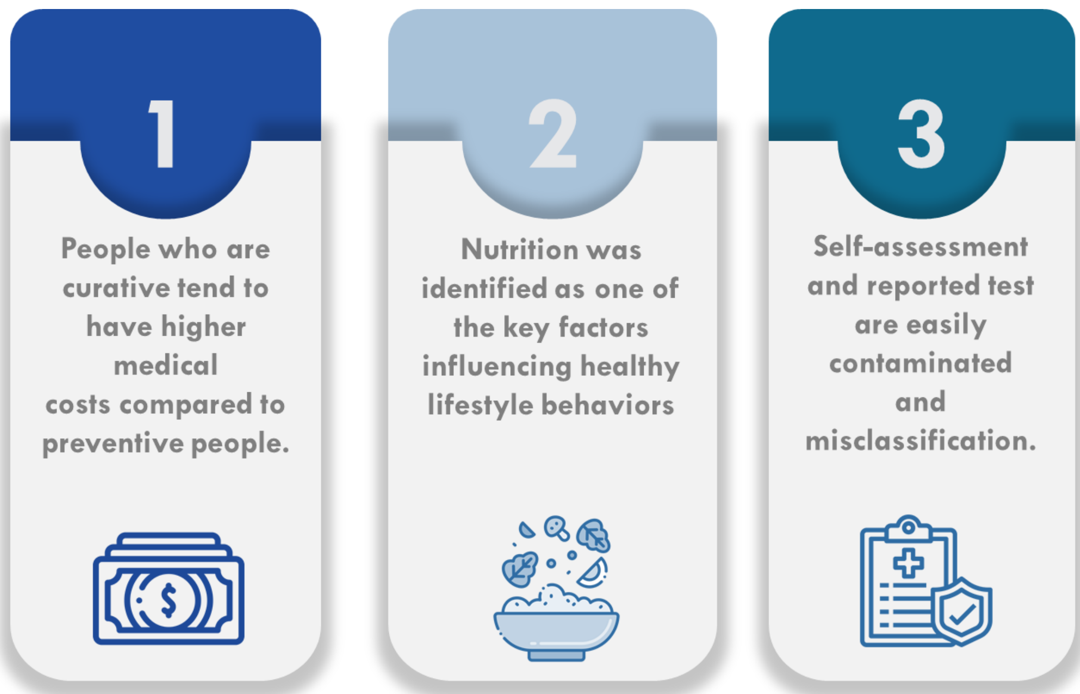
DATA SCIENCE PROJECT

HEALTH PROFILING CLASSIFICATION USING EEG BRAIN SIGNALS AND MACHINE LEARNING

- CUT AMALIA SAFFIERA -

BACKGROUND OF THE STUDY



There are two profiles regarding healthy lifestyle which are preventive and curative.

1. **A preventive people** have habit that maintains his health and avoid potential health problem or treating them early before feeling sickness or having symptoms.
2. **A curative people** have habit or lifestyle that not really maintains his health, just treat the disease after the pathological process has started.

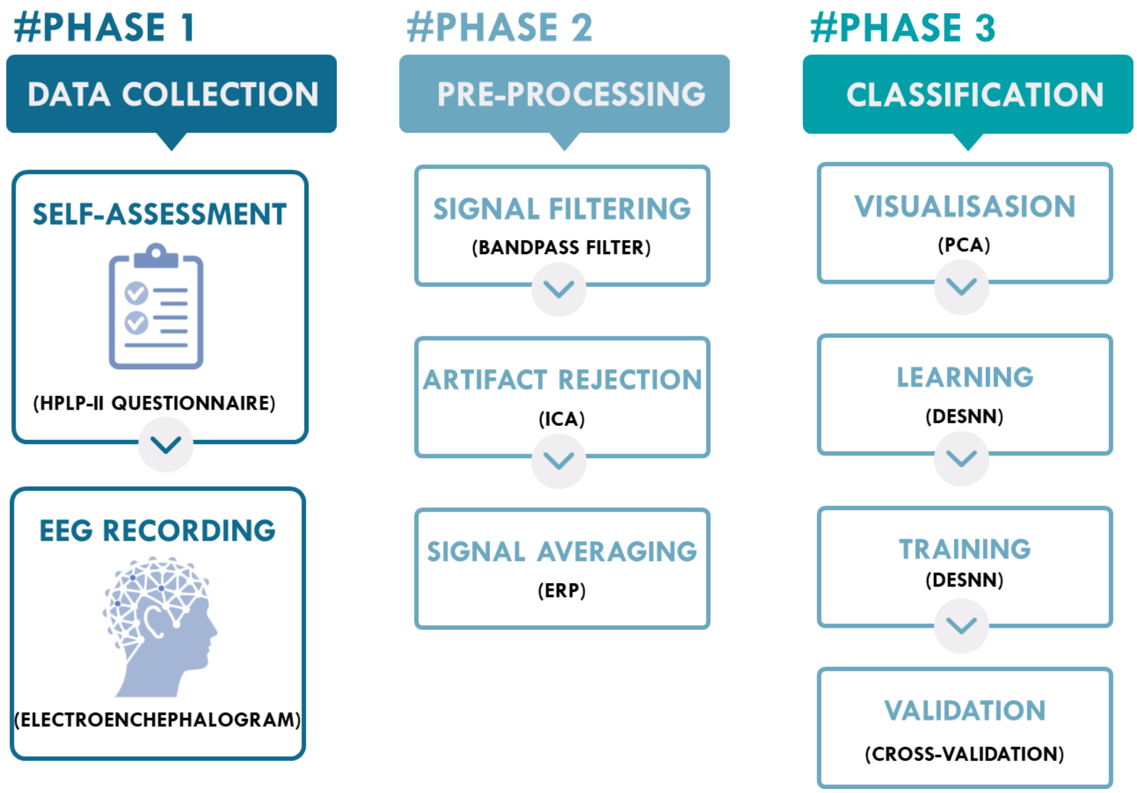
It is recorded that curative resulted **higher medical cost** compared to preventive one. Therefore, it is importance to increase the awareness of people regarding this matter.

