CONSUMER PREFERENCE DETECTION IN NEUROMARKETING USING EEG

A DISSERTATION

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF DEGREE

OF

MASTER OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

Submitted by:

Aditya Gupta

2K20/CSE/02

Under the supervision of

Dr. Divyashikha Sethia

(Assistant Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

May, 2022

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Aditya Gupta, Roll no 2K20/CSE/02, student of M.Tech (Computer Science and Engineering), do hereby declare that the project dissertation titled "Consumer Preference Detection in Neuromarketing Using EEG" which is submitted by me to the Department of Computer Science and Engineering, Delhi Technological University, Delhi, in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without prior citation. This work has not previously formed the basis for the award of and Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Aditya Gupta

Place: Delhi — Aditya Gupta

Date: 31/05/2022 — 2K20/CSE/02

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi – 110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "Consumer Preference Detection in Neuromarketing Using EEG" which is submitted by Aditya Gupta, 2K20/CSE/02, Department of Computer Science & Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge, this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Vinod Kumar

HoD CSE

Place: Delhi for Dr. Divyashikha Sethia

Date: Assistant Professor

Department of CSE

ACKNOWLEDGMENT

This project was an endeavour taken up by not just me, but also required the input of several other people. I have no words to express my gratitude for everyone who contributed to this project. Most of all, my sincere gratitude to Dr. Divyashikha Sethia, my research mentor, who continuously pushed me to achieve better and more rigorous results, and gave me the opportunity and guidance needed to pursue this project. I would also like to thank Chirag Ahuja, for helping me understand signal processing and machine learning in depth, without which I would not be able to deliver the results reported in this dissertation. I am highly indebted to the panel faculties during all the progress evaluations for their guidance, constant supervision and for motivating me to complete my work. I would also like to express my gratitude for my batchmates and my family, who motivated me at every step and made sure I did not leave any stone unturned in the process of completing this dissertation.

This has been an ecstatic learning experience, and I would like to express my gratitude to Delhi Technological University, which instilled in me the learnings and confidence needed to complete my M.Tech degree with flying colours.

Aditya Gupta

2K20/CSE/02

ABSTRACT

Electroencephalography (EEG) has made several strides in neuroscience. It is generally used to monitor, diagnose, and identify several neurological conditions. The advantage of EEG is that it can record brain waves with high resolution with an extremely ergonomic setup. Hence, it has become a favoured choice for multiple applications like detection of dementia, epilepsy, classification of motor imagery, neuromarketing, measuring cognitive attention, etc.

This dissertation focuses on the application of neuromarketing, which is a fusion of neuroscience and marketing. Debatably unethical, neuromarketing has quickly gained traction in industry after several experiments found success. An example is Frito-Lay, who used neuroimaging to entirely re-evaluate their approach to marketing. Customers were shown products with different packaging and colours, and their responses were recorded as positive, negative, or neutral. It was found that shiny packaging was not preferred, and matte was. Frito-Lay went on to scrap shiny packaging and adopted the matte look.

The publicly available dataset recorded by Yadava et al [1] has been used. Independent component analysis (ICA), empirical mode decomposition (EMD) and logistic regression were subsequently applied on the raw data, and the best f1 score is reported as 89.41%, which is superior to the method used by Yadava et al [1], who achieved an accuracy of 70%.

CONTENTS

Candidate's Declaration	i
Certificate	ii
Acknowledgement	iii
Abstract	iv
Contents	v
List of Figures	vii
List of Tables	ix
List of Abbreviations	x
CHAPTER 1 INTRODUCTION	1
1.1 Electroencephalogram	1
1.2 Neuroscience	1
1.3 Neuromarketing	2
CHAPTER 2 RELATED WORK	4
CHAPTER 3 TECHNIQUES USED	6
3.1 Bandpass Filter	6

3.2 Independent Component Analysis 8

3.3 Empirical Mode Decomposition 9

3.4 Logistic Regression 11

CHAPTER 4 PROPOSED WORK 13

4.1 Problem Statement 13

4.2 Proposed Methodology 13

CHAPTER 5 OBSERVATIONS AND ANALYSIS 16

CHAPTER 6 RESULTS 20

CHAPTER 7 CONCLUSION AND FUTURE SCOPE 22

REFERENCES 23

LIST OF FIGURES

3.1 Working of bandpass filter	6
3.2 Magnitude transfer function vs frequency for bandpass filter	7
3.3 Frequency response of Butterworth filter	7
3.4 Result of passing signal through Butterworth filter	8
3.5 Behaviour of signals after ICA	9
3.6 Flowchart of EMD algorithm	11
4.1 10-20 International System	14
4.2 Flowchart of proposed methodology	15
5.1 Amplitude vs Time Raw signal Channel AF3	16
5.2 Amplitude vs Time Raw signal Channel F7	16
5.3 Amplitude vs Time Filtered signal Channel AF3	17
5.4 Amplitude vs Time Filtered signal Channel F7	17
5.5 Amplitude vs Time Sub-bands Channel AF3	17
5.6 Amplitude vs Time Sub-bands Channel F7	17
5.7 Amplitude vs Time IMF 0 delta sub-band Channel AF3	18
5.8 Amplitude vs Time IMF 1 delta sub-band Channel AF3	18
5.9 Amplitude vs Time IMF 0 theta sub-band Channel AF3	18
5.10 Amplitude vs Time IMF 1 theta sub-band Channel AF3	18
5.11 Amplitude vs Time IMF 0 alpha sub-band Channel AF3	18

5.12 Amplitude vs Time IMF 1 alpha sub-band Channel AF3 18

5.13 Amplitude vs Time IMF 0 beta sub-band Channel AF3 18

5.14 Amplitude vs Time IMF 1 beta sub-band Channel AF3 18

5.15 Amplitude vs Time IMF 0 gamma sub-band Channel AF3 19

5.16 Amplitude vs Time IMF 1 gamma sub-band Channel AF3 19

LIST OF TABLES

6.1 F1 scores for Sub band IMF x Channel 20

6.2 Comparison of result with Yadava's work 21

LIST OF ABBREVIATIONS

1. EEG: Electroencephalogram
2. ICA: Independent Component Analysis
3. EMD: Empirical Mode Decomposition
4. ML: Machine Learning
5. IMF: Intrinsic Mode Function

CHAPTER 1

INTRODUCTION

1.1 ELECTROENCEPHALOGRAM

Our brain is composed of cells called neurons. They communicate via electrical impulses, and are active at all stages of a healthy person's life, whether conscious or unconscious. An electroencephalogram (EEG) is is a technique that uses tiny circular discs called electrodes to record electrical activity on the scalp, and this shows up as near sinusoidal wavy lines on a graph. They were recorded on paper with ink initially, but have progressed to become completely digital. Various devices such as Emotiv Epoc+ (which was also used by Yadava et al [1] to record their dataset) are now able to record EEG with high resolution and a low cost, ergonomic setup.

