Πολυτεχνείο Κρήτης

Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Πρόγραμμα Προπτυχιακών Σπουδών

Αυτοματοποιημένη Αναγνώριση Προσωπικότητας με χρήση Λειτουργικής Συνδεσιμότητας Εγκεφαλογραφήματος

Κωνσταντινίδη Ε. Παναγιώτα

Διπλωματική εργασία υποβληθείσα στο Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πολυτεχνείου Κρήτης για την απόκτηση του διπλώματος

Χανιά, Ιούνιος 2018

Κωνσταντινίδη Ε. Παναγιώτα

Αυτοματοποιημένη Αναγνώριση Προσωπικότητας με χρήση Λειτουργικής Συνδεσιμότητας Εγκεφαλογραφήματος

ΤΜΗΜΑ ΟΠΟΥ ΕΚΠΟΝΗΘΗΚΕ Η ΕΡΓΑΣΙΑ : ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΕΠΙΒΛΕΠΩΝ: ΜΙΧΑΛΗΣ ΖΕΡΒΑΚΗΣ
Αντιπρύτανης Πολυτεχνείου Κρήτης

Εγκρίθηκε από την τριμελή επιτροπή την................

....	....	....
Μ. Ζερβάκης	Ε. Πετράκης	Μ. Κλάδος

Αφιέρωση
Στη δεύτερη ‘‘μητέρα’’ μου

Ευχαριστίες

Σε αυτό το σημείο, θα ήθελα να ευχαριστήσω όλους αυτούς που με βοήθησαν στη διεκπεραίωση αυτής της εργασίας. Συγκεκριμένα, τον καθηγητή μου Μιχάλη Ζερβάκη (επιβλέπων) που με στήριξε από την πρώτη στιγμή και ήταν δίπλα μου για να μου λύσει κάθε πιθανή απορία καθώς και τον καθηγητή Μανούσο Κλάδο ο οποίος αν και μακριά συνέβαλε όσο ήταν δυνατό στην υλοποίηση της διπλωματικής. Επιπρόσθετα, όλα αυτά θα ήταν αδύνατο να γίνουν χωρίς τη συμβολή του Πολυτεχνείου Κρήτης που επί πέντε συνεχή χρόνια προσέφερε τις υποδομές ώστε να καταρτίσει όσο το δυνατόν καλύτερα τους υποψήφιους μηχανικούς. Τέλος, θα ήθελα να ευχαριστήσω όλους όσους στάθηκαν κοντά μου και με στήριξαν όλο αυτό το διάστημα, οικογένεια και φίλους.

Αυτοματοποιημένη Αναγνώριση Προσωπικότητας με χρήση Λειτουργικής Συνδεσιμότητας Εγκεφαλογραφήματος

Περίληψη

Σκοπός της παρούσας εργασίας είναι να εξετάσει κατά πόσο είναι εφικτό να ανιχνευθούν οι βασικές διαστάσεις προσωπικότητας χρησιμοποιώντας ως βάση τη νευροφυσιολογία και συγκεκριμένα την ανάλυση και επεξεργασία εγκεφαλογραφικών σημάτων. Δεδομένου ότι αποδεδειγμένα δεν είναι εφικτό να ανιχνεύσουμε στοιχεία της προσωπικότητας από εγκεφαλογραφικά δεδομένα που έχουν καταγραφεί σε κατάσταση ηρεμίας, προτείνουμε μία μέθοδο βασισμένη στην ιδέα της συναισθηματικής επεξεργασίας. Συγκεκριμένα, τα σήματα τα οποία επεξεργαζόμαστε έχουν καταγραφεί κατά τη διάρκεια ενός πειράματος όπου οι συμμετέχοντες παρακολουθούν βίντεο έντονου συναισθηματικού περιεχομένου. Υπάρχει μεγάλος όγκος δεδομένων, ωστόσο, σε αυτήν την εργασία, επιλέγουμε να αναλύσουμε εκείνα τα σήματα που θεωρούμε ότι προκαλούν τη μεγαλύτερη συναισθηματική διέγερση με βάση το περιεχόμενό τους ώστε να καταλήξουμε σε έγκυρα και εμπεριστατωμένα αποτελέσματα. Πιο αναλυτικά, λοιπόν, διαλέγουμε τα εγκεφαλογραφικά σήματα που προέρχονται από τα βίντεο που προκαλούν υψηλή διέγερση και καλύπτουν όλο το εύρος του "σθένους", δηλαδή την κλίμακα που φανερώνει πόσο δυνατό είναι το συναίσθημα που ο εκάστοτε συμμετέχων βιώνει. Τα χαρακτηριστικά που εξάγουμε από τα εγκεφαλογραφικά σήματα αφορούν στα νευρωνικά δίκτυα του εγκεφάλου και συγκεκριμένα στα χαρακτηριστικά που διαμορφώνουν τη λειτουργική του συνδεσιμότητα. Εξάγεται ένας μεγάλος αριθμός χαρακτηριστικών οπότε είναι αναγκαίο να περιοριστεί και να δοθεί έμφαση σε αυτά τα χαρακτηριστικά που κυριαρχούν και παρέχουν την περισσότερη πληροφορία. Η ταξινόμηση, που αποτελεί και το τελευταίο κομμάτι της έρευνας, καταλήγει για κάθε συμμετέχοντα, στον επιμέρους δυαδικό διαχωρισμό των διαστάσεων προσωπικότητας (υψηλό-χαμηλό). Συγκρίνονται θεμελιώδεις τεχνικές και αλγόριθμοι μηχανικής μάθησης προκειμένου να βρεθούν εκείνοι που αποδίδουν καλύτερα και καταλήγουν σε μεγαλύτερα ποσοστά ακρίβειας και εγκυρότητας. Οι αλγόριθμοι που μπορούν να διαχειριστούν μεγάλο όγκο δεδομένων και να αντιμετωπίσουν το απαιτητικό πρόβλημα του διαχωρισμού που επιθυμούμε είναι οι Μηχανές Διανυσμάτων Υποστήριξης (ΜΔΥ), η ταξινόμηση με βάση τον κανόνα εγγύτερου γείτονα καθώς και η Ανάλυση Γραμμικού Διαχωριστή. Τα κριτήρια επιλογής των αλγορίθμων όπως και ο ορισμός των παραμέτρων τους είναι άμεσα συσχετισμένα με την τάξη υπολογιστικής πολυπλοκότητας που τους διακρίνει όπως επίσης και με την ικανότητά τους να διαχειρίζονται αποτελεσματικά το μεγάλο όγκο δεδομένων προς επεξεργασία. Στο Πρώτο Κεφάλαιο της παρούσας εργασίας θα εξετάσουμε τις βασικές έννοιες που διέπουν την προσωπικότητα καθώς και μεγάλο κομμάτι της έρευνας που έχει ήδη διεξαχθεί και στοχεύει σε ανίχνευση προσωπικότητας ή/και συναισθηματικής κατάστασης βασισμένο σε νευροφυσιολογικά και μη σήματα. Επιπλέον,

είναι το σημείο που υπογραμμίζει τη σημασία και την καινοτομία της παρούσας έρευνας. Στο Δεύτερο Κεφάλαιο, θα παρουσιάσουμε αναλυτικά όλα τη διαδικασία που ακολουθήθηκε από την εξαγωγή των χαρακτηριστικών ως την ταξινόμησή τους. Στο Τρίτο Κεφάλαιο, θα παρουσιάσουμε και θα αναλύσουμε τα αποτελέσματα της εργασίας. Στο Τέταρτο και τελευταίο Κεφάλαιο, θα γίνει μια συνοπτική παράθεση των συμπερασμάτων που προέκυψαν από την εργασία και θα δοθεί κατεύθυνση για πιθανή μελλοντική έρευνα.

Σημαντικοί όροι: Προσωπικότητα, Επεξεργασία Εγκεφαλογραφικών Σημάτων, Συναισθηματική Επεξεργασία, Νευρωνικά Δίκτυα, Λειτουργική Συνδεσιμότητα, Μηχανική Μάθηση, Εξόρυξη Δεδομένων

Automatic Personality Recognition using EEG Functional Connectivity

Abstract

The aim of the present thesis is to examine whether it is possible or not to detect the dimensions of personality using the base of neurophysiology and specifically, the EEG signals processing and analysis. Given that it is not possible to detect any personality traits using EEG data recorded in resting state, we propose a method based on the concept of emotional processing. In particular, the signals we process are recorded during an experiment where participants watch highly emotional videos. Taking into consideration the large amount of data, we select to analyze only the signals considered to induce the highest emotional stimulation in order to result into more accurate and valid results. Therefore, we select the videos that are characterized by high arousal and range from low to high valence, that is the strength level regarding a particular emotion experience. The features extracted are associated with neural brain networks and specifically, connectivity patterns which define their functional connectivity. A large number of features is extracted that needs to be limited in order to emphasize on the dominant features which provide more information. Classification consists the last stage of the research and it results, for each participant, to the individual binarized discrimination of the personality dimensions (high-low). Multiple machine learning algorithms and methods are compared so as to result into those which outperform and provide higher accuracy and validity. The comparison criteria of classification algorithms as well as the definition of their parameters concern the computational complexity that characterizes them and their ability to deal with large amount of data effectively. In the First Chapter of the present work, we introduce the main terms related to personality and we make a review concerning the research already conducted which aims to personality traits or affective states detection using or not neurophysiological signals. Furthermore, it is the part which underlines the innovation and the impact of the present research. In the Second Chapter, we present a thorough analysis of the process followed beginning from the feature extraction to the classification. In the Third Chapter, we present and discuss the results. In the Fourth Chapter, we conclude the research and provide some orientations of future work.