In many previous studies, individual behavior toward a healthy lifestyle was assessed in a Self-Report. But, the self-reports of health-related behaviors are widely known of being vulnerable generating **various errors** in consequence of cognitive factors

With regard to profiling method, another alternative method emerges based on the knowledge that the different personality or is influenced by the perception formed by each individual, which comes from the **human brain**. EEG is one technique that have ability measure brain signal.

Hence the aim of this project is to identify and classify individual profile, namely preventive and curative using the EEG data. This project applied the **machine learning** technique for the classification to increase the **accuracy** of the previous assessment.

RESEARCH METHODOLOGY



There are 3 phases of this project:

Data Collection

There are two methods for collecting data. For the needs of psychological assessment, this study uses HPLP-II Questionnaire, that focus on nutrition dimension. And for the needs of neurological assessment, it is use electroencephalogram or EEG.

Pre-processing

This step starts with signal filtering using Bandpass filter. Then signal artefact with ICA to remove the artefact such as eye movement, body movement, and other noises identified. Next is signal averaging to convert EEG into the ERP brain signals.

Classification

First step is the visualization of dataset with PCA (Principal Component Analysis) to do visual inspection of distribution subject. Lastly, the classification using dynamic evolving spiking neural network (deSNN), including learning, training, classifying, and validating.

EEG DATA COLLECTION

PARTICIPANT

22 Undergraduate Student at International Islamic University Malaysia

EXCLUSION CRITERIA

(1) eating disorders
(2) substance addiction
(3) neurological disorders

REQUIREMENTS

stop consuming any food except mineral water (250 ml) within 3 hours prior to the session

BRIEFING

A description of the research is generally delivered to the subject to establish an adequate understanding of the research.

EXPERIMENT CONDITION CONTROL

Subjects have to sit facing the monitor calmly and comfortably during signal recording.
Temperature and light levels of the room are also regulated.

EXPERIMENTAL PROTOCOL

The procedure of the EEG data capturing is divided into three parts, specifically Baseline, Visual Task, and Baseline Extension.