While complex cognitive phenomena like motions cannot be captured, simple phenomena like arousal and valence can be, and this allows mapping of stimuli of interest to observed patterns in brain waves. Several studies have used recorded EEG to monitor and identify cognitive conditions such as epilepsy [2], sleep disorders [3], etc with remarkable success. The viability of EEG for cognitive studies has hence been confirmed.

1.2 NEUROSCIENCE

Cognition in human beings has been studied in great depth [4]. The authors of [5] found that roughly 90% of data processing in the brain is subconscious.

A great interest has developed in demystifying the secrets of the subconscious brain, and various technologies have emerged, that enable us to pursue this knowledge. This pursuit of knowledge has culminated in a field called neuroscience, which is a multidisciplinary field spanning physiology, molecular biology, physics, computer science, chemistry and mathematical modelling. Over time, the scope has broadened, and the field has proliferated quickly to branch out into several other subfields.

1.3 NEUROMARKETING

Combining neuroscience and marketing, the field of neuromarketing is based on the idea that cognitive activity can be used to predict consumer behaviour [6]. Since most decisions an individual makes are subconscious [7], if these decisions can be mapped to biofeedback signals, through biometrics like eye movement tracking, facial encoding, galvanic skin response and EEG, insight can be gained into the emotional and psychological state of a consumer. In particular, EEG measures cerebral cortical activity. This data is superficial and deep brain data is not available, so complex cognitive phenomena cannot be detected, like emotions. However, the high temporal resolution of EEG signals allows detection of stimuli of interest, like arousal [8].

Neuromarketing equips companies with the tools to gauge consumer preference more tangibly, and not only market, but also develop products more compatible with their target audience [9]. Conventional advertising campaigns on the other hand are expensive and do not clearly quantify user engagement holistically. Companies often use focus groups during market research; however, they are self-reported, and could be inaccurate. As a result, neuromarketing has quickly gained traction, and found success in multiple scenarios. Frito-Lay, for instance, discovered matte bags were more visually appealing to consumers than shiny bags, and within months, they scrapped shiny bags, and designed new bags. In another case, Hyundai took a sample space of thirty participants, fitted with EEG caps and presented them the prototype of a car for an hour, and detected their response through EEG and questionnaire. Another instance of successful neuromarketing is an experiment by PayPal. They found out that advertisements that were based on speed and convenience elicited a considerably superior response than those centred around safeness and secureness. They went on to redesign their ad campaign based on the results.

CHAPTER 2

RELATED WORK

Lee et al [10] define neuromarketing as using neuroscientific methods to perform analysis of, and gain insight into behaviour of the population with respect to commercial trends and institutions. A common argument against neuromarketing is that it reduces a consumer's ability to consciously reject a product, which leaves consumers vulnerable, and they can be easily taken advantage of by a company's campaign [11]. However, according to Lee et al [10], this view of neuromarketing as not ethical, riddled with flaws, and potentially harmful, does not apply to research performed for scholarly purposes. Instead, it is viable to think of neuromarketing as an extremely useful and significant area for research in the future, since it has the capability to allow us to identify important patterns in human behaviour, and help us gain insight into problems which have previously remained elusive. According to Hubert et al [12] individuals easily cave to peer pressure, and in pursuit of social acceptance, often lie, or provide false or partially correct information. Thus, genuineness is absent from the recorded responses, and are filtered by the interviewee's consciousness before being reported. These obstacles can be overcome through neuromarketing, since control over the data collected is extremely limited for research participants [12, 13].

The effect of price on consumers has been researched and investigated in great detail [13]. Vecchiato et al [14] illustrated how, through brain imaging, cerebral activity of an individual can be tracked, and subsequently, the emotional response elicited by different parts of an advertisement can be investigated, and their impact on cognitive processing can be gauged. Golnar-

Nik et al [15] extracted power spectral density features to emphasize separability between preferred or not preferred products as perceived by the consumer with significantly high accuracy (87%). Their results indicated that the Centroparietal regions and frontal regions are particularly interesting brain locations for identifying separability of a consumer's resting state versus an active cognitive state, i.e., for labelling a given product. The authors of [16] used several differentiation measures, namely, precision and recall, and accuracy. They also validated their results using LOOCV, holdout, and k-fold cross validation, to test performance of the classifiers they had developed, namely DNN, RF and KNN. The accuracies of these classifiers are 94%, 92% and 88% respectively. They show that RF is capable of achieving competitive results with DNN, even though KNN and SVM underperform when compared to DNN in terms of the exhibited measures of accuracy, recall, and precision.

Oon et al [17] classified products after clubbing them into certain categories, and subsequently found the characteristics of EEG signals for the "most preferable" products. They used DFA features as input of NN and k-NN classifier and got highest classification accuracy for fast food. Alpha waves recorded 80% accuracy, while the accuracy was 76% for beta and 72% for the combined spectrum of alpha and beta spectral bands.

Golnar et al [15] studied the effect of background colour on the costumers' preferences and showed that not only can it impact consumer preference (liking), but also that change of EEG power, particularly in electrode positions Cp3, Cpz, and Fp1, which are essentially the centroparietal and frontal regions of the brain, can be used as a good candidate for predicting subject decision-making. Yadava et al [1] collected data from 40 participants, after showing them a selection of 42 products. The like/dislike for each product was collected in a self-reported questionnaire. They cleaned the raw signal with Savitzky Golay filter, extracted features using discrete wavelet transform and trained a Hidden Markov Model (HMM) with the features, to achieve 70% classification accuracy for the 25 male participants in the study, and 63% accuracy for the 15 female participants.

CHAPTER 3

TECHNIQUES USED

3.1 BANDPASS FILTER

A bandpass filter is a device or construct that allows only certain frequencies within a range to pass through it. It is perhaps the most ubiquitously used signal processing technique, owing to low frequency noise components and line noise being present almost universally in all available raw data. It contrasts with low-pass and high-pass filters in that it blocks frequencies both above and below given thresholds. Fig 1 shows how a bandpass filter works.