Keywords: Personality, EEG signals processing, emotional processing, Brain Networks, Functional Connectivity, Machine Learning, Data Mining

Περιεχόμενα - Contents

Chapter 3

Results and Discussion

Chapter 4

Tables

Figures

Chapter 1

Personality Computing

1.1 Area of the study, Personality and Big-Five Traits Model

Personality is a psychological structure gaining lots of attention during the past few years as long as it becomes possible to explain the human behavior at a wide variety using few, specified, and stable characteristics[1]. It can be described as a pattern of an individual's thoughts, behaviors and feelings which make them special and separate them from the others [2]. The physiological and psychological systems of an individual define their kind of behavior and social acting[2]. In the study of personality, we have to deal with diverse factors including family, social and work environment, heredity and even geographical and physical condition that influence peoples' attitudes and preferences. Although, despite the complexity of personality, it can be considered as a worthwhile goal is that it appears to be such a good predictor of important life outcomes in relational, occupational and social functioning, physical morbidity, longevity and mortality. *Science* and *psychology* are the powerful tools which can contribute to the decrease of this complexity and the development of personality models in order to show us more clearly where *Personality Computing* is and how far can evolve [3]. It is a fact that Personality Computing approaches consist of a great help in any technology involving measuring, predicting and structuring the human behavior. Research related to personality and the models developed until now, has successfully predicted patterns of behavior and emotion through examination of diverse information in terms of data and methodologies. The results are strongly associated with significant aspects of life such as physical and psychological health, personal relations with family and others, social-antisocial behavior, criminal activity and political ideology [3] [4].

Measurement of personality traits and their detection has been an issue of research for several years. Experiments demonstrate that some specific traits appear regularly and they are independent in terms of situations and cultures [5]. These traits consist the *Five-Factor Model* or the *Big-Five* personality traits which are the dominant model in personality research and consist a great influence in the field of psychology [6]. Other models have been also developed but most of the time, they are based on the Big-Five theory [5].

More specifically, the Big-Five model is based on common language descriptors of personality. This model suggests that personality can be described in five dimensions which are **O**penness, **C**onscientiousness, **E**xtraversion, **A**greeableness and **N**euroticism (known also by the acronyms OCEAN).

- Openness describes the curiosity, the appreciation of art, adventure and liking of risk a person has.

- Conscientiousness is a measure of self-discipline, responsibility, organization and efficiency.
- Extraversion is about energy, sociability, talkativeness, activity and amount of positive feelings.
- Agreeableness is the tendency of being cooperative, helpful, kind, forgiving, trusting and the like.
- Neuroticism identifies the factor of vulnerability that describes a person and their tendency to psychological stress, namely levels of anxiety, instability and worrying.

These are the five dimensions of the model which describe most of the individual differences in human behavior and personality and their measurement is the main purpose in various areas and fields as we will present afterwards.

In this particular work, we aim to detect and predict personality traits using neurophysiological signals, and specifically Electroencephalogram (EEG) signals. The deployment of affective processing in the general concept of the research is a promising orientation as it will be explained further afterwards. What's more, the understanding of the brain networks functional connectivity is considered to be an appealing approach which provides homogeneous and accurate predictors.

1.2 Affective Computing and Human-computer Interactions

Nowadays, computers and the Internet are embedded in the daily fabric of our lives and consist an inseparable part of our daily routine. Technology is used to work, communicate, do research, even for entertainment and machines play an important role in every aspect of our life we can possibly think about, such as medical treatment, security, gaming, education or travel [7]. When we refer to behavior and attitude, it is impossible to omit the emotion factor, typically the *source* of our behavior. This event can be defined as an interaction of mind, brain and body[7]. The study of emotion could be involved in the term *Affective computing* which has also arisen in interest in the field of personality as long as the human emotional cues consist a significant factor of an in-depth analysis of human behavior and reactions. For instance, N. Al Moubayed et al. [8] used the facial images of individuals and aimed at their personality detection through the facial appearance. Results were satisfying since Openness, Extraversion and Neuroticism appeared to be strong predictors with high accuracy of 65%. Although, accuracy described Agreeableness and Conscientiousness did not even touch the 60%.

As it has been stated by Rosalind Picard in her book, affective computing "relates to, arises from and definitely influences emotions". The crucial role of emotion in intelligence is the main motive of computer scientists, technology developers and Artificial Intelligence researchers to focus on this field [7]. Therefore, Affective computing has become inevitably associated to Human Computer Interaction (HCI). In particular, HCI focuses on the design of computer technology and fespecially, the interaction between user and computer as a "simulation" of human-to-human dialogue. If a computer or a machine is capable of understanding the user's emotional state, more effective and healthy the human-computer interactions will become [7].

More specifically, S. H. Kahou et al. [9] used facial expressions, audio information and mouth features combined with convolutional networks so as to detect emotion in affective videos, namely Angry, Disgust, fear, Happy, Sad, Surprise, and Neutral. In the end, after various methods applied, the highest test accuracy (41.03) resulted from moderate random search. In this search, the different models were re-weighted regarding the emotion predictors with the "angry", "fear" and "disgust" emotion to be the leading for audio, mouth features, and convolutional networks combined with audio analysis. Average prediction method resulted in 37.17 test accuracy while the SVM classification provided only 32.69.

Microsoft Kinect [10] infrared depth sensing device to extract data from facial expression and sensors placed on hands, head and body. Speech features were also examined. Results demonstrated that combination of different modalities had an interesting performance but hand modality finally resulted in the highest accuracy anger detection (0.55) than the other visual modalities. Although, the most accurate anger prediction (0.73) is related to speech recognition.

M. Pantic et al. [11] state that the modeling of human perception using affective computing would largely benefit different research areas and technologies. Besides, as M. Carroll argues [12], in the process, HCI was oriented to satisfy the needs of all people regardless of age, physical needs and disabilities in a wide range including commerce, education, medical applications. Identifying behavioral cues could be of a great assistance in medical applications, enhance scientific understanding towards the development of patient friendly machines and improve the treatment process.

One of the most important work related to HCI and its effect in health care behavior was realized by H. Ronen et al. [13] who selected online feedback from patients to examine whether they were complied with medical prescriptions and in general, the intention to change their behavior. HCI introduces online feedback since it provides self-management for the health care systems and it outperforms in complex tasks. In this particular work, the type of online feedback selected is related to visual, interactive and personalized content. Visual content (**V**) is about visual representation of information, interactive content (**I**) is about the users' immediate control of actions and personalized content (**P**) describes the adjustment of general to personal information. This sort of feedback provides information regarding comprehension, self-efficacy that is peoples' self-confidence to control their own body functions and involvement, namely the level of physical, cognitive and affective participation in any activity. Participants were divided into three categories depending on the interface used (1.VIP feedback, 2.VI feedback, 3.P feedback). In the results, significant relations were observed among VIP level and Self-efficacy (.282, p = .002), Self-efficacy and intention to change (.220, *p*<.001), VIP level and involvement (.356, *p*<.001) and Involvement and Intention to change (.437, *p*<.001). Although, comprehension, self-efficacy and involvement appear to perform moderately in contrast to the statistically strong link between VIP level and Intention to change (.23, *p*=0.05).

When we talk about emotions, it is difficult to omit studies of Paul Ekhman, a great studier of emotion and its physiology. In particular, he argues that emotions can be detected in every part of our life that we consider significant such as family, work environment or friends [14]. He also states that positive emotions improve our quality of life and enhance communication whereas negative ones such as anxiety and stress, can cause real damage, even lead to mental illnesses and chronic diseases. Indeed, if we consider about experiments in the field of affective computing and HCI, it is easily observed that most of them aim at emotion detection and classification through the concept of *emotional processing*. In particular, emotional processing describes the process when we try to stimulate emotion via the display of affective videos, images and the like and then collect audio and/or visual, speech, facial expression and other features in order to detect the affective states of participants and if possible, personality traits according to their behavior.

Foa et al. [15] argue that anxiety occurs when our brain activates certain mechanisms in order to escape or avoid an undesirable situation. These mechanisms are the emotions represented by information structures in memory [15]. Therefore, emotional processing can be determined as the modification of memory structures that hide emotions. Teasdale after several years [16], proposes that effective emotional processing is capable of changing the cues triggered at times of potential relapse of people who have experienced a major depression. His work based on the Mindfulness-Based Cognitive Therapy had the main purpose of learning attentional skills to patients so that they can control themselves regarding their engagement in dysfunctional affective modes. Affect-related mental models resulted into the reduction of depression relapse from 66% of patients to 37%.

The idea of emotional processing is strongly related to *emotional regulation* [17], which followed by the strategies of expressive suppression and cognitive reappraisal. Dorota Szczygiel et al. [17] state that suppression can be described by the inhibition of ongoing emotionally expressive behaviors which leads to decreased emotional responses whereas the experience remains the same. On the other hand, reappraisal is an effort to reduce the impact of a stressful or disturbing situation by altering the way of thinking. As a consequence, the experience of negative emotion is reduced as well as the expressive responses.

Taking into account the definition of personality, it can be realized that individual differences lead to different emotion regulation outcomes. In this part, we can also observe the association of emotion regulation with human-computer interaction. Considering the work of M. Vuorela et al. [18] related to web-based learning in order to determine the emotional reactions of students towards the collaborative learning, emotion regulation and computer self-efficacy regarding their activity in the online environment. Computer self-efficacy describes the self-confidence of students towards their ability of using any part of the web environment. Their affective reactions were analyzed in the two-dimensional space of *valence and arousal*. In the end, reappraisal was the main emotional strategy (Z=-8.00, p<.01) comparing to suppression. Furthermore, positive affects and increased computer self-efficacy led to decreased arousal. Thus, mean valence was negatively related to mean arousal (r_s=-.55, p<.01, n=101) and computer self-efficacy was positively correlated with the mean arousal r_s=.31, p<.01, n=77) respectively.

1.3 Review in Personality Computing

A significant number of methods and technologies have been developed in order to detect and extract the behavioral cues related to the most important social behaviors, namely emotion, personality, status, dominance, persuasion [19]. Peoples' physical appearance, gesture, facial expressions, vocal behavior and peripheral physiological signals are some of the critical factors which are examined in order to construct effective personality models [20].

