix, $\mathbf{X}_u \in \Re^{(N+M) \times (N+M)}$, which in its main diagonal contains the weights for each training sample. In that case the extended matrix is modified as: $\mathbf{X}_e^{'} = \mathbf{X}_e\mathbf{X}_u$. The overall algorithm is provided in Alg. 1 for the case of balanced dataset (we called this algorithm *semiSRC*). In the case of unbalanced classification problem, the modified extended matrix is adopted (we called this modified algorithm, *un-semiSRC*).

IV. Results

The pre-processed and segmented EEG dataset consist of 340 labeled trials (from them 240 are labeled as RMB and the rest 80 as FRG), and, 85 unlabeled trials. Power EEG

Algorithm 1 Proposed Semi-supervised Sparse Representation Classification scheme (Adapted from [19])

Require: Training samples, $\mathbf{X}_\ell$, with its corresponding labels, ℓ, unlabeled training samples, $\mathbf{X}_o$, one test sample, $\mathbf{y}$.
1. Construct extended matrix $\mathbf{X}_e$ by concatenating matrices $\mathbf{X}_\ell$ and $\mathbf{X}_o$, $\mathbf{X}_e = [\mathbf{X}_\ell \; \mathbf{X}_o]$.
2. Construct graph Laplacian matrix, L (as reported in [19]).
3. Find $\hat{\mathbf{w}}$ uses Algorithm 2 reported in [19],
4. Calculate the residuals:
$r_c(\mathbf{y}) = \|\mathbf{y} - \mathbf{X}_e \delta_c(\hat{\mathbf{w}})\|_2$, $c = 1, \cdots, C$

Ensure: $class(\mathbf{y}) = \arg\min_c \{r_c(\mathbf{y})\}$

features are extracted from each EEG trial. More specifically, the energy of EEG signals in each channel is calculated. These results into 19 features characterizing each trial and they provide information about the activation patterns of the brain due to the naturalistic stimuli. The features are fed to the classifier in order to characterize the trial as FRG or RMB. In our analysis we compare the proposed algorithms (*semiSRC* and *un-semiSRC*) with three well known classifiers: the SVM with Radial Basis Functions (RBF) kernel [2], [19], [22], [23] (*SVM-RBF*), the k-Nearest Neighborhood (*kNN*) with $k = 1$ [2], [19], and the Graph-based Sparse Representations Classification (*GraphSRC*) scheme [19]. For the SVM-RBF algorithm, we take into consideration the unbalanced nature of our classification problem, and more specifically, we weight each class by using the inverse of the class distribution present in the training dataset. This weighting scheme was introduced to the algorithm through the cost matrix. A similar procedure was adopted for the proposed *un-semiSRC* algorithm. To train and evaluate the above classifiers the Leave-One-Out cross validation approach is used. Furthermore, special care must be taken related to the evaluation of the performance of classifiers due to the unbalanced nature of the used dataset. Under this view we calculate the Fscore [24], and the κ coefficient (or Kappa Value) [25]. These two measures are suitable for handling the unbalanced nature of the dataset and providing accurate evaluation of the models. Since, the Fscore describes the trade off (or the balance) between precision and recall, while, the κ coefficient describes the randomness of the agreement (agreement beyond chance).

In Table I we provide the obtained results of the five classifiers for the aforementioned performance measures. Furthermore, we have included the performance of the *Naive* classifier (a classifier that classifies all samples to the majority class). The main conclusion from the provided results is that all classifiers provide performance above chance level and better than the *Naive* classifier. This indicates that the experimental protocol did elicit the mental states of interest and that brain produces activation patterns that are different between the two groups (FRG vs RMB). Observe here, that this is consisted across a number of classifiers which indicates its universal nature. Furthermore, the *un-semiSRC* method provides the best performance among all methods since it presents better agreement between predicted and actual classes (higher κ coefficient) by correctly identifying both positive and negative instances (higher Fscore) compared to the other models. Comparing the various SRC versions, we can observe that the *semiSRC* present better performance from *GraphSRC* in terms of Kappa Value and Fscore, while the *un-semiSRC* presents the best performance among the SRC-based algorithms. The above observations show the usefulness of the weighting scheme into the SRC scheme that address the issue of skewed distribution of class labels for the particular classification problem, as well as, the usefulness of using unlabeled EEG trials.