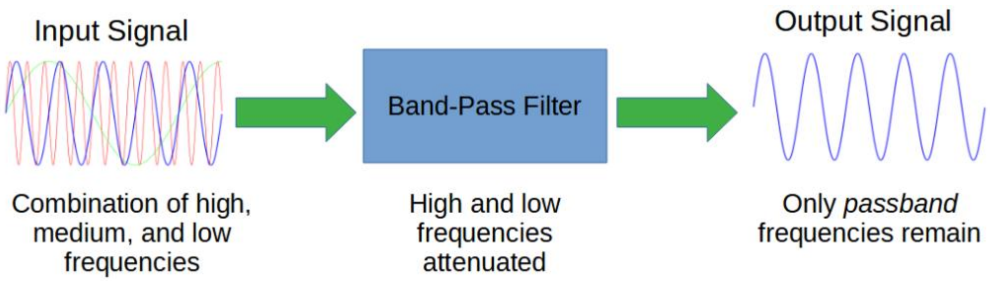

Fig 3.1

An ideal bandpass filter has a flat passband, i.e., all frequencies above or below the given threshold are attenuated, or flattened out, completely. However, this is not possible in practise. Fig 3.2 shows the magnitude transfer function vs frequency for a bandpass filter.

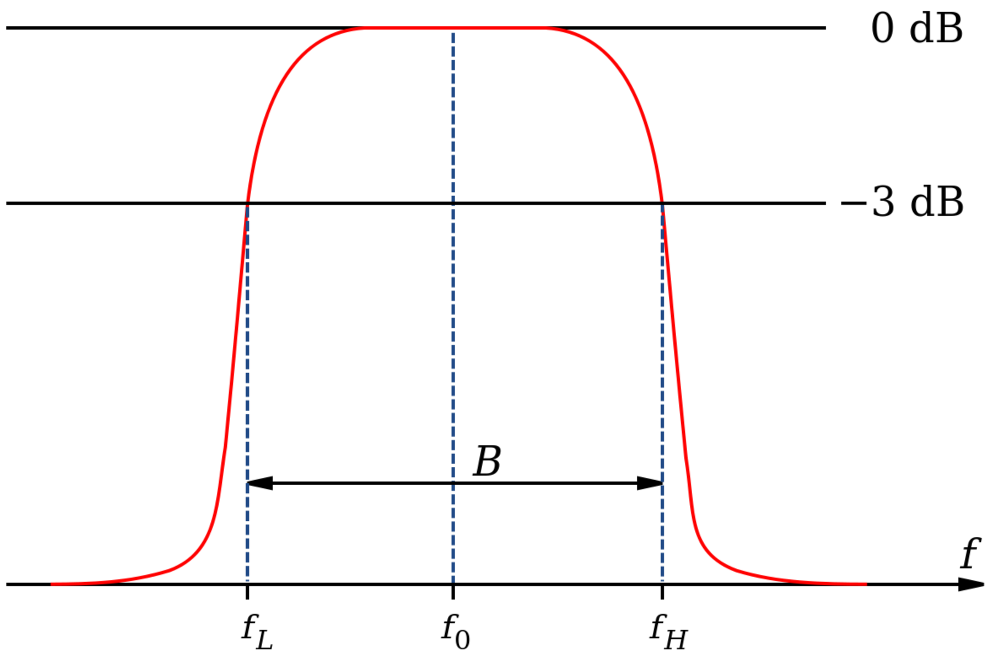

Fig 3.2

The most commonly used bandpass filter in signal processing is the Butterworth filter [18]. It is an attempt at approximating the ideal “brick wall” filter, i.e., the frequency response is as flat as possible. Fig 3.3 shows how the frequency response of a Butterworth filter changes with respect to order.

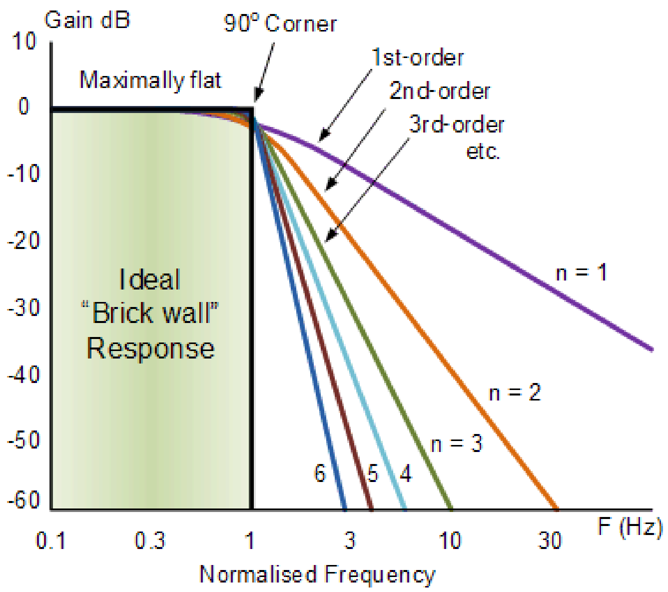

Fig 3.3

Fig 3.4 shows the result of an input signal passed through a bandpass filter.

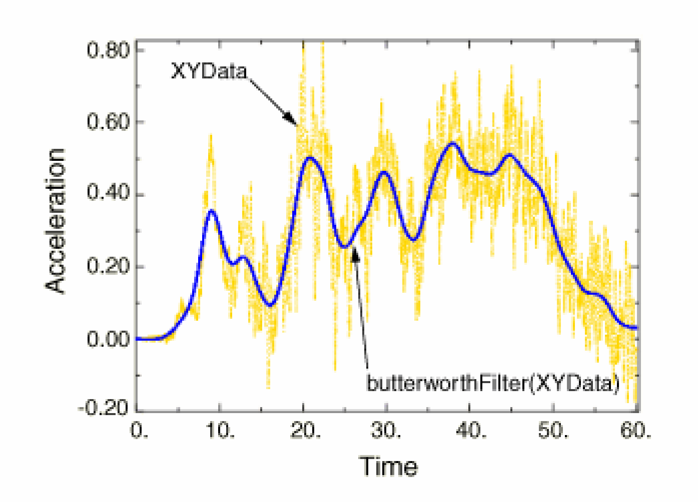

Fig 3.4

3.2 INDEPENDENT COMPONENT ANALYSIS

Independent Component Analysis (ICA) [19] is an algorithm used to separate an analog or digital signal into its additive subcomponents. It is a blind source separation technique, and assumes that the subcomponents are statistically independent from each other. This is usually true in case of EEG signals, since arousal/valence EEG is statistically independent from muscular artifacts such as eye blinks, or motor imagery. As a result, ICA has been adopted in EEG signal processing in several studies.