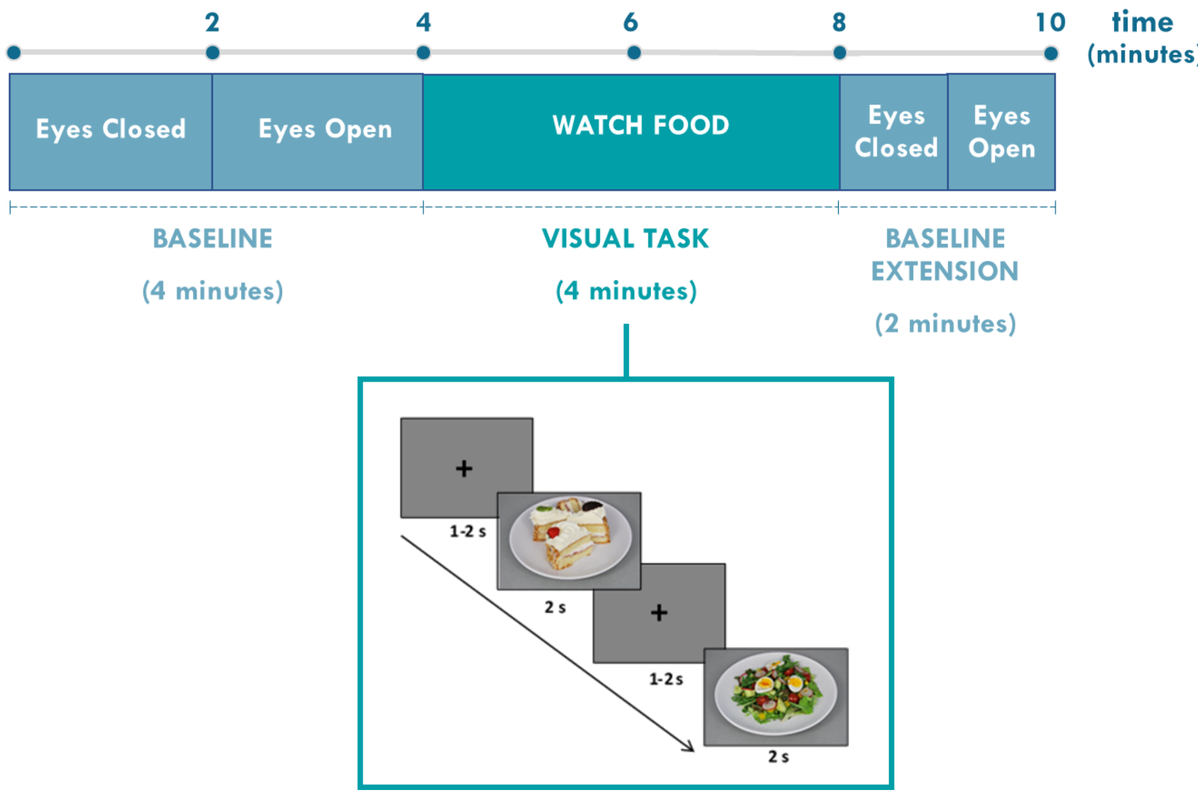
FILLING IN FORM

(1) Consent Form
(2) Additional Consent Form
(3) Personal Information
(4) Health and Hunger Conditions

ELECTRODE PLACEMENT

Using the International 10-20 system (19-channels)

EEG EXPERIMENT PROTOCOL



The experiment protocol is the instruction or condition that the participant will attend to and the EEG will record the brain response.

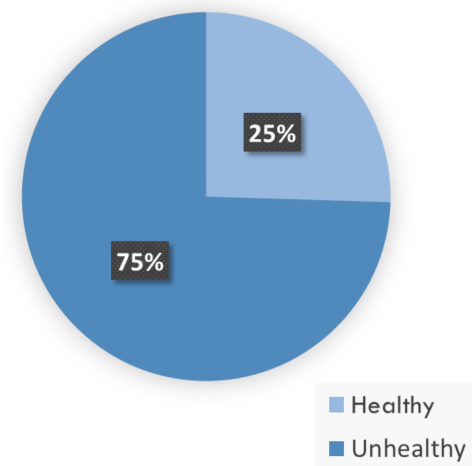
The first one is Baseline, recording two minutes for each session for **Eyes Closed** dan **Eyes Open**.

Then proceed to Visual Task named **Watch Food**. These stimuli present a set of food images (30 healthy food; 30 unhealthy food) with the aim of seeing one's perceptions and tendencies in choosing daily food. The images are shown in random order for 2000 ms each and preceded by a fixation cross (1000-2000 ms) for each image. Participants are instructed to view and pay attention to each image naturally the entire time it appeared on the screen. The food images were taken from the Standardized Food Images (SFI) database by Charbonnier et al. (2016). Nutritionists or experts were asked to choose food images that represented healthy and unhealthy lifestyles based on the HPLP questionnaire.

As closing it will be baseline extension, **Eyes Closed** and **Eyes Open** again in about 1 minute for each session.

RESULTS AND FINDINGS

A. HPLP-II RESULTS



RESPONDENT

208

Undergraduate Student

EXCLUSION

25

eating disorders, substance addiction, and neurological disorders

HEALTHY

53

Female = 41
Male = 12

UNHEALTHY

155

Female = 78
Male = 77

Screening Stage

HEALTHY
TOP 11 FOR HIGHEST SCORE
As Preventive

UNHEALTHY
TOP 11 FOR LOWEST SCORE
As Curative

The distribution of HPLP-II questionnaires that had been carried out for 3 weeks received responses from 208 undergraduate students in IIUM. After scoring, 53 people (F = 41; M = 12) had a healthy lifestyle, while 155 people (F = 78; M = 77) had an unhealthy lifestyle. As many as 25 people were excluded because of eating disorders, substance abuse or addiction, and neurological disorders.