In recent studies, information on personality traits derive from databases related to various aspects of technology usage. Firstly, *social media datasets* contain serious information which reveal specific personal cues as long as the way people use technology consist a way of personality externalization [5]. For instance, the activation of a Twitter account and the frequent 'tweets' lead to a strong correlation between linguistic cues and personality traits [21]. Specifically, linguistic cues were found to be strongly associated with neuroticism and agreeableness and the method resulted in high accuracy outcomes regarding the prediction of these traits. Other research, considering both Instagram and Twitter [22] makes an effort to propose personality models from features extracted related to image, linguistics using natural language processing resources (e.g. LIWC) and meta-features based on public user accounts such as number of followers and followees. This research, evaluating the Root Mean Square Error concludes that in general, the best performance is for extraversion (RMSE: 0.71) . Apart from this, agreeableness (RMSE: 0.55), neuroticism (RMSE: 0.73) and conscientiousness (RMSE: 0.65) are best predicted by combined features from Twitter (linguistic) and Instagram (caption of image features) while Twitter's linguistic and meta plus Instagram image features result in good prediction of openness (RMSE: 0.53) Furthermore, the use of Facebook and the privacy matter [24] reflect or not a tendency to self-disclosure and liking of self-exposure. Gina M. Chen [23] states that various features contribute to the construction of a personality model through Facebook, namely number of Facebook friends, posts, frequency of updates on friends, hours per week usage are some of the most important characteristics to take into account. In the particular research, Pearson's Correlation resulted in the strongest positive relationship of number of Facebook friends and extraversion (.47, p<.001) and Hierarchical Ordinary Least regression demonstrated, for the first model examined, a moderate positive relationship between extroversion and having more friends on Facebook (β=0.38, p <.001) and for the second model, extroversion was again the strongest predictor (β=0.34, p <.001).

Smartphones and their usage could not but be a part of the personality model construction since they consist a source of social interaction of a great importance. G. Chittaranjan et al. [25] perform a significant research related to the collection of contextual data by mobile phones in order to draw a connection between personality and behavioral aspects. Some of the main features extracted for the dataset refer to SMS (average SMS length, number of words, and number of outgoing/incoming SMS), calls (average call length, number of incoming/outgoing calls), applications (video/audio/music, mail, camera, youtube etc.) and Bluetooth usage. After gender-based classification, results showed that F measures scored low in agreeableness (0.49) and openness (0.54) but outperformed in extraversion (0.67), conscientiousness (0.62) and neuroticism (0.63) for women. As for male population, all the traits were better classified than extraversion (0.49) and conscientiousness and neuroticism were also misclassified with F-scores 0.55 and 0.54 respectively for the entire population.

1.4 Emotion and personality detection using EEG

Electroencephalogram (EEG) is the measurement of brain activity and its signals are proved to be very useful in recognition of emotional states [20]. EEG appears to have a great temporal resolution compared to other techniques such as PET or fMRI. Current emotion recognition techniques have widely applied the use of Electroencephalogram (EEG) signals in order to detect the changes in emotional states [26]. Besides, J. Atkinson et al. [26] state that the application of proper stimuli can lead to successful identification and classification of emotional types. Using kernel classifiers and EEG features, high accuracy results were demonstrated, that is 73% arousal accuracy for 2 emotional classes versus 62% arousal accuracy with Naïve Bayes and 73.14 % valence accuracy versus 67.6 %. N. Jadhav et al. [20] conducted a research including 4 emotions, namely Happy, Angry, Sad, Neutral using band power EEG features in pre and post-meditation. Classification of 2 emotional classes is the most significant in pre-meditation with accuracy: 79.17%, sensitivity: 72.92% and specificity: 85.42% and in post-meditation with accuracy: 59.38%, sensitivity: 45.83% and specificity: 60.43%. For 3 or more emotional classes, the method performs less effectively.

M. K. Abadi et al. [27] conducted a research regarding personality traits and affect using physiological signals (EEG, ECG, GSR) and 16 emotional videos. Affect was evaluated through Positive Affect and Negative Affect (PANAS) schedules. It was observed that peripheral physiological signals were strongly associated with different traits and especially extraversion with the highest F1-score of 0.7. Spectral power EEG features appeared to be related with Openness with F1-score of 0.69. As for PANAS, only positive videos resulted in important outcomes with the “funny” videos to be distinctive for general positive affect.

J. Wache et al. [28] conducted a significant experiment based on physiological responses to detect personality via the Big-Five personality model. Results concluded in baseline accuracy and F1-score for Conscientiousness and Openness in 0.53 and 0.5 for the other traits. High recognition performance concerns the trait of Conscientiousness which is greater or equal to 0.63 and reaches up to 0.91 considering affective responses to all videos. Agreeableness is also a strong predictor for the High Valence - High Arousal quadrant scoring 0.84. Low recognition performance is observed in Conscientiousness and Neuroticism with overall F1-scores of 0.31 and 0.63 respectively.

Bocharov et al. [29] studied one of the most known mental illnesses, called depression. The basic mean of the experiment was the implicit emotion processing of facial expressions combined with EEG oscillatory dynamics. Reduced sensitivity to positive stimuli was observed in mild and borderline depression individuals. Sensitivity was increased in responses to angry faces and decreased to happy faces regarding people in high depression while for low depression scorers the opposite was observed.

In the process, Herbert et al. [30], in an EEG study which has as main purpose the detection of emotional experience through emotional processing(emotional pictures), present the results detected using physiological signals namely EEG and ECG. The study concludes in high arousal ratings for pleasant pictures (Mean=6.26, Standard Deviation=0.14,p<.01) and unpleasant pictures (Mean=4.9, Standard Deviation=0.22,p<.01) while neutral pictures achieved lower values. The ECG resulted in good heartbeat perceivers with F1-score above 0.85 and poor heartbeat perceivers with F1-score below 0.85.

Moreover, Nina Zollinger [31] in her master thesis describes a research related to music stimuli as the choice of emotional processing and tries to stimulate brain in order to detect emotions generated using different types of music. Three types of physiological mechanisms are used namely EEG, heart rate and skin conductance recordings. As for the valence dimension analysis, results demonstrated significant correlation for sad music (r=.52) and happy music (r=.47). Psycho physiological data that is heart rate and skin conductance measurement did not lead in a significant association whereas EEG data resulted into increased theta-band(3-8 Hz) activity for happy music.

Karl Guiseffi [32] also conducts a research using EEG, although this time in a different field of interest, the *political science sector* regarding citizens' voting and participation. Guiseffi contends that brain activity is related to political attitudes and behaviors so he tries to detect brain processing differences through face processing. In the end, non-voters but active citizens were found correlated with the emotion of "disgust" (r=-.25, p<.022, n=83) while voters and active citizens were found strongly related with the "anger" emotion (r=-.09, p<.434, n=78).

Brain Networks and Functional Connectivity

Brain activity consists a complex process which can be better understood through its functional anatomy on the basis of its structure [34]. Brain functions accomplish a variety of tasks [33], such as perception or cognition performing commonly in different brain regions. Brain Networks are described as unique and non-ovelapping sets of brain regions and their architecture is described in the terms of structural-functional *connectivity*. Neuroscience is generally focused on the biology of brain in detail by considering the topology of brain networks and the detection of connectivity patterns [35].

Functional brain processes form the brain's network architecture [34] so that people are capable of performing adaptability, destruction resistance or effective transmission of a message. Park and Friston [34] state that functional connectivity is regularly examined through the nodal activities based on blood-oxygenation level-dependent (BOLD) fMRI or EEG coherence signals during task performance or resting state. Additionally, when brain connectivity is examined, the *graph theory* is applied most of the time and then the network is represented of undirected connections and functional connectivity correlations [35].

C. A. Frantzidis et. Al [36] aimed at the detection of Alzheimer's disease (AD) through neuropsychological examination of the participants to determine their generic cognitive status and EEG data acquisition. Apart from the healthy people and mild AD patients, there were also amnestic Mild Cognitive Impairment (aMCI) patients. Small-world brain architecture statistics are demonstrated as a function of group and density for each network feature. Results showed important effects of group (V=0.44, $F_{(6,120)}$ =5.63, p<0.0001), density (V=0.999, $F_{(9,\ 53)}$ =8766.855, p<0.0001) and group-density interaction (V=0.541, $F_{(18,108)}$ =2.225, p<0.006). Additionally, there were no significant differences among aMCI and mild AD patients.

It has also been observed that biological processes, especially these realized within the brain lead to human behaviors and experiences [37]. More specifically, DeYoung and Gray contend that biological functions of the brain are the base of *Personality Neuroscience* and suggest that the biological sources of individual differences regarding psychology and behaviors can be easily detected through neuroscientific methods. The most important among them are considered to be neuroimaging (PET/MRI),molecular genetics, electrophysiological processes, analysis of endogenous psychoactive substances. DeYoung, after further research, also states [37] that the main purpose in this field of research is not only the determination of the biological systems underlying beyond the traits but also the detection of the parameters which differentiate one person from another in order to construct personality trait models.

The development of Brain-Computer Interfaces has become of a great assistance towards EEG research and Neuroscience. It is based on biofeedback, autonomic function learning and motor neurons learning. Specifically, N. Birmauer et al. [38] based their research on the development of a BCI system for a motor disease of unknown aitiology, Amyotrophic Lateral Sclerosis (ALS). Although, after the EEG data collection, research did not result in acquisition of table communication with the system. The fact becomes obvious considering the high error rate (>80%) even for highly trained patients. Apart from this, detection of emotional state was implied through the projection of affective pictures. Results demonstrated low arousal in images with social context, more positive responses in positive slides and less negative responses in negative slides showing that patients appeared more positive in their emotional state than people in healthy controls.