The mathematical basis behind ICA is described as follows. Let the measured data be

$x(t) = [x_1, x_2, \ldots x_n]^T$

where the horizontal dimension of x represent a single electrode's signal, and the vertical dimension represent signals of all electrodes at an instantaneous point of time. Let motor imagery-related brain sources be

$s(t) = [s_1(t), s_1(t), \ldots s_N(t)]^T$

The relationship between them and observed EEG signals on the scalp $x(t)$ can be expressed as follows

$$x(t) = As(t)x(t) = As(t) \tag{1}$$

where

$A = [a_1, . . . , a_N]$

is a matric with a constant coefficient and its dimesion is NxN. It can also be termed as the mixing matrix. Each column vector

$aj \ (j = 1, . . . , N)$

of A is a pattern in space that represent the independent components of the overall electrodes attached to the scalp and the values are projection weights of each electrodes. To achieve a discriminability between observed signals *x(t)*, the inverse matrix of A, i.e., the unmixing matrix W, needs to be manipulated to obtain N estimated source signals

$u(t) = u_1(t),....u_N(t)T$

as independently as possible, giving:

$$W = A^{-1}W = A^{-1} \qquad (2)$$

and the estimated source signal is obtained as

u(t)=Wx(t)=s(t) [20]

Ideally, ICA decomposes a multivariate signal into independent non-Gaussian subcomponents. Fig 3.5 shows how multivariate signals behave when passed through an ICA filter.

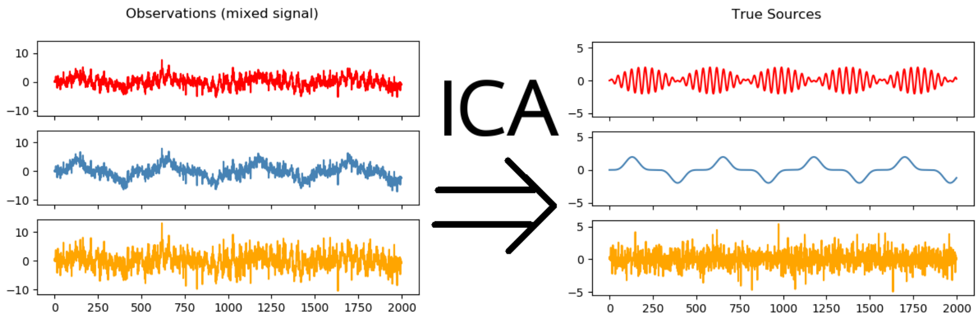

Fig 3.5

3.3 EMPIRICAL MODE DECOMPOSITION

Empirical Mode Decomposition (EMD) [21] is a technique used in signal processing that allows isolation of instantaneous frequency in a signal. EMD decomposes a given signal into Intrinsic Mode Functions (IMF). The basis formed by these IMFs must be complete and nearly orthogonal basis. IMFs must satisfy 2 conditions:

1) Difference of the number of extrema and the number of zero-crossings in the entire signal must be zero or one.
2) Say there is an envelope defined by local maxima, and another defined by the local minima. The mean value of this envelope must be zero at any given point.

Extracting an IMF is called sifting. The local extrema are identified in the dataset, the local maxima are connected using a cubic spline line as upper envelope, and the local minima are connected using a cubic spline line as the lower envelope.

The mean of the data is m_1, and the difference between the data and the mean is the first component, h_1.

$X(t) – m_1 = h_1$

After sifting for one iteration, h_1 satisfies the definition of an IMF, and another crest becomes a local maximum. Treating h_1 as data for subsequent iterations,

$h_1 – m_1 = h_{11}$

After sifting for k times, h_1 becomes an IMF, and the first IMF component of the data.

$c_1 = h_{1k} = h_{1(k-1)} – m_{1k}$

Fig 3.6 shows the flowchart for EMD algorithm

Since EMD is designed keeping in mind nonstationary and nonlinear data, it has found extensive use in EEG signal processing.