Twenty-two participants (Female = 11; Male = 11) with the **lowest** or highest score were chosen as representative subjects for each group. High scores (scores ≥ 23) are categorized as preventive groups, and low scores (scores ≤ 19) are categorized as curative groups.

B. EEG RESULTS

PREVENTIVE

P3 COMPONENT		
Channels	**Stimuli**	
	Healthy Food	Unhealthy Food
P3	1.05	0.56
P4	0.79	0.93
P7	0.91	0.17
P8	0.58	0.58
O1	1.61	1.14
O2	2.49	1.53
Mean	**1.24**	**0.82**

LPP COMPONENT		
Channels	**Stimuli**	
	Healthy Food	Unhealthy Food
P3	0.88	0.87
P4	1.05	1.23
P7	0.69	0.38
P8	1.04	1.32
O1	0.49	0.26
O2	1.47	0.5
Mean	**0.94**	**0.76**

CURATIVE

P3 COMPONENT		
Channels	**Stimuli**	
	Healthy Food	Unhealthy Food
P3	1.96	2.83
P4	2.23	2.52
P7	2.78	3.04
P8	3.81	3.48
O1	3.76	3.48
O2	3.35	4.94
Mean	**2.98**	**3.38**

LPP COMPONENT		
Channels	**Stimuli**	
	Healthy Food	Unhealthy Food
P3	2.2	2.51
P4	2.89	2.73
P7	2.51	3.12
P8	4.32	4.09
O1	2.76	2.53
O2	2.78	3.7
Mean	**2.91**	**3.11**

As an initial analysis, MATLAB was applied to derive grand-averaged ERP waveforms across 22 subjects. The grand-averaged ERP is conducted for each group, preventive and curative.

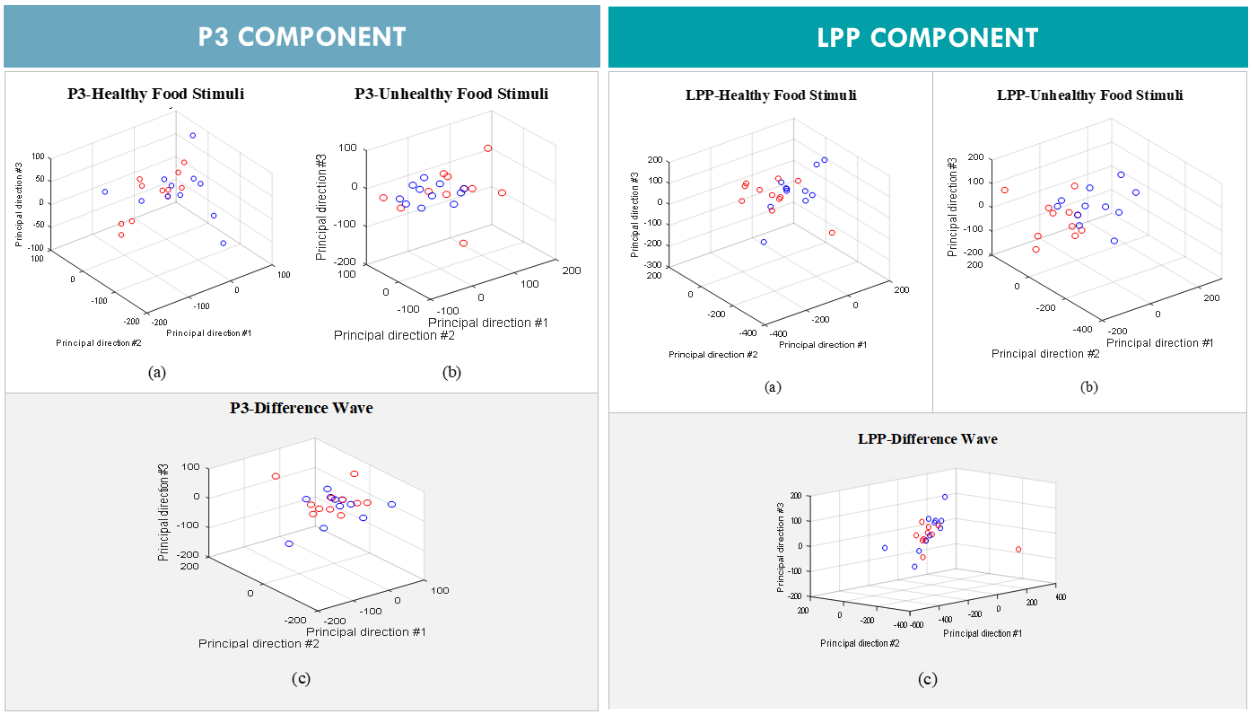
ERP result presented into 2 approaches: **P3 and LPP component**. Because P3 and LPP are known as the two robust ERP components and are generally used in investigating the cognitive function. ERP results found that these 6 channels (P3, P4, P7, P8, O1, O2) produces the greater amplitude than another channel. Therefore, further examination is conducted of those channels.