1.5 Proposed Approach and the Impact of the research

Neuroscientific methods using neurophysiological data and in particular, EEG have made a long way towards the detection of emotion and affective states. Although, no research conducted managed to detect *personality* traits through EEG physiological signals. In particular, K. Korjus et al. [39] have reached to a significant conclusion regarding EEG data by arguing that resting state EEG data and its power spectrum cannot contribute to the detection of emotional states and the prediction of personality traits. Specifically, in this particular research, a large dataset was analyzed consisted of eyes open and eyes closed resting EEG recordings. After the data was extracted, the process of classification followed with various combinations of classifiers and features. There was no significant classification rate and after further methods applied, misclassification rates were still the result. As a consequence, we observe that resting-state EEG cannot be of assistance in the prediction of personality traits, thus it is necessary to introduce other techniques to result in a significant personality assessment through EEG.

The reason we insist on EEG is simple. EEG signals are considered to be necessary and helpful in personality detection firstly, because of their high temporal resolution. Additionally, it is widely accepted that EEG recordings are more credible and objective than other methods or techniques. More specifically, the use of self-assessment questionnaires or annotations realized by the person themselves or an external annotator may contain false or misleading information and an important error factor respectively. In contrast, the EEG signals recorded through an experiment are trustworthy and the external factors that may affect them can be seriously reduced by proper preprocessing. Last but not least, their main benefit is their cost effectiveness and practicality. In particular, it would not be wise to use the fMRI technique in Human-Computer Interaction applications and experiments, since it consists a non affordable solution.

As it can be observed from the previous analysis, there has been significant research conducted related to affective computing and personality traits detection which is strongly associated with Human-Computer Interactions. Various methods and multiple datasets have been examined in order to develop accurate and strong predictors of personality based on features either obtained by physiological responses or extracted from other important sources such as social media platforms and smartphones. In addition, the thorough study accomplished in the field of Neuroscience and Brain Networks is considered to be a valuable contribution regarding the differentiations of human behaviors and aspects of personality. In particular, the functional connectivity features of the human brain consist powerful measures which combine psychology and neuroscientific methods and have a lot to reveal in the study of the 'so-called' Personality Neuroscience that is constantly gaining interest. Last but not least, the concept of emotional processing is implied in an increasing number of experiments as long as it leads to further affective stimulation and assists the detection of emotional states.

The contribution of this work is the attempt to detect personality traits using EEG signals through emotional processing for the first time, having in mind that this would not be accomplished in case of resting-state EEG. Personality recognition is based on the dominant Big-Five personality traits model and the dataset used is **A** dataset for **M**ultimodal research of affect, personality traits and mood on **I**ndividuals and **G**r**O**up**S** (AMIGOS) [43]. In particular, we focus on the EEG modality since the EEG signals are recorded using the low cost Emotiv EPOC Neuroheadset[1] and this may lead the project to a cost effective, quick, portable and useful application in the study of Personality. In the end, EEG recordings provided a large amount of brain connectivity features, proved to be very useful in personality traits detection in various combinations and frequency bands. After multiple algorithm selections and trials in the *Matlab* environment, we found the algorithms which outperformed in the prediction of each dimension of personality separately and provided the highest detection accuracy. The results were really significant since 4 out of 5 dimensions, namely Extroversion, Neuroticism, Agreeableness and Conscientiousness appeared to be accurately detected (> 80%) while Openness was the dimension with the lowest but not insignificant accuracy (75.7%).

From here on, section 2, materials and methods employed for the data compilation are discussed as long as the proposed methodology, feature extraction, selection and classification. Section 3, details and discusses the results. Finally, Section 4 concludes the research.

[1] https://www.emotiv.com/epoc/

Chapter 2

Materials and Methods

2.1 Personality recognition

The Big-Five personality traits as we observed of all the work review are dominant and their detection in different fields and areas develops the whole concept of *Personality Computing*.

The tool that is most used in the measurement of Big-Five is the *questionnaires* [5] where people have the opportunity to self-assess their behavior with Likert scales (from *"Strongly disagree"* to *"Strongly agree"*). Each question rating contributes to the score of a particular trait. One of the main parameters of the questionnaires is the validity coefficient. Specifically, the question item is correlated to the trait it describes. Therefore, the significance or not of a correlation results in a strong or weak prediction of a trait, respectively and describes the value of validity coefficient, which in the end, determines also the validity of the questionnaire. Questionnaires are divided into first person that is the self-assessments where people rate their own behavior, and third person where people are requested to assess a given individual and attribute to them specific traits that may suit their personality. It is a fact that self-assessments have been questioned for their validity since answers might not be objective. Nevertheless, high correlations between self-assessments and observers' assessments led to the wide acceptance of questionnaires.

Alessandro Vinciarelli [40] makes an effort to examine the matter of Social Perception in terms of Automatic Personality Perception that is the prediction of personality traits people attribute to others. He uses Speaker Personality corpus and Face Personality corpus which contain speech samples and face images assessed from 11 observers respectively. Results showed the highest accuracy for the prediction of Agreeableness (78.5%), with Conscientiousness coming into the second place with 72.5% accuracy for the Speaker Personality Corpus. As for the Face Personality Corpus, 67.1% accuracy was observed for Neuroticism while Agreeableness was the least accurate predictor (59.2%).

DeYoung [41] has described that The Big-Five traits are also hierarchically organized *"ranging from narrow-bandwidth constructs to broad meta-traits"*. This means, that the concept of personality and its measurement can be described at different levels of abstraction and fidelity. As for the fidelity spectrum, DeYoung states that Conscientiousness in large samples may be related to Agreeableness and Neuroticism forming the meta-trait of *"Stability"* or *"Alpha"*, associated with central serotonergic functions. On the other hand, Extraversion and Openness may be related in terms of the meta-trait of *"Plasticity"* or *"Beta"*, associated with dopaminergic functions. G. C. Wright [42] describes that hierarchical organization of traits explains certain types of repeated findings. Take as an example the morbidity and mortality in physical and mental disorders which are strongly detected in the terms of Stability. He underlines that, depending on the characteristic we need to predict, a broader meta-trait should be targeted.

2.2 Description of the Dataset AMIGOS

Most of the databases available intend to detect people's affective responses and personality traits as individuals or limited number groups while it is more effective to detect emotions, moods and reactions in social contexts where social interaction takes place and changes the experiment's performance [43]. What's more, none of these databases examined personality computing from the prospect of combined personality and affect. This is why we choose to use in our approach **A** dataset for **M**ultimodal research of affect, personality traits and mood on **I**ndividuals and **G**r**O**up**S** (AMIGOS) [43]. It is about a dataset consisting of diverse aspects of personality and emotion as long as it contains information coming from multimodal neurophysiological signals recordings, namely Electroencephalogram (EEG), Electrocardiogram (ECG) and Galvanic Skin Response (GSR), personality traits questionnaires, anonymized participants' data, mood self-assessment, internal and external annotations and video recording that is frontal HD, full-body and depth videos[43], covering a wide range of personality computing.

Two experiments were realized so that all that kind of information was gathered. At first, the short videos experiment took place where 40 subjects watched 16 short videos (video-duration<250s) of affective content extracted from movies so that specific affective states are elicited. The participants had to describe their feelings through the video processing by selecting the basic emotions generated namely Happiness, Sadness, Fear, Disgust, Anger and Neutral and assess each video as far as valence, arousal, dominance, familiarity and liking are concerned. Secondly, the long video experiment was realized where 37 participants of the former experiment watched 4 long videos (video-duration>14min) of affective content extracted from movies. The video duration's main purpose this time is to elicit various affective states and stimulate different emotions through the story and its content. For the long experiment, 17 of these participants were selected to watch the videos as individuals while the other 20 participants were gathered in 5 groups consist of 4 people per group. Participants followed the same process as before for emotion selection and video assessment. Moreover, as internal annotations, J. A Miranda-Correa et al. [43] describe the self-assessment of affective states participants performed at the start of the experiments and at the end of each video while external annotations is the off-line evaluation the videos by 3 annotators as far as valence and arousal are concerned. To be more specific, *valence* and *arousal* are the main terms which describe emotional experiences. Valence is about positive or negative affectivity while arousal describes how calm or excited someone will be after being exposed in specific stimuli or information (strength-intensity of the emotional state). This two terms form a two dimensional space that consists a common framework when dealing with emotional responses and contributes to the best understanding of emotion [44]. Apart from the experiment processing, participants' personality was profiled through the Big-Five model and their mood through the

Positive Affect and Negative Affect Schedules (PANAS). Finally, wearable commercial sensors are used for the recording of the neurophysiological signal (EEG, ECG, and GSR).

In particular, there is a standard process followed in this research. In the beginning, internal and external annotations are compared. After that, affective responses of both experiments regarding the social context (individual or group setting) are analyzed and the final step is the attempt for personality detection through neurophysiological signal considered either as single modalities or in different combinations. The results demonstrate a strong association as far as internal and external annotations of valence and arousal are concerned with external annotation being a good affective state predictor. Furthermore, differences are revealed associated with the individual or group setting and correlations are observed between the big-five personality traits and the PANAS schedules. Last but not least, the combination of personality traits, PANAS and social context recognition through neurophysiological signals seriously contributes to the prediction of extraversion, emotional stability, positive and negative affect (with EEG) and the prediction of conscientiousness and openness (GSR & ECG).