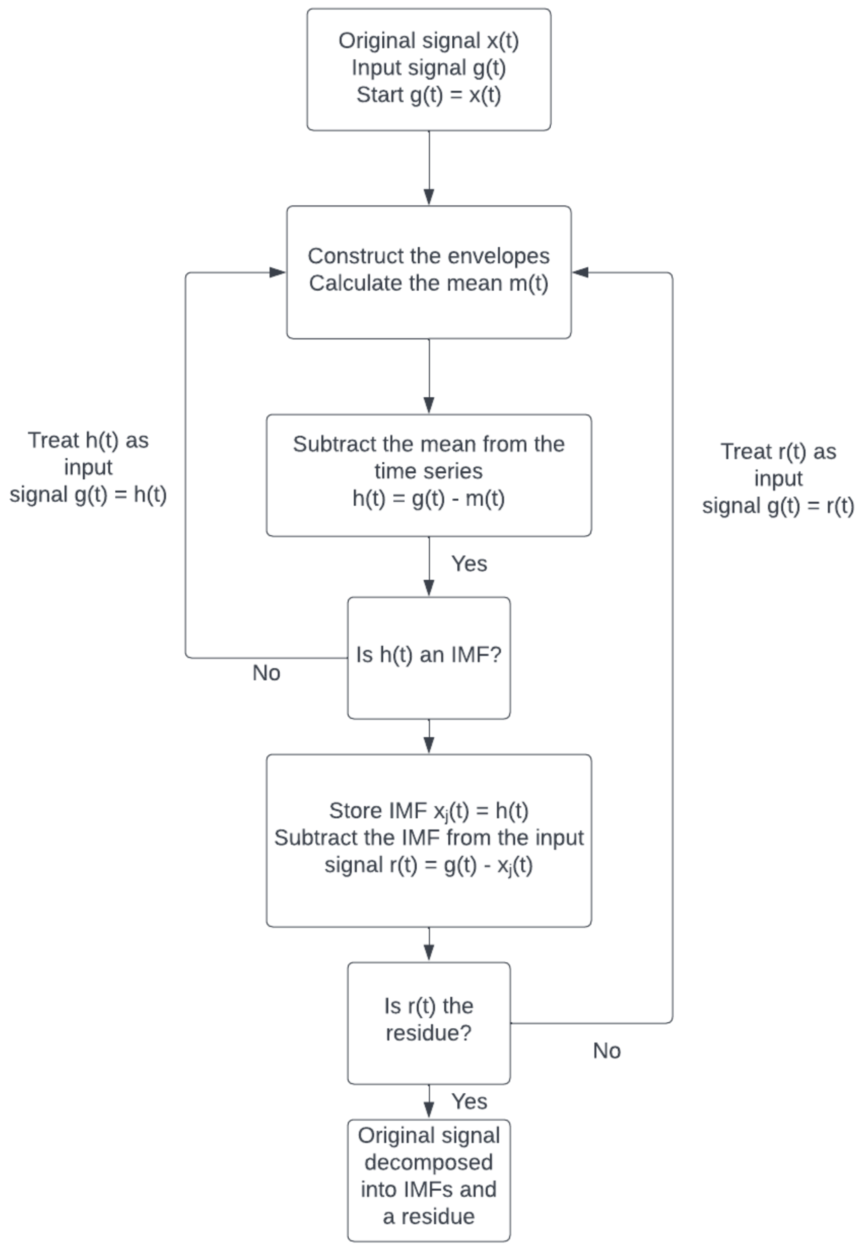

Fig 3.6

3.4 LOGISTIC REGRESSION

Logistic regression [22] is a popular Machine Learning (ML) algorithm and is a supervised learning technique. It attempts to predict a dependent variable using a set of independent variables. In EEG, the dependent variable would be the label, and independent variables are various source signals associated with the particular data label.

Mathematically, let the equation of a straight line be

$$y = b_0 + b_1x_1 + b_2x_2 ...+ b_nx_n$$

In Logistic regression, y can only be between 0 and 1, so we divide the equation by *(1-y)*

y/1-y; 0 for y = 0, and infinity for y = 1

Taking the logarithm of the equation, we get the final equation for logistic regression as

$$\log [y/(1-y)] = b_0 + b_1x_1 + b_2x_2 ...+ b_nx_n$$

It is very similar to linear regression; however, logistic regression is used expressly to solve classification problems.

CHAPTER 4

PROPOSED WORK

4.1 PROBLEM STATEMENT

Manufacturers spend a lot of resources in identifying and gauging consumer interest and preference, to better design products aligned with consumer needs. However, self-reported questionnaires are often misleading and biased. Neuromarketing allows this gap to be bridged, and research along the direction pursued in this thesis, and related literature, can benefit industry greatly.

After perusing the available literature, a research gap was identified. Blind source separation is a commonly used tool for signal processing in EEG analysis, and for neuromarketing in particular, ICA has been found to be greatly favoured. However, Empirical Mode Decomposition (EMD) is another blind source separation technique, that separates a signal's components into an orthogonal basis. The hypothesis was that EMD can be used in conjunction with ICA to offer better discriminability between source signal and noise in EEG signals for detecting arousal, and the hypothesis has been validated through the results achieved in this paper.

4.2 PROPOSED METHODOLOGY

The publicly available dataset recorded by Yadava et al [1] was used. Yadava et al [1] captured data from 40 participants using an Emotiv EPOC+ device. This device has 14 EEG channels at AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 as per International 10-20 system. Fig 4.1 shows a diagram of international 10-20 system. The sampling frequency is 128Hz per channel. Items were displayed to users and EEG signals were captured simultaneously, following which they were presented a questionnaire to record their like or dislike for the product.

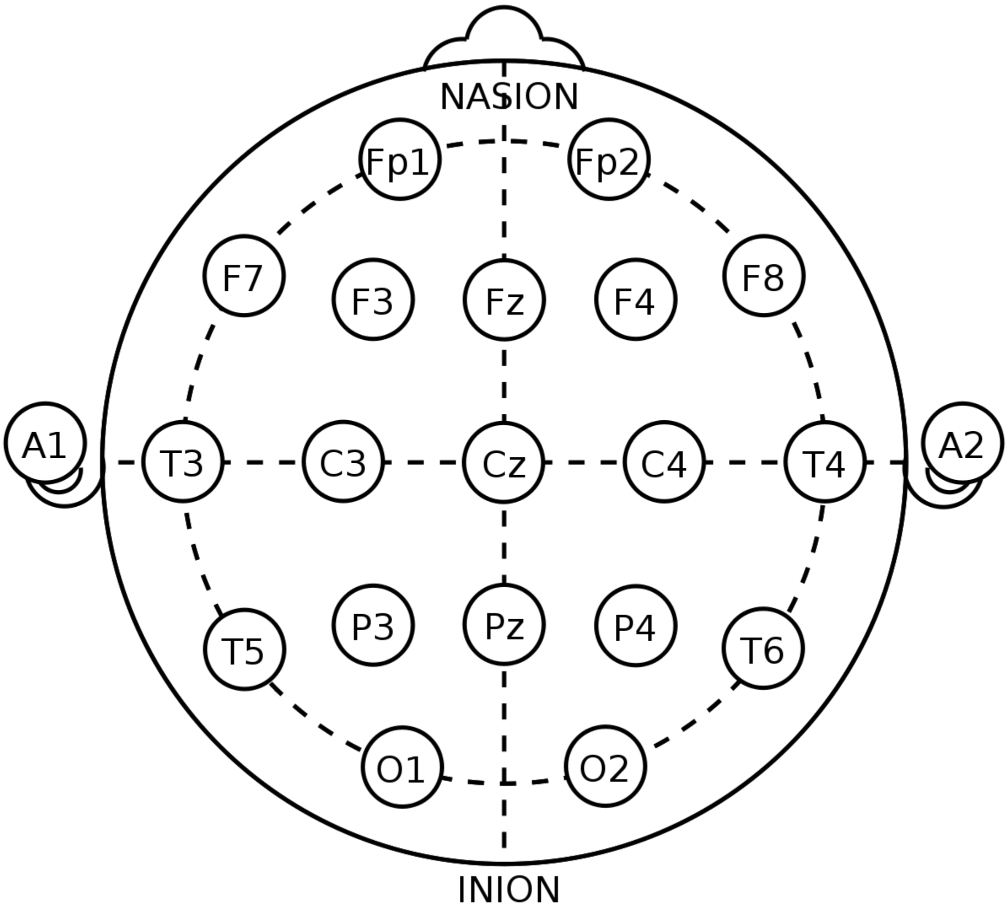

Fig 4.1

The raw EEG signals for each subject were passed through a bandpass filter due to evident low frequency noise component. The resulting signals are assumed to have statistically independent non-Gaussian subcomponents. ICA was applied on the filtered signal to obtain source signals. These source signals were split into energy bands, namely delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-35 Hz) and gamma (35-49 Hz), and each signal associated with the band was decomposed using EMD. Then features were extracted from each IMF for each channel. The list of features is as follows:

1) Statistical Features (mean, median, standard deviation, skew, kurtosis, max, min)
2) Hjorth features (activity, complexity and mobility)
3) Spectral Power

This resulted in 14-channels x 5-bands x total-IMFs data frames with the features calculated as the columns. A logistic regression model was trained using each data frame separately, and f1 score was calculated as the metric. Fig 4.2 shows a flowchart of the proposed methodology.