TABLE I
CLASSIFICATION ACCURACY RESULTS

	Fscore	Kappa Value
Naive	0.4343	0
SVM-RBF	0.6324	0.2687
kNN	0.5866	0.1740
GraphSRC	0.6042	0.2091
semiSRC	0.6222	0.2448
un-semiSRC	**0.6399**	**0.2836**

We perform additional experiments by investigating the performance of classifiers in various frequency bands of EEG signal (i.e. EEG bands). Again, the performance was evaluated by using the aforementioned measures. The Frequency bands that we choose are corresponding to the basic EEG rhythms (delta:1-4Hz, theta:4-8Hz, alpha:8-12Hz, beta:13-30Hz and gamma:>30Hz). The obtained results are provided in Tables II, for the Kappa Value, and III, for the Fscore. The best performance for Kappa Value was obtained by *un-semiSRC* in the theta band, while, the best performance in terms of Fscore metric was obtained by *semiSRC* in the alpha band. It is worth to mention here that all algorithms in all frequency bands provided Kappa Value larger than 0.1, besides the *SVM-RBF* in the theta band. Furthermore, the *un-semiSRC* method is the only method that presents fair agreement (Kappa value>0.2) in more than one frequency band, while the other SRC versions and the SVM-RBF present slight agreement (Kappa value:0-0.2) in all frequency bands. With respect to the Fscore measure, the *un-semiSRC* and the *semiSRC* are the only methods that presents Fscore >0.6 in at least one frequency bands. Also we can observe that all models present Fscore values larger than the *Naive* model. Note here that the *Naive* model has the same Fscore value (0.4343) for all reported classification problems. Finally, one significant property of the proposed algorithm is that it is able to use pre-stimulus, hence unlabeled in our case, brain data to solve the classification problem. This property is evident in Tables II and III. More specifically, we can observe that the *semiSRC* algorithm has better performance from *GraphSRC* in all frequency bands, and, in addition it has better performance than the *SVM-RBF* algorithm in most of the frequency bands.

By taking a look at Tables II and III we can observed that there is not a single classifier that performs best at all frequency bands. For example, at frequency band 1-4Hz the

TABLE II
CLASSIFICATION ACCURACY RESULTS PER EEG BAND/RHYTHM - KAPPA VALUE

	SVM-RBF	*kNN*	*GraphSRC*	*semiSRC*	*un-semiSRC*
1-4Hz	0.1920	0.1920	0.1559	0.1668	0.2189
4-8Hz	0.0872	0.1319	0.1528	0.1619	**0.2197**
8-12Hz	0.1649	0.1873	0.1974	0.2179	0.1857
13-30Hz	0.1500	0.1824	0.1552	0.1751	0.2090
30-45Hz	0.1500	0.1920	0.1090	0.0959	0.1592

TABLE III
CLASSIFICATION ACCURACY RESULTS PER EEG BAND/RHYTHM - FSCORE

	SVM-RBF	*kNN*	*GraphSRC*	*semiSRC*	*un-semiSRC*
1-4Hz	0.5959	0.5959	0.5749	0.5817	0.6007
4-8Hz	0.5430	0.5657	0.5725	0.5794	0.6033
8-12Hz	0.5787	0.5937	0.5956	**0.6066**	0.5886
13-30Hz	0.5679	0.5910	0.5768	0.5871	0.6024
30-45Hz	0.5679	0.5960	0.5528	0.5460	0.5784