Stimuli, tools and display

As for the *stimuli selection*, there are two sets of affective videos corresponding to the experiments, short video and long video respectively. The set of short videos was initially annotated by 72 volunteers on the valence-arousal scale and the video annotation is followed by classification of each video into one of four quadrants of the two dimensional valence-arousal space, that is HVHA, HVLA, LVLA, LVHA , where H,L,V,A stand for High, Low, Valence and Arousal respectively. After that, the three most appropriate videos, as far as the origin of scale is concerned, of each quadrant are selected and one video corresponding to each of the four quadrants is selected, as well. Thus, we have a total of 16 videos. As for the long videos set, IMDB Top Rated Movies list was the source of 8 video extracts of movies. The content does not require special knowledge so that it could be widely understood and was highly affective. After that, classification is performed by four researchers in order to categorize the video segments into the suitable quadrant of valence-arousal space. In the end, the process results in the selection of 4 videos regarding the four different quadrants. As far as the *neurophysiological signals* are concerned, they are all recorded using wearable sensors. This fact positively affects the experiments as long as wireless technology of the sensors offers increases flexibility. More specifically, EEG was recorded with Emotiv EPOC Neuroheadset and it is described by 14 channels according to the 10-20 system [45], sampling frequency of 128 Hz and 14 bit resolution. The channels used are AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4. Shimmer 2R Platform extended with an ECG module board (256 Hz, 12 bit resolution) contributes to the ECG recordings and the process followed is described by the placement of two electrodes at the right and left arm and a third electrode placed at the ankle. Heart rate and the full ECG QRS complex are recorded. Shimmer 2R Platform extended with a GSR module board (128 Hz, 12 bit resolution) performs the GSR recordings with two electrodes placed at the left hand and fingers. As for the frontal HD face video, a JVC GY-HM150E camera is placed below the screen and a Microsoft's Kinect V1 is placed above the screen to record both RGB and full depth body videos. Moreover, stimuli presentation, synchronization of signals and self-assessment acquirement are realized with one PC (Intel Core i7, 3.4 GHz). The PC and Shimmer sensors are connected via Bluetooth and the Emotiv headset is connected to the PC wirelessly. A 40-inch screen (1280 x 1024) is selected for the video display which is presented covering the largest area of the screen possible while black background covers the rest area. Every subject has an average distance of 2 meters from the screen and stereo speakers with a loud volume level are implemented.

Protocol of the experiments

40 participants from the age of 21 until the age of 40 were involved in the experiments which took place in a lab. At first, they were informed of the experiments and guided so as to complete the self-assessment form and use properly the affective scales. Before the recording, the placement of sensors went ahead. In the short experiment, participants watched the 16 selected videos randomly and completed a self-assessment in the end of each video. As for the long experiment and group setting, people in groups interacted among them at a sufficient level as long as most of them had a sort of communication before or similar cultural background. This time, participants had to evaluate their emotions at the beginning and at the end of the video display. The experiment was divided into two sessions of two long videos presentation each. A break of 15 minutes was offered between the two sessions. As soon as the experiment was accomplished, each subject was about to complete online questionnaires regarding personality traits and PANAS schedules.

Internal Self-assessment and External Annotations

The internal annotation had to do with the participants' levels of arousal, valence, dominance, liking, familiarity and basic emotions. For all these dimensions, apart from basic emotions, there is a scale from 1 to 9. In particular, 1 in arousal scale stands for "very calm" while 9 stands for "very excited". Valence ranges from "very negative" (1) to "very positive" (9), dominance from "full of emotion" (1) to "in full control of emotions" (9) and liking from "do not like" (1) to "like" (9). As for the familiarity, 1 stands for "Never seen it before" and 9 stands for "Know the video very well". In the end, after completing this part, participants had to select one or more, if they wished, of the basic emotions. Speaking of external annotations, they were performed off-line on the frontal HD face videos on the valence-arousal scale ranging from -1 (low valence/arousal) to 1 (high valence/arousal).

Personality traits and Mood

An online form questionnaire regarding the Big-Five traits is applied for the measurement of personality traits. A 7-point scale of liking is used for the rating of ten descriptive adjectives and after that, the mean calculation follows. The online form PANAS questionnaire is the mean for mood assessment and two 10 question sets describe the positive and the negative affect respectively. This time, a 5-point intensity scale is used for the rating and in the end, results derive from the total of ratings.

Results

Single modalities as well fusion of modalities are applied to different scenarios for the classification process which result to important outcomes. First of all, EEG performs better regarding valence and arousal recognition with ECG coming at the second place. Furthermore, valence and arousal prediction is slightly better in the short experiment compared to the long one. The method also appears to be really effective regarding the prediction of extroversion, emotional stability, positive and negative affect while agreeableness and conscientiousness prediction is not successful through EEG. What's more, conscientiousness and openness are detected through GSR and conscientiousness is detected through ECG.

2.3 Proposed Methodology

Dataset formulation of personality motifs

We select to examine the data of the *short video experiment* so as to reduce the samples we process and enhance the computational speed of the algorithms used. As it was mentioned in the dataset description before, each participant completed the Big-Five questionnaire. Therefore, we choose to use the mean personality scores for each of the five dimensions of personality concerning each subject separately in order to result into binarized (high and low) levels for the OCEAN dimensions. The high and low class division is achieved with *k*-means clustering.

First of all, *k*-means is considered to be an algorithm used in unsupervised machine learning [46]. It is about a method that consists a partitioning analysis technique widely used in data mining. This particular method main purpose is to synthesize a dataset partition of n points into a k clusters set, initially randomly selected, where k stands for number of partitions [46]. As it is stated by S. H. Al-Harbi et al. [46], a single cluster representation is related to the centre or one of the points in the cluster with a total minimum distance of the other points. In the beginning, the algorithm performs k partitions and then, cluster optimization is realized through an iterative process. Nevertheless, in the research conducted [46], *k*-means appears to be a method suitable for classification as well, through supervised learning.

As far as the algorithm itself is concerned, the process is described as follows [47]. In the beginning, n patterns are divided into k clusters in a d dimensional space. This leads to the extraction of k centers, one for each dataset partition. The k_{th} center is located at the centroid and each of the n patterns is assigned to the nearest cluster, taking the minimum distance into account. In the process, the membership function $m(C_j|x_i)$ in each cluster C_j is computed for each pattern x_i, namely this function defines the proportion of pattern x_i belonging to the j_{th} cluster C_j. In general, if the pattern x_i, is closest to the pattern C_j according to the already mentioned minimum distance then $m(C_j|x_i) = 1$, else $m(C_j|x_i) = 0$. After that, the centres are computed again in order to find the new centres v_j and calculate the square error E with the following equations

$$v_j = \frac{\sum_{i=1}^{n} m(C_j|x_i)x_i}{\sum_{i=1}^{n} m(C_j|x_i)} \text{ for } j = 1, \ldots, k.$$

$$E = \sum_{j=1}^{k} \sum_{x_i \in C_j} \|x_i - v_j\|^2 \text{ for } i = 1, \ldots n; j = 1, \ldots k.$$

The last step repeats until the point of convergence when patterns are no longer reassigned to new clusters, value of E gets below a certain threshold or the number of defined iterations is reached [47].

Before applying k-means clustering, we also checked the performance of median dichotomization. The only difference between the two clustering methods is that for median clustering, the sum of squared Euclidean distances has to be replaced with the sum of absolute distances [48]. Thus, the minimizing centroid for a cluster is the median of the points in the cluster and not the mean [48]. The computation of the median leads to increased time complexity while k-means is faster. The main advantage of median compared to k-means is its robustness to outliers [49]. Although, this is not an issue in this project since the mean personality scores range from 1 to 7 and therefore, there is no risk of having outliers. Figure 1 describes the distributions of the mean personality scores and the results of the clustering methods regarding the mean trait score.

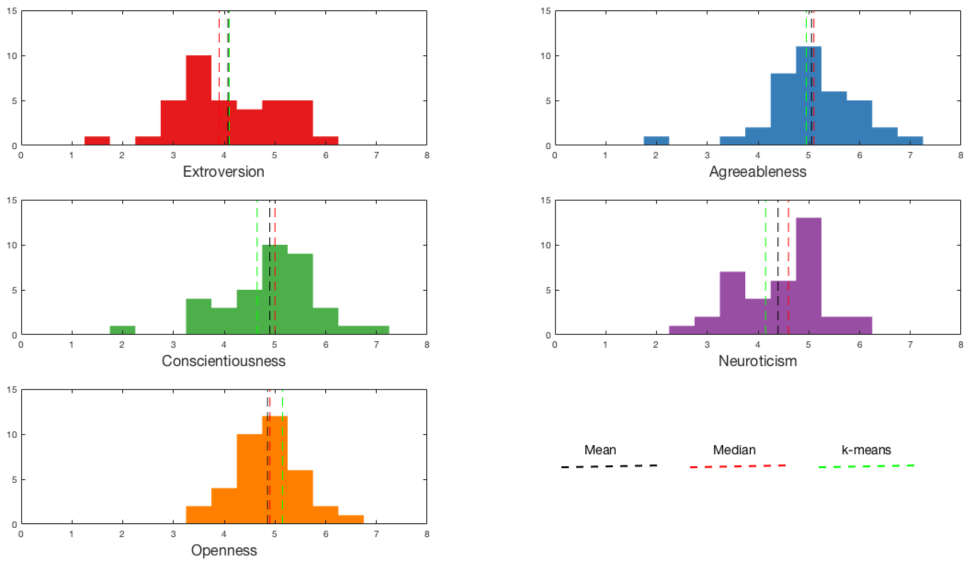

Figure 1

This figure demonstrates personality traits distributions, the mean trait score and the results of *k*-means and median clustering.

Generally speaking, it has been observed that *k*-means clustering [50] is an algorithm that performs better and results in better outcomes when a large amount of data has to be processed. Apart from this, it is an algorithm easy to implement, simple, efficient and provides empirical success [51]. In particular, research regarding personality segmentation [50] using *k*-means clustering resulted into clearly distinguished personality patterns comparing to other methods such as Ward's hierarchical clustering method.

Taking all the above into consideration, it becomes obvious that *k*-means is a suitable algorithm for our research. The class division will be produced using average thresholding through the mean personality scores provided by the AMIGOS dataset. The low/high clustering for each dimension adds significant value to the initial purpose of the project that is the classification of the five dimensions of personality separately. Although, it is not possible to perform classification for all participants as long as personality ratings for 3 participants are missing. Therefore, we proceed to the clustering regarding the rest 37 participants.

In the process, after *k*-means clustering, we continue with the feature extraction based on the EEG modality and the detection of significant functional connectivity patterns in the Matlab environment. We limit the research to the short videos corresponding in the **HVHA** and **LVHA** quadrants since we expect the high arousal to be more helpful and stimulating regarding the detection of affective states. In particular, as C. Lithari et al. [52] have documented, high arousal increases the global efficiency of functional networks since it guarantees a more efficient communication between nodes. In the beginning, we extracted features related to the time-frequency domain. However, we decided to omit these features and emphasize on the brain connectivity features since the EEG functional connectivity is expected to result into a homogeneous and clear conclusion for the dimensions of personality.