Fig 4.2

CHAPTER 5

OBSERVATIONS AND ANALYSIS

This section reports the observations made as the proposed methodology was carried out.

Fig 5.1 and 5.2 show the Amplitude vs Time plot of raw signal for Channel AF3 and F7 respectively. It is immediately clear that there is a low frequency noise component present.

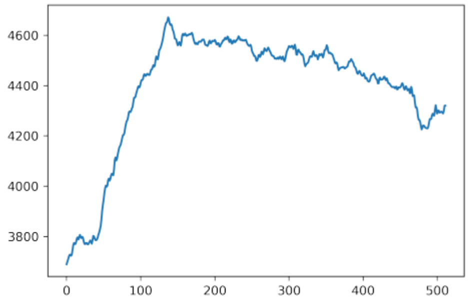

Fig 5.1

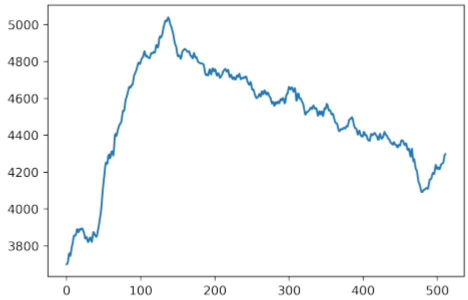

Fig 5.2

To remove this noise component, the raw signals are passed through a 6^{th} order Butterworth bandpass filter of 0.5-49 Hz, followed by ICA. Fig 5.3 and 5.4 show the Amplitude vs Time plot of filtered signal for Channel AF3 and F7 respectively. These plots show that there is still a high frequency noise component present in the filtered signal.

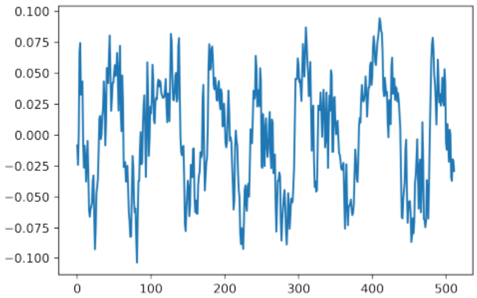

Fig 5.3

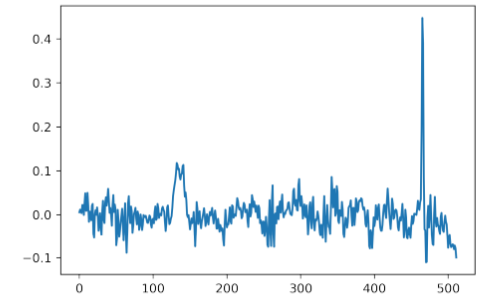

Fig 5.4

For better separability and discriminability, the filtered signals are split into sub-bands of delta, theta, alpha, beta, and gamma. Fig 5.5 and 5.6 show the Amplitude vs Time plot of sub-bands for Channel AF3 and F7 respectively. The subplots are in the same order from top to bottom. It can be observed that delta, theta, and gamma have high noise components, but alpha and beta bands appear closer to clean signals.

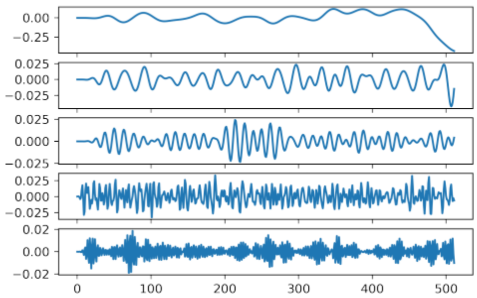

Fig 5.5

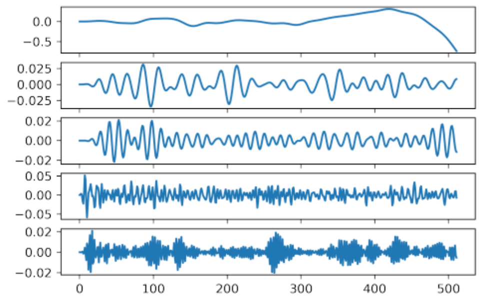

Fig 5.6

Using EMD, each of these sub-band signals are decomposed into their respective IMFs. Fig 5.7 to 5.16 show the Amplitude vs Time plot of IMF 0 and IMF 1 of Channel AF3 respectively.

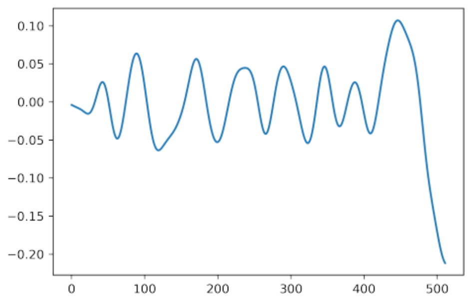

Fig 5.7 (AF3 IMF 0 delta sub band)

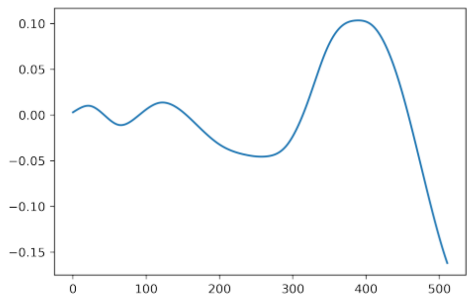

Fig 5.8 (AF3 IMF 1 delta sub band)

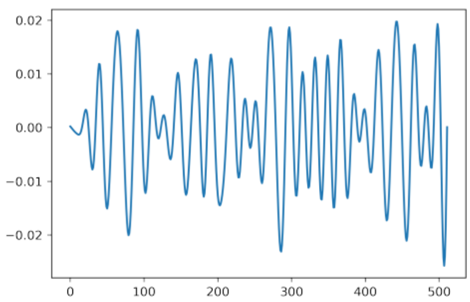

Fig 5.9 (AF3 IMF 0 theta sub band)

