2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

MarketBrain: An EEG based intelligent consumer preference prediction system

Fazla Rabbi Mashrur[1], Mohammad Tohidul Islam Miya[2], Ferdousi Sabera Rawnaque[3], Khandoker Mahmudur Rahman[2], Ravi Vaidyanathan[4], Syed Ferhat Anwar[5], Farhana Sarker[6], Khondaker A. Mamun[1]

***Abstract*— The traditional marketing research tools (Personal Depth Interview, Surveys, FGD, etc.) are cost-prohibitive and often criticized for not extracting true consumer preferences. Neuromarketing tools promise to overcome such limitations. In this study, we proposed a framework, MarketBrain, to predict consumer preferences. In our experiment, we administered marketing stimuli (five products with endorsements), collected EEG signals by EMOTIV EPOC+, and used signal processing and classification algorithms to develop the prediction system. Wavelet Packet Transform was used to extract frequency bands ($\delta, \theta, \alpha, \beta_1, \beta_2, \gamma$) and then statistical features were extracted for classification. Among the classifiers, Support Vector Machine (SVM) achieved the best accuracy (96.01±0.71) using 5-fold cross-validation. Results also suggested that specific target consumers and endorser appearance affect the prediction of the preference. So, it is evident that EEG-based neuromarketing tools can help brands and businesses effectively predict future consumer preferences. Hence, it will lead to the development of an intelligent market driving system for neuromarketing applications.**

I. INTRODUCTION

Marketing professionals' primary objective is to present their product in such a way that it elicits expected consumer response. To achieve this, companies adopt marketing techniques such as product promotion, celebrity endorsement, and various offers. To measure the effectiveness, they generally perform one-on-one interviews, general polls, and focus group discussion [1] to measure the consumer preferences towards these methods. However, even though, these techniques are straightforward, sometimes they are expensive to adopt [2]. In addition, they generate results that could contain biases, making them seem unreliable [1]. Therefore, there is a necessity for automatic prediction of consumer preference harnessing the power of technology. Hence neuromarketing is introduced, the fusion between neuroscience and conventional marketing.

*This research is supported by Institute of Advanced Research, United International University, Project Code No. IAR/01/19/SE/10.

[1]F. R. Mashrur and K. A. Mamun are with Advanced Intelligent Multidisciplinary Systems Lab, Institute of Advanced Research, United International University, Dhaka, Bangladesh

[2]M. T. I. Miya and K. M. Rahman are with School of Business and Economics, United International University, Dhaka, Bangladesh

[3]F. S. Rawnaque is with Graduate Program in Acoustics, The Pennsylvania State University, University Park, Pennsylvania, USA

[4]R. Vaidyanathan is with Department of Mechanical Engineering, Imperial College London, London, United Kingdom

[5]S. F. Anwar is with Institute of Business Administration, University of Dhaka, Dhaka, Bangladesh

[6]F. Sarker is with Department of Computer Science and Engineering, University of Liberal Arts Bangladesh, Dhaka, Bangladesh

Neuromarketing uses the brain's electrical activity, imaging, or other activity measurement technology to measure consumer response towards marketing stimuli. Literature suggests that the frontal lobe is mostly responsible for decision making and likeability of a product [2]. Moreover, among all the existing data acquisition methods to observe the human brain response, electroencephalogram (EEG) is the most cost-effective and portable with high temporal resolution. As a result, the use of EEG has recently increased in neuromarketing research.

Several experiments relating to neuromarketing using EEG were implemented to assess how advertising design can influence consumer decision-making and shopping behavior. Lee [3] showed that improved theta activation is correlated to the frontal brain area for the induction of empathy and increased buying rates. Research also performed automatic identification of preference prediction using machine learning. Yadava et al. [4] proposed a predictive machine learning model for 'likes' and 'dislikes' classification. While current research has shown the advantages of EEG signals for neuromarketing applications, there is still an absence of an automated framework for consumer preference prediction using marketing stimuli. In this work, we proposed a prediction algorithm that can identify consumer preference from EEG signals while administering marketing stimuli. We also performed extensive experiments on various classifiers to find the best classifier for the framework.

II. MATERIALS & METHODS

Figure 1 demonstrates the block diagram of the proposed framework. In this section, we discuss the participants, stimuli, and methodological description of our work.