The feature extraction is followed by the feature selection since a large number of features is extracted from each participant. This is the reason why feature selection is an indispensable part of the process in order to reduce the amount of data and the computational complexity of the algorithms used in the classification section. In particular, the *10 best features* are selected that may concern a different frequency band or brain region. The selection is performed firstly for the 4 HVHA and 4 LVHA videos respectively and later for the fusion of them. The feature selection of the fusion of the two quadrants is expected to result into the most significant features concerning our research and play a dominant role in the accuracy of classification algorithms.

The last step of the methodology is the classification. Using the *classification learner* of the Matlab environment, we proceed to different algorithm selections and parameter optimization. Multiple trials are expected to lead us to the algorithms which outperform in the prediction of each dimension of personality and provide the highest detection accuracy. The classification and results is the most crucial part of the project as they determine the success or failure of the initial hypothesis, namely the personality traits detection using the EEG modality through emotional processing.

The flowchart below describes briefly the proposed methodology process.

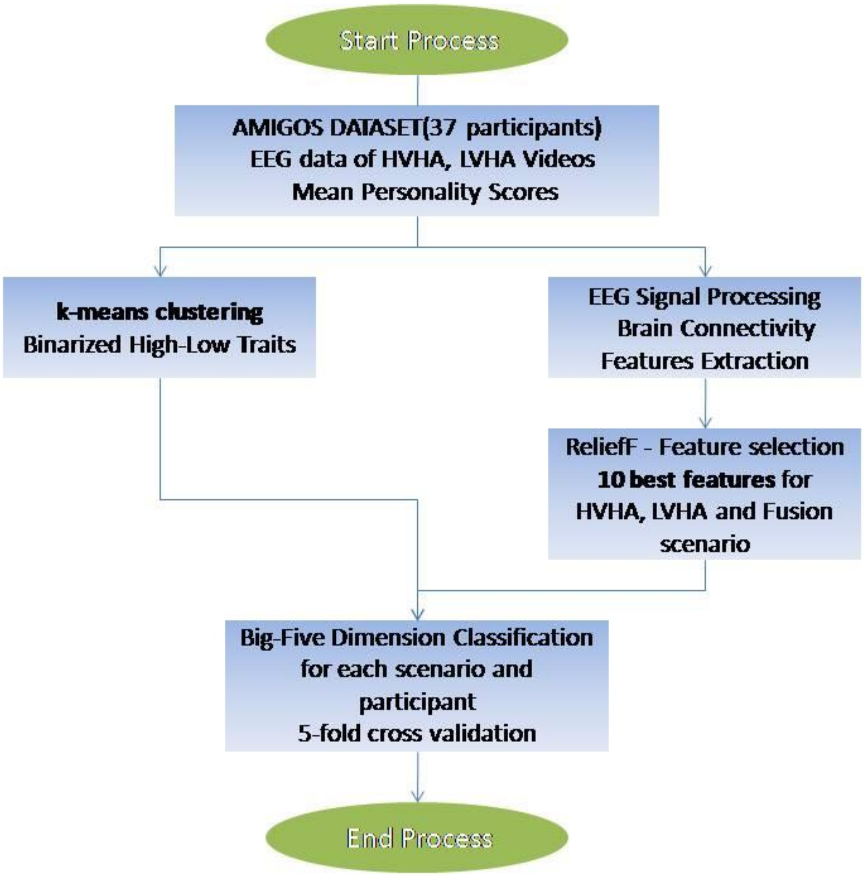

Figure 2

Methodology Process Flowchart

2.4 Feature Extraction

As we already mentioned before, the brain network architecture is based on connectivity data that is specific neurobiological and significant network measures concerning structural and functional connectivity. The neurobiological interpretation of network topology can be described in terms of brain mapping methods, anatomical parcellation schemes and connectivity measures [35].

In particular, L. Pessoa [35] thoroughly describes the brain network as a real-world complex system that consists of nodes-vertices and links-edges which connect the nodes. Nodes represent usually a brain region while links represent the structural, functional and effective connections. The proper construction of a brain network implies that these anatomical and functional connections subject to coherent patterns. Links differentiation depends on their weight and directionality. Binary links imply the absence or presence of connections. In contrast to this, the weight parameter underlies information regarding connection strength. Specifically, a structural's area size, density or coherence is determined by structural network weights whereas magnitudes of correlational interactions are described by functional network weights. Weights are informative concerning various aspects of network organization and strongly affect the non-significant links' filtering. The directionality parameter may be absent. Directed links may represent structural connections. Although, taking into consideration the large number of reciprocal connection, undirected links also provides useful information.

The network can be described in terms of *global* and *local* brain connectivity measures. The measurement of network elements, for instance nodes and links, forms the connectivity profiles. The distribution of all these network elements provides a global description of the network and it is defined commonly by its mean, mostly when the distribution is homogeneous. The measures also vary in terms of binary/weighted and directed/undirected variants.

Feature extraction from the AMIGOS dataset was implemented using the software package *Brain Connectivity toolbox*[2], an open source Matlab toolbox. Specifically, brain connectivity features are extracted from the recordings from the 14 EEG channels *after preprocessing*, namely downsampling, filtering, EOG removal and segmenting. We select to examine the data of the *short video experiment* so as to reduce the samples we process and enhance the computational speed of the algorithms used. For each of the 37 subjects participating in the short video experiment, the preprocessed data consist of three 20 matrix lists, 1 matrix related to each of the 20 videos (trials), that is 16 for the short experiment and 4 for the long experiment and 2 matrices for external annotation and self-assessment values. We further limit our research to the 8 trials related to the short video experiment. We choose to examine only the ***HVHA*** (4 videos) and ***LVHA*** (4 videos) quadrants since these are the quadrants where significant differences in emotional states are more obvious and easier to be detected. In particular, it has been documented in the AMIGOS dataset [43], that high or low levels of valence can hardly be elicitated with low arousal, so this is why we select to omit the videos stimulating low levels of arousal. Subsequently, we subject the video list which includes the samples (EEG recordings) to further process. However, the list contains samples of 17 channels, including 2 ECG and one GSR channels apart from the EEG. Therefore, the 3 last channels will be omitted. The reason why we emphasize on the EEG signals is that the

[2] https://sites.google.com/site/bctnet/

EEG modality outperformed in the prediction of valence and arousal in both experiments [43] and it is possible to lead the research in a cost effective, easy to use and quick application.

The features we extracted are based on the most significant brain connectivity measures [53]. In particular, node *degree* is calculated in order to compute the number of links connected to that node, namely its neighbors. Weighted variant of degree of the node *i* connected to the nodes *j* is: $k_i^w = \sum_{j\in N} W_{ij}$, where W_{ij} represents the connection weight. Network degree distribution is the degrees of all nodes which determine the network resilience and the weighted variant is described as $P(k^w) = \sum_{k'\geq k^w} p(k')$, where $p(k')$ is the probability of a node having degree k' [54]. By calculating its mean, we are led to the measurement of *density* that is the fraction of present connections to possible connections in the network. The calculation of density comes along with the *total number of edges and total number of vertices* in a specific graph, which also consist two global network features alike the density. In addition, the weighted variant of degree describes the *strength*, namely the sum of the weights of the links connected to a specific node.

It is very important to note that groups of brain regions, known as clusters or modules, interact in order to perform specific tasks. This specialized processing comprises the functional segregation. *Clustering coefficient* is a local measure of functional segregation and it is described by the fraction of triangles around a particular node, $C^w = \frac{1}{n}\sum_{i\in N}\frac{2t_i^w}{k_i\,(k_i-1)}$ [55]. It is equivalent to the fraction of node's neighbors that are neighbors of each other. A large number of triangles indicates segregation. *Transitivity* represents the ratio of triangles to triplets in the network and it is considered as a variant of clustering coefficient. The weighted variant of transitivity we calculate is represented by $T^w = \frac{\sum_{i\in N} 2t_i^w}{\sum_{i\in N} k_i\ (k_i - 1)}$ and it is not available for individual nodes [56].

In addition, another measure of segregation is the *community structure* and *modularity*. Community structure defines the composition of densely interconnected brain region groups and leads to the optimum subdivision of the network into non-overlapping groups of nodes. This means that the number of within-group edges is maximum and the number of between-group edges is minimum. Modularity is the global feature calculated, based on the definition of community structure and represents the degree to which further subdivision may proceed so as to avoid possible overlapping [57]. It is given by $Q^w = \frac{1}{l^w}\sum_{i,j\in N}\left[W_{ij} - \frac{k_i^w\,k_j^w}{l^w}\right]\delta_{m_i,m_j}$, where l^w is the sum of all weights in the network, m_i and m_j are the modules containing the nodes i and j respectively and the portion $\delta_{m_i,m_j} = 1$ when $m_i = m_j$, otherwise 0.

It is now important to introduce the term of functional integration, namely the capability of information processing through communication between different brain regions. Routes of information flow in structural networks may be represented by *paths* concerning specific nodes and links while in functional networks, paths reveal statistical associations.

Therefore, we consider very useful to compute the weighted *characteristic path length* that is the average shortest path between all pairs of nodes in the network [58] and consists a global connectivity measure given by $L^w = \frac{1}{n}\sum_{i \in N} \frac{\sum_{j \in N, j \neq i} d_{ij}^w}{n-1}$, where d_{ij}^w is the weighted distance between the nodes i and j. By calculating the inverse of the characteristic path length, we are led to the *global efficiency* [59], namely$E^w = \frac{1}{n}\sum_{i \in N} \frac{\sum_{j \in N, j \neq i} (d_{ij}^w)^{-1}}{n-1}$. Another parameter we choose to calculate is the *local efficiency*, which is practically the global efficiency but computed on node neighborhoods and it is strongly related to the clustering coefficient. It is also essential to calculate nodal *eccentricity* (ecc), namely the maximal shortest path length between a node and any other node, as well as the minimum and maximum eccentricity that is the *radius* and the *diameter* respectively.