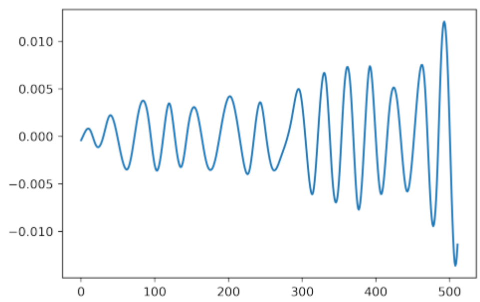

Fig 5.10 (AF3 IMF 1 theta sub band)

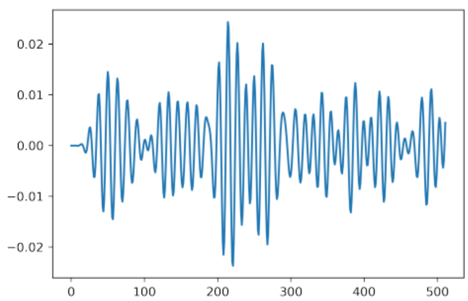

Fig 5.11 (AF3 IMF 0 alpha sub band)

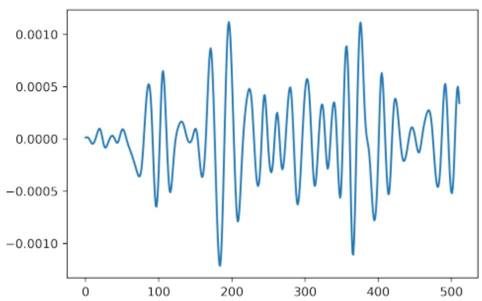

Fig 5.12 (AF3 IMF 1 alpha sub band)

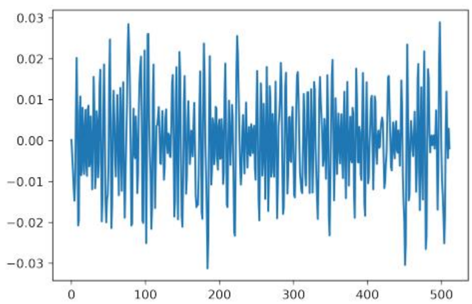

Fig 5.13 (AF3 IMF 0 beta sub band)

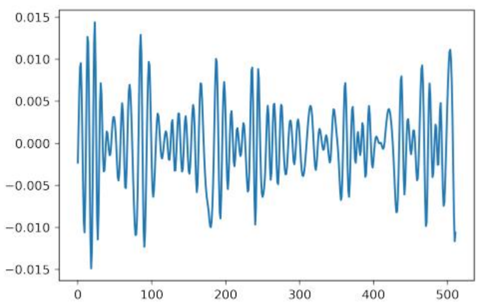

Fig 5.14 (AF3 IMF 1 beta sub band)

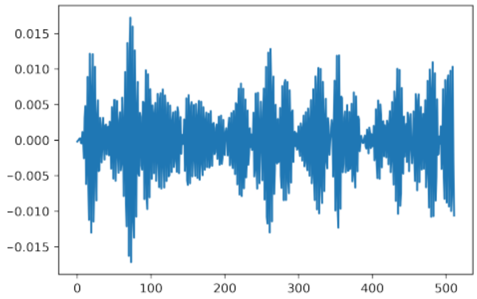

Fig 5.15 (AF3 IMF 0 gamma sub band)

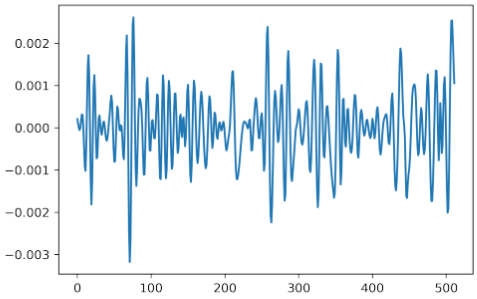

Fig 5.16 (AF3 IMF 1 gamma sub band)

The signals in figures of sub-bands look relatively clean, so features were extracted from them and a logistic regression model was trained using the feature matrix calculated.

CHAPTER 6

RESULTS

The results obtained have been tabulated in Table 6.1. The best result is reported for IMF 0 for Alpha band of channel O2. This is a channel related to the occipital lobe, and it is a reasonable result because product images would cause ocular stimulus.

It can also be observed that the first IMF of Alpha and Beta bands give better accuracy for the dataset used. Occipital, parietal and frontal lobes particularly have higher accuracies. This indicates that consumer choice is a largely active cognitive process, but also affected by subconscious decision making.

	AF3	F7	F3	FC5	T7	P7	O1	O2	P8	T8	FC6	F4	F8	AF4
Delta IMF 0	56.12	35.75	42.01	39.64	44.25	51.63	59.56	21.25	53.85	48.67	59.34	41.53	50.94	56.0
Theta IMF 0	65.81	64.97	60.84	65.81	63.16	66.45	64.95	68.75	63.72	64.1	64.29	66.22	63.46	69.68
Theta IMF 1	51.34	46.35	51.03	49.21	51.13	46.47	52.46	48.41	47.06	38.77	54.15	40.93	45.34	46.09
Alpha IMF 0	85.71	84.56	88.37	86.47	87.12	84.87	86.79	89.41	86.47	85.82	85.07	85.93	87.45	85.19
Alpha IMF 1	46.09	37.56	42.98	46.28	42.45	54.84	49.6	46.03	47.86	42.11	48.53	45.3	43.84	52.67
Alpha IMF 2	56.88	57.94	40.85	55.26	47.62	45.71	48.13	44.44	53.05	43.72	49.12	56.8	53.77	41.88
Beta IMF 0	80.57	86.69	86.59	85.27	87.12	83.27	88.12	88.72	87.79	84.13	85.49	86.92	86.26	82.18
Beta IMF 1	46.15	50.92	52.08	39.84	46.22	45.71	50.81	40.85	48.91	45.9	45.6	44.98	47.77	51.18
Beta IMF 2	50.0	49.81	42.98	34.91	46.64	42.74	41.35	46.83	55.78	45.9	34.41	45.13	46.86	44.72
Gamma IMF 0	70.15	71.57	73.02	70.37	70.12	71.65	71.65	71.07	70.99	72.56	68.28	71.43	72.33	71.65
Gamma IMF 1	51.19	51.28	42.42	53.28	31.16	41.15	46.27	48.09	45.69	51.16	51.15	50.0	49.37	46.34
Gamma IMF 2	50.79	55.48	44.96	43.2	56.92	54.61	38.87	42.98	43.62	45.34	51.75	43.04	45.76	42.11

Table 6.1

The third IMF and onwards have been discarded due to the decreasing trend observed in accuracy. The reported results show that the first IMF for any band has better accuracy, and is a more suitable choice for training models of neuromarketing EEG data.