According to the table, both of P3 and LPP Component show the mean amplitude generated in preventive group is greater for the healthy food images. And conversely for the curative group, a greater mean amplitude is generated for unhealthy food.

These results are suspected to be partly due to the fact that the food picture is strongly linked to food choices in daily life. This might reflect their **preference in choosing food** because in the preventive profile that more maintain their health has a greater amplitude on healthy food. Whereas for curative that is not really maintaining their daily food, shows the greater when seeing unhealthy food.

C. PCA RESULTS

PCA RESULTS



Stimuli conditions:

1. Healthy Food
2. Unhealthy Food
3. Difference Wave (Healthy Food – Unhealthy Food)

= PREVENTIVE

= CURATIVE

PCA results visualize the distribution of subject dataset of this 3 D graph. The blue circle refers to preventive, and red circle refers to curative. This visualization conducted on 3 stimuli condition.

1. Healthy food stimuli -> brain data responses when receive the healthy food stimuli
2. Unhealthy food stimuli -> brain data responses when receive the healthy food stimuli
3. Difference wave -> healthy food stimuli minus unhealthy food stimuli

Based on the 3D visualization shows the **most excellent performance** on **unhealthy food** on **LPP wave** in quantifying the profiles. As we can see in this graph, these 2 groups look separately. Thus, visually this graph has ability to distinguish a person's profile related to their lifestyle.

Whereas in another condition it is very difficult to draw the boundaries to divide the two regions on the red and blue circles. Because some point the red and blue circles is very close and overlap one another.

D. DESNN RESULTS

P3 COMPONENT

Accuracy	Healthy Food	Unhealthy Food	Difference Wave
Class 1	18.18%	36.36%	27.27%
Class 2	63.64%	45.45%	27.27%
Overall	**40.91%**	**40.91%**	**27.27%**

LPP COMPONENT

Accuracy	Healthy Food	Unhealthy Food	Difference Wave
Class 1	27.27%	50.00%	18.18%
Class 2	36.36%	33.33%	45.45%
Overall	**31.82%**	**41.67%**	**31.82%**

CLASS 1 = PREVENTIVE
CLASS 2 = CURATIVE

Here are the classification results using deSNN on Neucube architecture:

On P3 component, the classification accuracy for the healthy and unhealthy food shows better compared to difference wave. We found also the accuracy rate on classifying class 1 is very low. It means it's very difficult to identify the preventive subject considered because the subject who represent preventive are not good enough to form distinctive characteristics.

On the LPP component, the classification accuracy for the **unhealthy food performs best classification** than others. It supports the **PCA result** that we can see here also the unhealthy food stimuli produce better. The-accuracy rate is **relatively low** most likely due to twenty-two subjects are considered **inadequate to form a stable profiling method**. The diversity of characters, preferences, experience, and other aspects related to food makes profiling difficult to do with a small number of subjects. Supposedly, in this profiling case, more participants were needed as samples for learning and training datasets.

CONCLUSIONS

The EEG greater amplitude produced in the healthy food stimuli by the preventive and in the unhealthy food stimuli by the curative were suspected to be partly due to the fact that the food picture is intimately linked to food choices in daily life.

By using PCA, the unhealthy food stimuli provided excellent performance and are suitable in distinguishing the profile groups.

DeSNN result that shows the classification accuracy in unhealthy food stimuli is optimum than other conditions. The accuracy rate might get higher if the number of samples is increased and the quality of the subject is upgraded

LIMITATION & FUTURE WORKS

LIMITATION

Insufficient data and the subject's qualifications do not present the profile strongly.

The subjects are considered not qualified to be optimal to be a representative of each group, especially in preventive groups.

FUTURE WORKS

The protocol design for ERP experiments can be constructed more to unhealthy food stimuli.

Stimuli can be more varied by holding instructions such as directions for :

(1) thinking about the taste of the food
(2) imagining eating the food
(3) thinking about the short and long term consequences of eating the food.

The environment for distributing the questionnaires should be specified by setting targets for certain populations, for example sports areas or student's degree in health field for finding the extremely preventive people.