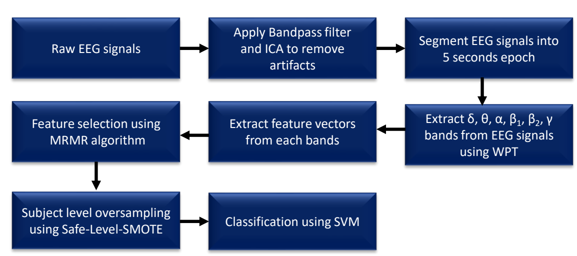

Fig. 1. Block diagram of the proposed MarketBrain

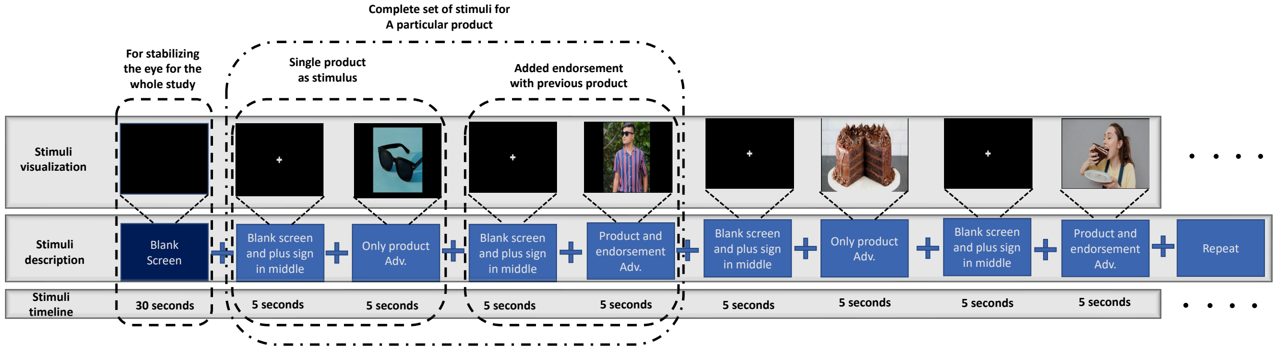

Fig. 2. Stimuli sequence and timing of presentation during electroencephalography data collection. Initially, a blank screen for visual stabilization for the participants. Then complete set of stimuli for a particular product (first a product followed by it's endorsement) appeared six times in random fashion throughout the experiments.

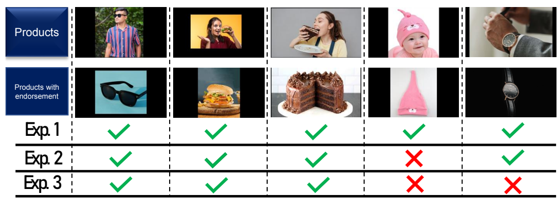

Fig. 3. All stimuli with corresponding experimental setup. In exp.1 all the products were used while in exp.2 (hat) and exp.3 (hat and watch) were excluded.

A. Participants

In this study, five healthy subjects (age: 20 ± 4 years, weight: 68 ± 12 kg) participated with no history of neurological disorder. Before enrollment, all participants gave their informed consent in accordance with the Helsinki Declaration and Neuromarketing Science and Business Association Code of Ethics.

B. Stimuli Description and Data Collection

In this study, we used five products with the corresponding endorsement with each product. An endorsement is a form of advertising which influences buyers positively towards the products. Usually, in real-life setup celebrities endorse a product. However, in our cause, we intentionally used neutral endorsement for avoiding biasing effect among the participants. In Fig 3, the products are in the first row, namely, sun glass, burger, cake, baby hat, and watch, and the endorsement of the products are in the second row. As our participants are young, a baby hat would not be appropriate with age structure. Again, for watch endorsement, the full endorser appearance is not visible which leads us to perform three different sets of experiments (exp.). In exp.1, we used all the products and products with endorsement. After that, in exp.2, we removed the baby hat and its endorsement from the analysis. Finally, in exp.3 we removed both the baby hat, watch, and their endorsement. Note that while collecting data, we used exp.1 stimuli (all the stimuli), and later while doing analysis we removed stimulus for exp.2 and exp.3.

We divided the data collection into three stages. In stage 1, the experimenter described the participants about the stimuli so that they are comfortable while watching those on screen. In stage 2, participants sat in front of a monitor that showed the stimuli with 75-100 cm distance. EEG signals (128 Hz) were collected from the participants using Emotiv Epoch+ headset while they were watching the stimuli. It is a wearable headset that contains 14 channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4) according to the 10–20 electrode system. For this work, we used eight frontal channels as they showed better performance in previous studies [3], [4]. Illustrated in Fig 2, each product has been shown followed by its endorsement for five seconds throughout the experiment. Note that, for each product, we randomized stimuli where one stimulus has been shown six times for each participant. In stage 3, we gave the participants a questionnaire where each stimulus was presented with a question "How much do you prefer (like) this product?". Subjects gave a rating on a Likert scale of 1 to 10 (low to high) that has later been converted to a binary form: low preference (1 to 5) and high preference (6 to 10).