Node *centrality* is also a significant parameter to calculate regarding the detection of important brain regions, known as *hubs*, which are really important as far as the facilitation of functional integration and the network resilience are concerned. At first, node *degree* that was mentioned before, is a dominant measure of centrality since high-degree nodes interact with many other nodes in the network, especially in structural networks where degree is a really sensitive measure. In particular, *within-module degree z-score* (a localized version of degree centrality) describes within-module connectivity [60] and the weighted variant is given by $z_i^w = \frac{k_i^w(m_i) - \bar{k}^w(m_i)}{\sigma^{k^w}(m_i)}$, where m_i is the module that contains node i, $k_i^w(m_i)$ is the within-module degree of i, $\bar{k}^w(m_i)$ and $\sigma^{k^w}(m_i)$ are the mean and standard deviation of the m_i degree distribution respectively. Last but not least, the nodal *eigenvector centrality* is measured that is a self-referential measure of centrality. More specifically, high eigenvector nodal centrality implies the connection of the node to other nodes with high eigenvector centrality.

The differentiation of intermodular interconnections is denoted by the complementary *participation coefficient* (PC) [60] described by $y_i^w = 1 - \sum_{m \in M}(\frac{k_i^w(m)}{k_i^w})^2$, where M is the set of modules and $k_i^w(m)$ is the number of links between i and all the nodes in a given module m. Consequently, provincial hubs, namely nodes with high within-module degree z-score and low participation coefficient, facilitate the modular segregation while connector hubs, that are nodes with high participation coefficient, facilitate global intermodular integration.

Apart from measures based on degree, it is useful to compute another significant parameter, known as *betweeness centrality* (BC), that is the fraction of all shortest paths in the network that contain a specific node. This parameter is based on the concept that central nodes appear in many short paths and "control" the information flow. The undirected variant of betweeness centrality of node i is calculated as $b_i = \frac{1}{(n-1)(n-2)}\sum_{\substack{h,j \in N \\ h \neq j, h \neq i, i \neq j}} \frac{\rho_{hj}(i)}{\rho_{hj}}$, where ρ_{hj} is the number of shortest paths between h and j and $\rho_{hj}(i)$ is the number of shortest paths between h and j that include node i [61]. The weighted variant of betweeness centrality implies the calculation of weighted path lengths.

Network resilience reflects the vulnerability of the network. *Assortativity coefficient* is considered to be a useful measure of network resilience and it is described as a correlation coefficient between the degrees of all nodes on two opposite ends of a link. It is described as follows:

$$r^w = \frac{l^{-1}\sum_{(i.j)\in L} w_{ij}\, k_i^w k_j^w - [l^{-1}\sum_{(i,j)\in L}\frac{1}{2}w_{ij}(k_i^w + k_j^w)]^2}{l^{-1}\sum_{(i,j)\in L}\frac{1}{2}w_{ij}((k_i^w)^2 + (k_j^w)^2) - [l^{-1}\sum_{(i,j)\in L}\frac{1}{2}w_{ij}(k_i^w + k_j^w)]^2}$$

where $k_i^w(m)$ and k_j^w are the numbers of links between i and j respectively and all the nodes in a given module m [62]. For instance, a positive assortativity coefficient implies that nodes with a particular degree are likely to be connected with nodes with a similar degree and that may lead to resilient core of mutually interconnected high-degree hubs. Otherwise, high-degree nodes are vulnerable and widely distributed.

The features extracted are presented in Table 2.1 and they are divided into local and global measures.

Table 2.1

Local and global features extracted in the feature extraction process

Local Features	**Global Features**
Betweeness centrality	Maximized Modularity
Within module degree z score	Transitivity
Participation coefficient	Assortativity
Eigenvector centrality	Characteristic path length
Clustering coefficient	Efficiency
Degree	Radius
Strength	Diameter
Eccentricity	Density
Efficiency	Number of Edges
	Number of Vertices

As we stated in the beginning, we extracted the features discussed above using the open source Matlab toolbox which contains Matlab functions in order to calculate all the measures we need. It is also necessary to mention that EEG spectrum ranges from 4 to 45 Hz therefore, we divide it into *seven frequency bands* that represent the brain rhythms, namely theta [4-7 Hz], alpha_1 [8-9 Hz], alpha_2 [10-11 Hz], sensor motor rhythm (smr) [12-14 Hz], beta [15-29 Hz] , gamma [30-45 Hz] and full spectrum (fs) [4-45 Hz]. Each one of the seven frequency bands is represented in MATLAB as a symmetrical and up-triangle coherence matrix, comprising of information related to the 14 EEG channels. In particular, coherence estimates the consistency of relative amplitude and phase between any pair of signals in a given frequency band and its value indicates whether a relationship between a pair of signals can be approximated by a linear transformation. Thus, seven 14x14 *weighted and undirected graphs* are generated based on the coherence matrix. Matrices are thresholded to 0, namely every negative number is transformed to 0. Then, using the Brain Connectivity toolbox package, each of these features presented in Table 2.1 is calculated. Local measures result in 14 values corresponding to the 14 different nodes that concern a specific graph while global measures concern the whole network and result in a single value. What's more, since the matrices are symmetric, we use only the upper-triangular matrix that represents the weights of the edges of the network that are considered equally important regarding the brain functional connectivity and this is why we take them into consideration in the feature extraction part. Each 14x14 graph corresponds to 105 weight values. Thus, for each of the 37 participants, we extract

$$(9 \text{ local } \times 14 \text{ values} + 10 \text{ global values}) \times 7 \text{ bands} \times 14 \text{ channels}$$
$$= 13.328 \text{ brain connectivity features}$$

&

$$105 \text{ weight values } \times 7 \text{ bands} \times 14 \text{ channels} = 10.290 \text{ weight features}$$

Finally, for each of the 37 participants, a total of **23.618** features are extracted. It is expected that many of these features are not related and do not provide a significant information concerning brain connectivity patterns. Therefore, it becomes obvious that feature selection is more than necessary in order to focus on the important features and enhance the computational speed of the project. Feature selection with Relief-F algorithm is described in the following section.

2.5 Feature Selection

When large amount of data needs to be processed, further feature subset selection might be necessary in supervised learning algorithms. The selection algorithms ought to be effective and accurate. R. Durgabai [63] describes why he proposes the *relief-F* algorithm instead of the original *relief* algorithm as better feature selection method. Firstly, he points out that relief algorithm has low evaluation accuracy since the feature weight vector is calculated under random and uncertain instances. In addition, it cannot deal with missing values data or more than two-class problems. Therefore, in order to make the result more stable and accurate and solve other matters, he recommends the relief-F extension.

The input of the algorithm is an attribute values vector for each training instance and the class value. Its output is defined as the estimation quality vector w of attributes.

The algorithm follows certain steps:

- An instance r_i is randomnly selected
- K of its nearest neighbors of the same class are detected (nearest hits h_j)
- K of its nearest neighbors of the different classes are detected (nearest misses $m_j(C)$)
- Estimation vector w is updated depending on the values ofr_i, h_j and $m_j(C)$. The update implies the average of contribution of nearest hits and misses. As for the misses, the contribution of each class is weighted with the prior probability of that class $P(C)$, already estimated from the training set. Additionally, a $diff_f()$ function [64] is used which calculates the distance of two samples for a feature f.
- Process repeated for m times.

The selection of k hits and misses differentiates the algorithm from the initial relief algorithm. It consists the step which leads to the greater robustness of the algorithm as far as the noise is concerned. Locality of the estimates is controlled by the user defined parameter, commonly set to 10. Changes in the differential function used in the update, deals with incomplete data and missing values are treated in a probabilistic way. In particular, the probabilistic method[3] ensures that even if an object from a specified class is randomly selected, it results in a more than zero probability of the prescribed kind.

Z. Wang et al. [64] worked on classification of high resolution remote sensing image and combined fuzzy classification with ReliefF feature selection. The overall accuracy resulted in 81.6% and kappa coefficient value reached 0.791 showing the enhancement of classification quality.

S. Gilbert Nancy et al. [65] worked on cancer classification comparing different feature selection methods, namely Fast Correlation Based Filter (FCBF), ReliefF, Random Selection and Support Vector Machine Recursive Feature Elimination incorporated with T-statistic (SVM-t-RFE). The results were really important regarding the final classification accuracy and the number of features selected for each algorithm. Specifically, ReliefF resulted in the smallest number of features selected that is 246 in contrast to FCBC (249), SVM-t-RFE (407) and Random (283). Additionally, reliefF accuracy reached 74.12% while the other methods did not achieve such effectiveness (FCBC=57.12%, SVM=t-RFE =53.44%, Random=61.23%). Last but not least, classification accuracy of KP-SVM resulted in nearly 65% *before* feature selection while *after* ReliefF feature selection reaches 73% accuracy!

In general, ReliefF is considered to be an effective algorithm. It selects good features [65], deals with various data types, it is not limited in case of continuous or discrete datasets and it is suitable for multi class problems [64]. It also increases the classifier's efficiency [65].

[3] https://en.wikipedia.org/wiki/Probabilistic_method

All the above reach the conclusion that Relief-F is the suitable feature selection algorithm and therefore, it is applied before the stage of classification. In particular, we use Relief-F to select the *10 best features* firstly for the HVHA and LVHA quadrants separately and then, for the fusion of them. This is a process repeated for a given dimension of personality. This leads to the ranked features that will be classified afterwards, namely 10 for each dimension and each case (HVHA, LVHA, both HVHA & LVHA) and consequently 30 for each of the 5 dimension of personality (Agreeableness, Neuroticism, Openness, Conscientiousness, Extroversion).