Table 6.2 compares the accuracy reported by Yadava et al [1] and the one reported in this work. It is immediately clear that the proposed methodology vastly outperforms the method used by Yadava et al [1].

Paper	Preprocessing	Machine Learning Model	Accuracy
Yadava et al [1]	Discrete Wavelet Transform	Hidden Markov Model	70%
Current Work	ICA, EMD	Logistic Regression	89.41%

Table 6.2

CHAPTER 7

CONCLUSION AND FUTURE WORK

This work used ICA in conjunction with EMD to outperform the results obtained by the original authors of the dataset used. An extensive literature survey revealed that such a technique has not been used before in neuromarketing EEG studies. It was verified that centro-parietal lobes are highly suitable for the task performed. It was also observed that Alpha and Beta bands are the most suitable choices for training this model. Occipital lobe was found to give best results for neuromarketing data. The results also indicate that the first IMF contains the most useful data for predicting consumer preference.

In the future, the delta, theta, and gamma sub-bands can be further pre-processed to yield better results. The accuracies from said bands are below 50%, which means the model is underfitted and the complexity of the data is unable to be captured. A more rigorous feature matrix can be extracted to improve results. The performance of Logistic Regression can be compared against other baseline ML algorithms, like K-Nearest Neighbours, Random Forest, and State Vector Machine.

REFERENCES

[1] M. Yadava, P. Kumar, R. Saini, P. P. Roy, and D. Prosad Dogra, "Analysis of EEG signals and its application to neuromarketing," *Multimedia Tools and Applications*, vol. 76, no. 18, pp. 19087–19111, 2017.

[2] S. J. Smith, "EEG in the diagnosis, classification, and management of patients with epilepsy," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 76, no. suppl_2, pp. ii2–ii7, 2005.

[3] C. P. Derry, M. Davey, M. Johns, K. Kron, D. Glencross, C. Marini, I. E. Scheffer, and S. F. Berkovic, "Distinguishing sleep disorders from seizures," *Archives of Neurology*, vol. 63, no. 5, p. 705, 2006.

[4] P. D. Harvey, "Domains of cognition and their assessment," *Dialogues in Clinical Neuroscience*, vol. 21, no. 3, pp. 227–237, 2019.

[5] S. Agarwal and T. Dutta, "Neuromarketing and consumer neuroscience: Current understanding and the way forward," *DECISION*, vol. 42, no. 4, pp. 457–462, 2015.

[6] W. M. Lim, "Demystifying neuromarketing," *Journal of Business Research*, vol. 91, pp. 205–220, 2018.

[7] N. I. A, "The Influence of Subconscious Mind on Human Behavior", *JPCBR*, vol. 2, no. 2, Dec. 2017.

[8] G. Chanel, Karim Ansari-Asl, and Thierry Pun, "Valence-arousal evaluation using physiological signals in an emotion recall paradigm," *2007 IEEE International Conference on Systems, Man and Cybernetics*, 2007.

[9] C. Morin, “Neuromarketing: The New Science of Consumer Behavior,” *Society*, vol. 48, no. 2, pp. 131–135, 2011.

[10] N. Lee, A. J. Broderick, and L. Chamberlain, “What is ‘neuromarketing’? A discussion and agenda for future research,” *International Journal of Psychophysiology*, vol. 63, no. 2, pp. 199–204, 2007.

[11] R. M. A. R. K. WILSON, J. E. A. N. N. I. E. GAINES, and R. O. N. A. L. D. P. A. U. L. HILL, “Neuromarketing and consumer free will,” *Journal of Consumer Affairs*, vol. 42, no. 3, pp. 389–410, 2008.

[12] M. Hubert and P. Kenning, “A current overview of Consumer Neuroscience,” *Journal of Consumer Behaviour*, vol. 7, no. 4-5, pp. 272–292, 2008.

[13] T. H. A. Bijmolt, H. J. Van Heerde, and R. G. M. Pieters, “New empirical generalizations on the determinants of price elasticity,” *Journal of Marketing Research*, vol. 42, no. 2, pp. 141–156, 2005.

[14] G. Vecchiato, L. Astolfi, F. De Vico Fallani, J. Toppi, F. Aloise, F. Bez, D. Wei, W. Kong, J. Dai, F. Cincotti, D. Mattia, and F. Babiloni, “On the use of EEG or MEG brain imaging tools in neuromarketing research,” *Computational Intelligence and Neuroscience*, vol. 2011, pp. 1–12, 2011.

[15] P. Golnar-Nik, S. Farashi, and M.-S. Safari, “The application of EEG power for the prediction and interpretation of consumer decision-making: A neuromarketing study,” *Physiology & Behavior*, vol. 207, pp. 90–98, 2019.

[16] M. Aldayel, M. Ykhlef, and A. Al-Nafjan, “Deep learning for EEG-based preference classification in neuromarketing,” *Applied Sciences*, vol. 10, no. 4, p. 1525, 2020.

[17] H. N. Oon, A. Saidatul, and Z. Ibrahim, “Analysis on non-linear features of electroencephalogram (EEG) signal for neuromarketing application,” *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)*, 2018.

[18] S. D. Roy, “On maximally flat sharp cutoff low-pass filters,” *IEEE Transactions on Audio and Electroacoustics*, vol. 19, no. 1, pp. 58–63, 1971.

[19] B. Ans, Y. Coiton, J.-C. Gilhodes, and J.-L. Velay, “A neural network model for temporal sequence learning and Motor Programming,” *Neural Networks*, vol. 7, no. 9, pp. 1461–1476, 1994.

[20] J. Ruan, X. Wu, B. Zhou, X. Guo, and Z. Lv, “An automatic channel selection approach for ICA-based motor imagery brain computer interface,” *Journal of Medical Systems*, vol. 42, no. 12, 2018.

[21] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[22] S. Sperandei, “Understanding logistic regression analysis,” *Biochemia Medica*, pp. 12–18, 2014.