C. Pre-processing of EEG signals

The collected EEG signals were pre-processed and analyzed using MATLAB 2020a (MathWorks, Natick, MA) software along with EEGLAB [5].
At first, We used a sixth-order bandpass Butterworth filter to extract the EEG signals between 0.5 and 48 Hz along with notch filter to remove 50. Afterward, independent component analysis was used to remove the contamination of eye artifacts, line noise, and movement artifacts.

D. Wavelet Packet Transform (WPT)

WPT has the ability to transforms a signal from a time domain into a time-scale domain [6]. It decomposes a signal into approximation and detail using highpass and lowpass filters which keep both time and frequency information. Let, $W_{m,n}(k), n = 0,..,2m-1$, denote the WPT coefficients at level i. We used these equations to calculate the WPT coefficients.

$$W_{m,2n}(k) = \sum_{l=0}^{L-1} h(l) W_{m-1,n} \left(2k+1-\mathrm{lmodN}_{n-1}\right) \quad (1)$$

$$W_{m,2n+1}(k) = \sum_{l=0}^{L-1} g(l) W_{m-1,n} \left(2k+1-\mathrm{lmod}\, N_{n-1}\right) \quad (2)$$

Where $k = 1 \ldots N$ and $N_n = N/2_n$. $g(l)$ and $h(l)$ are the impulse responses of highpass and lowpass filters of the wavelet packets respectively [6]. In this study, we used Meyer wavelet to compute the sub-bands as it has performed better in previous research [7] with EEG signals. In our study, we decomposed the signal in five levels and extracted various bands namely $\delta = 0-4\text{Hz}, \theta = 4-8\text{Hz}, \alpha = 8-12\text{Hz}, \beta_1 = 12-20\text{Hz}, \beta_2 = 20-32\text{Hz}, \gamma = 32-64\text{Hz}$.

E. Features Extraction

We extracted various sets of features from each band. Let, Y be the set of EEG signals from all the subjects, $Y = [y_1, y_2, ..., y_t]$, where t is the number of subjects. Again, y yields q number of distinct frequency bands, $y = [X_1(i), X_2(i), ..., X_q(i)]$ where i is the sample of EEG signals i=1,2,....,N

- Average power (ψ): $\psi = \frac{1}{N}\sum_{i=1}^{N}|X(i)|^2$
- Relative power (ζ): $\zeta = \frac{\psi}{\rho}$, where ρ is the total power of y_t. We also calculated all possible combination of distinct ratios ($\frac{\delta}{\theta}, \frac{\delta}{\alpha}, \frac{\delta}{\beta_1}, \frac{\delta}{\beta_1}, \frac{\delta}{\gamma}, \frac{\theta}{\alpha}, \frac{\theta}{\beta_1}, \frac{\theta}{\beta_2}, \frac{\theta}{\gamma}, \frac{\alpha}{\beta_1}, \frac{\alpha}{\beta_1}, \frac{\alpha}{\gamma}, \frac{\beta_1}{\beta_2}, \frac{\beta_1}{\gamma}, \frac{\beta_2}{\gamma}$) for both ψ and ζ.
- Arithmatic mean (μ): $\mu = \frac{1}{N}\sum_{i=1}^{N}X(i)$
- Modified mean absolute value (MMAV):
$$MMAV = \frac{1}{N}\sum_{i=1}^{N}w_i\left|X(i)\right|;$$
$$w_i = \begin{cases} 1, & \text{if } 0.25N \leqslant i \leqslant 0.75N \\ 4i/N, & \text{elseif } i < 0.25N \\ 4(i-N)/N, & \text{otherwise} \end{cases}$$
- Standard deviation (λ):
$$\lambda = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X(i) - \frac{1}{N}\sum_{i=1}^{N}X_i\right)^2}$$
- Skewness (S): $S = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i-\mu}{\lambda}\right)^3$
- Kurtosis (κ): $\kappa = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{X_i-\mu}{\lambda}\right)^4$
- Median (ν): $\nu = X(\frac{i}{2})^{th}$
- Teager–Kaiser energy operator (ξ) measures the instantaneous change of energy which is defined as:
$$\xi = \log\left(\sum_{i=1}^{N-1}X^2(i) - X(i-1)X(i+1)\right)$$
- Normalized 1st and 2nd difference (χ_1 and χ_2) is calculated by:
$$\chi_1 = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|X(i+1) - X_i|}{\lambda}$$
$$\chi_2 = \frac{\frac{1}{N-2}\sum_{i=1}^{N-2}|X(i+2) - X_i|}{\lambda}$$
- Temporal moment (TM) is calculated in 3rd order which is expressed by: $TM = \left|\frac{1}{N}\sum_{i=1}^{N}X(i)^3\right|$
- Shannon entropy (SE): $SE = -\sum_{i=1}^{N}\frac{X^2}{\sum_{i=1}^{N}X^2}\log_2\frac{X^2}{\sum_{i=1}^{N}X^2}$
- Threshold crosssing (TC) is the number of times that amplitude values cross zero amplitude from a thereshold value $T = 4 \times \frac{1}{10}\sum_{i=1}^{10}X(i)$ which iterated $N-1$ times. Each time the below condition satisfies, the value of TC increases by 1.
$$TC = \begin{cases} 1, & \text{if X}(i) > T \text{ and X}(i+1) < T \\ & \text{or X}(i)\text{i} < T \text{ and X}(i+1) > T, \\ 0, & \text{otherwise.} \end{cases}$$
- Slope sign change (SSC) is also similar to TC with different set of thereshold value = 0.01. The satisfying conditions are below:
$$SSC = \sum_{i=2}^{N-1}\left[f\left[\left(X(i) - X(i+1)\right) \times \left(X(i) - X(i+1)\right)\right]\right]$$
$$f(x) = \begin{cases} 1, & \text{if } x \geqslant \text{threshold} \\ 0, & \text{otherwise} \end{cases}$$