best classifier is the *un-semiSRC*, while at frequency band 30-45Hz the *kNN* provides the best performance. As we see some situations (frequency bands in our analysis) favor a specific classifier, and some other not. Hence, it raises the issue of classifier's robustness, as stated in [26], or how well a classifier behaves on average in scenarios where favors other classifiers. A quantity to check the robustness of an classification algorithm among a given group of algorithms is the ratio of the performance measure in a given situation to the largest performance measure for the particular situation [26]. Under this view, the robustness of the best classifier is equal to 1 and all other classifiers will have values smaller than 1. This quantity shows us how much deviate a specific classifier from the best achievable performance. In Fig. 2 we depict the distribution of each classifier's robustness over all frequency bands. The robustness has been computed for the Kappa Value metric. From this figure we can concluded that the most robust classifier is the *un-semiSRC* with a robustness of 0.9468 for the particular experiments.

A. Discussion

A significant property of the SRC schemes, as well as the kNN schemes, is that, while they need training data to perform the decision, they do not need any training procedure to tune the model's parameters (at least in principle). Clearly someone could used the training data to find for example the optimal neighborhood in kNN, however such cases are beyond our scope here. The above property affects significantly the modeling approach since the model's parameters need not to be tuned (ie. retraining the model) if the data distributions have deviated significantly from those of the original training set. However, the above property comes with increased computational cost since an optimization procedure is executed each time we test a new instance.

The necessity to retrain the model has serious implications in scenarios where the calibration of the model (the acquisition of new data and the retraining procedure) is a time consuming and ”costly” procedure. A situation which is presented many times in BCI related tasks. EEG signals are highly subject-specific and vary considerably even between recording sessions of the same subject within the same experimental paradigm. To minimize the EEG variability effects, a calibration phase on the beginning of each session is used in order to optimize models' parameters. Considerable effort have been devoted to reduce the time of calibration phase by utilizing Transfer Learning techniques, however even then, a few training samples are still required to retrain the model. However, the proposed SRC scheme doesnot need to be retrained, hence, it contributes to the efficient design of zero training BCI systems [27].

It is important to provide comments about the semi-supervised ability of our algorithms. We can observed how easily our framework utilizes the pre-stimulus condition of the brain, without the need to have labels for it. Semi-supervised learning (SSL) concerns the problem of how to improve classifiers' performance through making use of prior knowledge from unlabeled data. SSL classification algorithms are divided into two large groups [28]. This division is related on the way each algorithm treats the unlabeled trials. In the first group belongs algorithms that treat unlabeled trials as the test trials and predict their labels during the training, while the second group predicts the labels for unlabeled trials as well as the new test trials through the training procedure. As we see in both groups it is important to predict the labels of unlabeled trials. Also, the unlabeled trials are connected implicitly with the classification problem, but we are unable to have their labels *before* the training procedure and utilize them *during* the training procedure. However, in our case the unlabeled trials can not be connected implicitly (but explicitly) with the classification problem i.e. we do not desire and/or we are not able to assign any label in these trials. While these trials are not directly connected with the classification problem, there is an indirect connection which is coming from neuroscience stating that brain status before the stimulus presentation affects the memory performance of the subject. Furthermore, from a data analysis perspective, the pre-stimulus (unlabeled) trials/data can be used to construct a more informative prior distribution of stimulus (labeled) data. Overall, under the above views, our algorithms utilize the unlabeled data to constrain the data

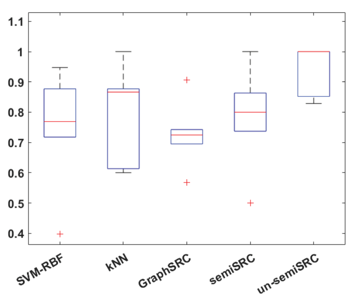

Fig. 2. Robustness of the classifiers using the Kappa Value metric.

distribution of the labeled data.

Closing this section, we want to note that the process of memorizing items (in our study TV ads) is more complex than other brain processes such as motor imagery tasks or Steady State Visual Evoked Potentials tasks, hence our lower performance levels compared to these tasks. However, restricting our view to memory tasks only, then our findings indicate that the prediction of memory performance using EEG signals in neuromarketing scenarios is a viable machine learning task. Our work contribute to the current literature of memory prediction using EEG trials [13], [15], [17]. Furthermore, by taking into account the unbalanced nature of the problem and the pre-stimulus status of the brain, the *un-semiSRC* algorithm provides results beyond random agreement, and much better performance than classical machine learning algorithms.