F. Feature Selection and Oversampling

We used the maximum relevance minimum redundancy algorithm [8] which ranks the features based on mutual information and correlation. Here, it gave 19 best features which were significantly different from others. We use these features for further classification. As we were performing binary classification, we oversampled these features using safe level SMOTE [9] on a subject basis. This is because there was a class imbalance on subject-level preference. We oversampled each class to 45 instances creating 90 samples for each subject.

G. Support Vector Machine (SVM) and other classifiers

SVM is a supervised classification algorithm, and can be used with different kernels (linear, polynomial, radial) [10], [11]. As it performs well with a small number of training instances and a huge number of features [11], we trained and tested SVM with RBF to classify between low and high preference towards the marketing stimuli. We also performed hyperparameter tuning sigma parameter and C parameter with a range of $1e-3$ to $1e3$ by Bayesian optimization in MATLAB [12]. In our experiment, SVM-RBF kernel was validated using 5-fold cross-validation for obtaining a robust estimation of the classification performance. The RBF kernel can be defined as:

$$K\left(x, x'\right) = \exp\left(-\gamma\left\|x - x'\right\|^2\right) \tag{3}$$

In this study, we used various classifiers namely, naive Bayes (NB), decision tree (DT), linear discriminant (LD), bootstrap aggregating (Bagging), logistic regression (LR), k-Nearest Neighbors (kNN), and AdaBoost [13].

TABLE I

PERFORMANCE OF THREE EXPERIMENTS USING SVM CLASSIFIER. BEST RESULTS ARE BOLDFACED FOR CORRESPONDING EXPERIMENT

Experiments	Acc.	Sen.	Spec.
Exp.1	89.90 ± 1.22	90.87 ± 1.25	88.74 ± 1.14
Exp.2	91.90 ± 0.89	92.75 ± 0.95	91.27 ± 0.81
Exp.3	**96.01±0.71**	**96.42±0.78**	**95.56±0.68**

III. RESULTS AND DISCUSSION

A. Performance of Support Vector Machine

From Table I, it is evident that exp.3 achieved the best result (boldfaced) with an accuracy of 96.01±0.71, sensitivity 96.42±0.78, and specificity 95.56±0.68 among the three experiments. It is also evident that exp.2 achieved better performance than exp.1. In exp.1, the performance was the lowest when all the stimuli were used. In exp.2, the performance improved after excluding the baby hat and its endorsement which were reported as irrelevant by our target consumers. In exp.3, the result improved on a much larger scale after excluding the watch and its endorsement which did not have a full endorser appearance. Herein the participants were limited in young adults, in future, diverse subject group should be added.

B. Performance for various classifiers

As *no free lunch theorem* [14] suggests no classifier will yield good result universally. This is because predictions might perform better in one domain while poor in others. For this, we tested various classifiers with the same validation scheme. As SVM achieved the best result with exp.3, we used 7 classifiers with selected features shown in Fig. 4. These classifiers were naive Bayes (NB), decision tree (DT), linear discriminant (LD), bootstrap aggregating (Bagging), logistic regression (LR), k-Nearest Neighbors (kNN), and AdaBoost. For each classifier, we used the bayesian optimization set by MATLAB for finding the best model tuning the hyperparameters [12]. Note that, we performed 5-fold cross-validation and iterate 20 times for each classifier to understand the stability of the model. We reported the accuracies in Fig. 4 to observe the performance variation where the linear classifiers performance were poor compared to the non-linear classifiers. NB, DT, LD, LR achieved an accuracy of less than 85% while AdaBoost, bagging achieved better performance. However, in the case of AdaBoost, the range of accuracies varied the most for its sensitivity to noisy data and outliers. KNN performs amongst the non-linear classifiers while SVM outperforms all others for having the hyper-plane feature alongside the RBF kernel.

IV. CONCLUSION AND FUTURE WORK

In this study, we proposed a framework, MarketBrain, to predict consumer preference in marketing stimuli from EEG signals using machine learning algorithms. The results suggest that the proposed approach is effective and complement the conventional methods to assess the expected customer response to marketing stimuli. We also found that age group

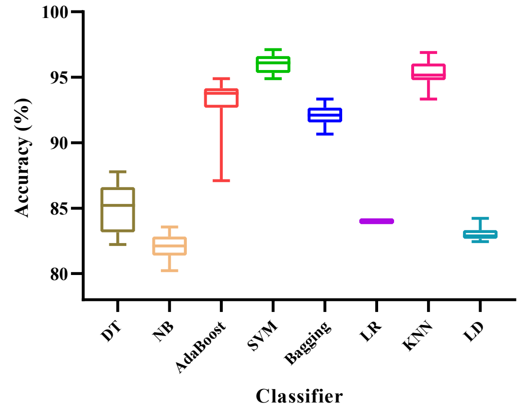

Fig. 4. Performance of the proposed framework for various classifiers shown in box plot. SVM performed better than other classifiers

of consumer and endorser appearance affect in predicting preference regarding target marketing. However, our limitation was having 5 participants which can be increased in future studies. Finally, it is evident that the neuromarketing tool is effective to predict future consumer preferences and help brands to forecast consumer behaviors.

REFERENCES

[1] J. Hulland, H. Baumgartner, and K. M. Smith, "Marketing survey research best practices: evidence and recommendations from a review of jams articles," *Journal of the Academy of Marketing Science*, vol. 46, no. 1, pp. 92–108, 2018.

[2] F. S. Rawnaque et al., "Technological advancements and opportunities in neuromarketing: a systematic review," *Brain Informatics*, vol. 7, no. 1, pp. 1–19, 2020.

[3] E.-J. Lee, "Empathy can increase customer equity related to pro-social brands," *Journal of Business Research*, vol. 69, no. 9, pp. 3748–3754, 2016.

[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, no. 18, pp. 19 087–19 111, 2017.

[5] A. Delorme and S. Makeig, "Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of neuroscience methods*, vol. 134, no. 1, pp. 9–21, 2004.

[6] D. B. Percival and A. T. Walden, *Wavelet methods for time series analysis*. Cambridge university press, 2000, vol. 4.

[7] K. A. e. a. Mamun, "Decoding movement and laterality from local field potentials in the subthalamic nucleus," in *2011 5th International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2011, pp. 128–131.

[8] M. Radovic et al, "Minimum redundancy maximum relevance feature selection approach for temporal gene expression data," *BMC bioinformatics*, vol. 18, no. 1, pp. 1–14, 2017.

[9] C. Bunkhumpornpat, K. Sinapiromsaran, and C. Lursinsap, "Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem," in *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 2009, pp. 475–482.

[10] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[11] V. Vapnik and R. Izmailov, "Knowledge transfer in svm and neural networks," *Annals of Mathematics and Artificial Intelligence*, vol. 81, no. 1, pp. 3–19, 2017.

[12] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *arXiv preprint arXiv:1206.2944*, 2012.

[13] G. Bonaccorso, *Machine learning algorithms*. Packt Publishing Ltd, 2017.

[14] D. H. Wolpert, "The lack of a priori distinctions between learning algorithms," *Neural computation*, vol. 8, no. 7, pp. 1341–1390, 1996.