V. Conclusions

In this work we propose two new SRC schemes that are used for the prediction of memory performance. The basic contributions of these schemes are: the ability to dealt with unlabeled data (semi-supervised ability) and with the unbalanced nature of the data. The provided results in a neuromarketing scenario show the usefulness of our approach. In the future, first, we intend to release a comprehensive version of the dataset to the scientific community. Second, to extend the current algorithms using the ideas of kernel and riemannian geometry [29]. And third, to study and propose a new metric related to the concept of memory performance.

References

[1] D. Ariely and G. S. Berns, "Neuromarketing: the hope and hype of neuroimaging in business," *Nature Reviews Neuroscience*, vol. 11, pp. 284–292, 2010.

[2] V. Khurana, M. Gahalawat, P. Kumar, P. P. Roy, D. P. Dogra, E. Scheme, and M. Soleymani, "A survey on neuromarketing using eeg signals," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 4, pp. 732–749, 2021.

[3] F. P. Kalaganis, K. Georgiadis, V. P. Oikonomou, N. A. Laskaris, S. Nikolopoulos, and I. Kompatsiaris, "Unlocking the subconscious consumer bias: A survey on the past, present, and future of hybrid eeg schemes in neuromarketing," *Frontiers in Neuroergonomics*, vol. 2, 2021.

[4] M. Yadava, P. Kumar, R. Saini, P. P. Roy, and D. P. Dogra, "Analysis of eeg signals and its application to neuromarketing," *Multimedia Tools and Applications*, vol. 76, pp. 19087–19111, 2017.

[5] K. Georgiadis, F. P. Kalaganis, K. Riskos, E. Matta, V. P. Oikonomou, I. Yfantidou, D. Chantziaras, K. Pantouvakis, S. Nikolopoulos, N. A. Laskaris, and I. Kompatsiaris, "Neuma - the absolute neuromarketing dataset en route to an holistic understanding of consumer behaviour," *Scientific Data*, vol. 10, 2023.

[6] A. Hakim, S. Klorfeld, T. Sela, D. Friedman, M. Shabat-Simon, and D. J. Levy, "Machines learn neuromarketing: Improving preference prediction from self-reports using multiple eeg measures and machine learning," *International Journal of Research in Marketing*, vol. 38, no. 3, pp. 770–791, 2021.

[7] G. Vecchiato, L. Astolfi, F. D. Vico, F. Cincotti, D. Mattia, S. Salinari, R. Soranzo, and F. Babiloni, "Changes in brain activity during the observation of tv commercials by using eeg, gsr and hr measurements," *Brain topography*, vol. 23, pp. 165–79, 06 2010.

[8] C. Christoforou, T. Papadopoulos, F. Constantinidou, and M. Theodorou, "Your brain on the movies: A computational approach for predicting box-office performance from viewer's brain responses to movie trailers," *Frontiers in Neuroinformatics*, vol. 11, 12 2017.

[9] J. P. Dmochowski, P. Sajda, J. Dias, and L. Parra, "Correlated components of ongoing eeg point to emotionally laden attention - a possible marker of engagement?," *Frontiers in Human Neuroscience*, vol. 6, 2012.

[10] G. Vecchiato, J. Toppi, L. Astolfi, F. D. V. Fallani, F. Cincotti, D. Mattia, F. Bez, and F. Babiloni, "Spectral eeg frontal asymmetries correlate with the experienced pleasantness of tv commercial advertisements," *Medical biological engineering computing*, vol. 49, pp. 579–83, 02 2011.

[11] G. Vecchiato, W. Kong, A. G. Maglione, and D. Wei, "Understanding the impact of tv commercials: Electrical neuroimaging," *IEEE Pulse*, vol. 3, no. 3, pp. 42–47, 2012.

[12] A. Y. Shestyuk, K. Kasinathan, V. Karapoondinott, R. Knight, and R. Gurumoorthy, "Individual eeg measures of attention, memory, and motivation predict population level tv viewership and twitter engagement," *PLOS ONE*, vol. 14, pp. 1–27, 03 2019.

[13] T. Kang, Y. Chen, S. Fazli, and C. Wallraven, "Eeg-based prediction of successful memory formation during vocabulary learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 11, pp. 2377–2389, 2020.

[14] S. Hanslmayr and T. Staudigl, "How brain oscillations form memories — a processing based perspective on oscillatory subsequent memory effects," *NeuroImage*, vol. 85, pp. 648–655, 2014. New Horizons for Neural Oscillations.

[15] E. Noh, G. Herzmann, T. Curran, and V. R. de Sa, "Using single-trial eeg to predict and analyze subsequent memory," *NeuroImage*, vol. 84, pp. 712–723, 2014.

[16] E. Noh, M. V. Mollison, G. Herzmann, T. Curran, and V. R. de Sa, "Towards a passive brain computer interface for improving memory," 2014.

[17] S. Charbonnier, R. N. Roy, R. Dolezalová, A. Campagne, and S. Bonnet, "Estimation of working memory load using eeg connectivity measures," in *International Conference on Bio-inspired Systems and Signal Processing*, 2016.

[18] C. Shen, L. Chen, Y. Dong, and C. Priebe, "Sparse representation classification beyond 1 minimization and the subspace assumption," *IEEE Transactions on Information Theory*, vol. 66, no. 8, pp. 5061–5071, 2020.

[19] V. P. Oikonomou, K. Georgiadis, F. Kalaganis, S. Nikolopoulos, and I. Kompatsiaris, "A sparse representation classification scheme for the recognition of affective and cognitive brain processes in neuromarketing," *Sensors*, vol. 23, no. 5, 2023.

[20] T. Shu, B. Zhang, and Y. Tang, "Sparse supervised representation-based classifier for uncontrolled and imbalanced classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 8, pp. 2847–2856, 2020.

[21] S. Blum, N. S. Jacobsen, M. G. Bleichner, and S. Debener, "A riemannian modification of artifact subspace reconstruction for eeg artifact handling," *Frontiers in Human Neuroscience*, vol. 13, 2019.

[22] Z. Wei, C. Wu, X. Wang, A. Supratak, P. Wang, and Y. Guo, "Using support vector machine on eeg for advertisement impact assessment," *Frontiers in Neuroscience*, vol. 12, 03 2018.

[23] K. Georgiadis, F. P. Kalaganis, V. P. Oikonomou, S. Nikolopoulos, N. A. Laskaris, and I. Kompatsiaris, "Rneumark: A riemannian eeg analysis framework for neuromarketing," *Brain Informatics*, vol. 9, 2022.

[24] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, pp. 427–437, 2009.

[25] A. Ben-David, "Comparison of classification accuracy using cohen's weighted kappa," *Expert Syst. Appl.*, vol. 34, pp. 825–832, 2008.

[26] J. H. Friedman, "Flexible metric nearest neighbor classification," 1994.

[27] M. Krauledat, M. Tangermann, B. Blankertz, and K.-R. Müller, "Towards zero training for brain-computer interfacing," *PLoS ONE*, vol. 3, 2008.

[28] B. Jiang, H. Chen, B. Yuan, and X. Yao, "Scalable graph-based semi-supervised learning through sparse bayesian model," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 12, pp. 2758–2771, 2017.

[29] K. Georgiadis, F. P. Kalaganis, V. P. Oikonomou, S. Nikolopoulos, N. A. Laskaris, and I. Kompatsiaris, "Harneshing the potential of eeg in neuromarketing with deep learning and riemannian geometry," in *Brain Informatics* (F. Liu, Y. Zhang, H. Kuai, E. P. Stephen, and H. Wang, eds.), (Cham), pp. 21–32, Springer Nature Switzerland, 2023.