2.6 Classification

Introduction to Classification learner

The Classification Learner[4] is an application in the Matlab environment which is used in order to train models through supervised machine learning and then, classify the given data. The selection of this tool provides the opportunity to proceed in feature selection, validation schemes specification, model training and last but not least, result assessment. Furthermore, we are capable of choosing from several classification types, such as Support Vector Machines (SVMs), Decision Trees (DTs) or k-Nearest Neighbors (kNNs) and *ensemble methods* that is bagging, boosting or random space.

2.6.1 Support Vector Machines

In our project, the dominant classifier is **Support Vector Machine** (SVM). SVM offers a principled approach to machine learning problems and we select it since it is considered as a suitable algorithm for binary classification (High/Low) [66] that is the desired result in this type of research. In particular, the concept of SVMs is based on an independent training dataset and a discriminant function [67] which can correctly predict labels for newly acquired instances. In particular, when applying a discriminant classifier like SVM, less computational resources and training data are required especially for a multidimensional space. It also implies the solution of the convex optimization problem which leads to the same optimal hyperplane parameter in contrast to perceptrons or genetic algorithms which lead to various hyperplanes in order to minimize error during training. The best hyperplane for an SVM is the one with the largest margin between the two classes [66].

In general, M. Awad et al. [67] state that SVMs are considered to be probably the most popular machine learning approach for supervised learning because of their robustness, good generalization ability and unique global optimal solutions. Furthermore, L. Auria et al. [68] also contend their suitability for binary classification tasks and underline the capability of dealing with non-linearity via engineering a *kernel*.

[4] https://www.mathworks.com/products/statistics/classification-learner.html

The mathematical model of Linear SVM is defined as follows [69]. Given a training set $\{y_i, \vec{x}_i\}_{i=1}^{l}$, where the input $\vec{x}_i \in \mathbb{R}^n$ and the output $y_i \in \{-1, +1\}$. If there is a hyperplane dividing all the points $\vec{x}_i$ into groups correctly, the aim is to find the maximum distance between the hyperplane and the nearest point $\vec{x}_i$ from either group. The optimal hyperplane is the solution of the constraint optimization problem defined as :

$$\min \ \frac{1}{2}\|\vec{\omega}\|^2 + C\sum_{i=1}^{l}\xi_i$$

$$\text{s.t.} \quad y_i[(\vec{\omega}\cdot\vec{x}_i) + b] \geq 1, \ \ i = 1,2, \dots l$$

$$\xi_i \geq 0, \quad i = 1,2, \dots l$$

where $C > 0$ is the penalty parameter and $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_l)$is the slack variable. Lagrangian multiplier method transforms the problem into a dual problem:

$$\max \quad \sum_{i=1}^{l} a_i - \frac{1}{2}\sum_{i,j=1}^{l} a_i\, a_j y_i y_j\left(\vec{x}_i \cdot \vec{x}_j\right)$$

$$\text{s.t.} \ \sum_{i=1}^{l} a_i\, y_i = 0,$$

$$0 \leq a_i \leq C, i = 1,2, \dots l$$

where $a_i \geq 0$ are the Lagrangian multipliers of samples $\vec{x}_i$. Cases when $a_i = 0$ are not part of the solution. Therefore, the classification decision function is:

$$f(x) = \ sign\left[\sum_{i=1}^{l} a_i y_i (\vec{x}_i \cdot \vec{x}) + b\right]$$

When non-linearity is a part of the problem, the only difference compared to the linear model is that we firstly perform data mapping to another high-dimensional space H, using a non-linear mapping called Φ. After that, the linear model is used again to perform classification in the space H. The kernel function k that is introduced is a symmetric, semi-positive definite function satisfying the Mercer theorem and converts the classification decision function as:

$$f(x) = \ sign\left[\sum_{i=1}^{l} a_i y_i \boldsymbol{k}(\vec{x}_i \cdot \vec{x}) + b\right]$$

The *Gaussian* kernel function which is intended to measure the similarity between $\vec{x}$ and $\vec{x}_i$ and it is mostly used in our project is described as:

$$k(\vec{x}, \vec{x}_i) = \exp(-\gamma\|\vec{x} - \vec{x}_i\|^2), \text{ where } \gamma > 0$$

As it can be observed, the Gaussian kernel only depends on the Euclidean distance between $\vec{x}$ and $\vec{x}_i$ and it is based on the assumption that similar points are found close to each other in the feature space [70].

Moreover, a commonly used kernel function that we select to implement in our project is the *Cubic* kernel [70], namely a 3rd-degree polynomial kernel function respectively described as:

$$k(\vec{x}, \vec{x}_i) = (\vec{x} \cdot \vec{x}_i + 1)^3$$

In general, SVMs provide high accuracy rate and tend to avoid *overfitting*. Moreover, the "easy-to-decide" binary classification enhances the speed of the algorithm [66]. Furthermore, the convexity problem results into a unique optimal solution which also make the algorithm robust in contrast to other methods, such as Neural Networks that deliver multiple solutions. As for the kernel introduction is considered to be beneficial as far as the SVM performance is concerned [68]. At first, kernel increases SVMs' flexibility since it makes the classification type capable of dealing with non-linear and non-monotone data distributions. Apart from this, kernel SVMs provide a good generalization if parameters C and γ are properly tuned and enhance the algorithm's robustness, even when the training sample has some bias.

Therefore, taking into consideration the benefits of the Gaussian kernel, we choose to apply specific variants of Gaussian SVMs, namely Coarse Gaussian SVM, Medium Gaussian SVM and Cubic SVM. In particular, Medium Gaussian SVM performs medium distinctions with kernel scale set to $\sqrt{P}$ and it provides medium model flexibility while Coarse Gaussian SVM makes coarse distinctions between classes, with kernel scale set to $\sqrt{P} * 4$ and provides low model flexibility. In both cases, P is the number of predictors. All the above perform fast and deal with memory usage appropriately in binary classification tasks.

2.6.2 Decision Tree

Decision Tree is an algorithm that aims to find the optimal partitioning of the space of possible observations by performing subsequent recursive splits [71]. It is a method widely used in data mining applications [72] because of their simplicity and effectiveness [73]. In particular, M. Magnani et al. [73] state this algorithm generates understandable models, does not demand prior assumptions on data distributions and it is fast to build while B. Patel et al. [71] underline that it resembles a tree structure where nodes represent a test on an attribute, branches denote the outcome of a test and leaves represent the class labels [71]. R. Barros et al. [74] denote that a decision tree can be represented by a graph $G = (V, E)$ where Vstands for Vertices (nodes) and E stands for Edges, that satisfies specific criteria. We select this algorithm since it is robust to noise, it provides low computational cost and deals with redundant attributes [75].

Hunt's algorithm is the main approach to the Decision Tree technique [76]. Given that $\mathbf{X_n}$ is the set of training instances related to n node and $c = \{c_1, c_2, \dots c_k\}$ is the set of class labels concerning k classes, the algorithm is divided in two basic steps that concern the assignment of the instances in $\mathbf{X_n}$ [75][76]. At first, the stopping criterion is checked, namely the case that all the instances in $\mathbf{X_n}$ belong to the same c_i class. This implies that all leaf nodes need to be pure. If so, n is a leaf node with label c_i. If the stopping criterion is not satisfied, the algorithm checks whether the instances belong to more than one classes. If yes, an attribute test condition is selected in order to form partitioned instance subsets. Each outcome of the test corresponds to the generation of a child node to which the instances of $\mathbf{X_n}$ are splitted regarding the outcomes. Then, the algorithm is recursively iterated on each child, namely the splitting is repeated on all new nodes.

In order to assess the quality of the algorithm, it would be beneficial to measure the overall quality of a splitting [73]. M. Magnani et al. [73] consider Entropy at node n as

$$E(n) = \sum_{all\ classes} -p(c_i|n) \log_2 p(c_i|n)$$

where $p(c_i|n)$ denotes the percentage of records belonging to class c_i at node n. Therefore, the overall quality of a splitting is characterized by the weighted sum of Entropy of the new generated nodes, and thus it is possible to define the level of impurity [75]. If N is the total number of instances related to the parent and N_i is the total number of instances related to the i_{th} child node after splitting, the impurity of split node S can be defined as [73] :

$$I(S) = \sum_{i \in [0,k]} \frac{N_i}{N} E(n_i)$$

In general, it can be observed that this classification method presents several weaknesses. Although, the algorithm evolution has overcome these problems. For instance, the stopping criterion requires a class absolutely pure, a constraint that possibly leads to overfitting. Therefore, the stopping criterion could be slightly altered by stopping the tree growth when reaching a specific impurity level or by employing a pruning step [74]. This is a very significant step as long as it leads to the tree size reduction, the increase of its readability and prevents from overfitting [73]. Apart from this, the selection of the attribute test condition has been an ambiguous matter. Initially, Hunt employed a cost-driven function to deal with tree partitioning [76]. In the process, information theory based functions were applied [77].

In this project, we employ the Fine Tree method, a variant of Decision Tree which performs fast prediction, it is easy to interpret and provides high flexibility with maximum number of splits set to 100 [66].

2.6.3 k-Nearest Neighbors

The k-Nearest Neighbors (kNN) classifier is considered to be one of the simplest machine learning algorithms, suitable for pattern classification problems. It is a conventional non-parametric classifier that results in good performance for optimal values of k [78]. In particular, in this algorithm, an instance is classified by a majority vote of its neighbors and it is then assigned to the most common class among its kNN [79]. k is a positive integer, usually small, and in case $k = 1$, the instance is assigned to the class of the nearest neighbor.

C. Li et al. [79] describe the mathematical background as follows. Given a binary classification problem and a training set $S = (x_i, y_i)$ for $i = 1,2, \dots, N$, where x_i represents the d-dimensional feature vector and $y_i \in \{+1, -1\}$ is related to the observed class labels, the kNN algorithm constructs a local subregion $R(x) \subseteq \mathbb{R}^d$ of the d-dimensional input space, situated at the estimation point x. In particular, $R(x)$ is the predicting region which includes the closest training points to x and it is described as:

$$R(x) = \left\{\hat{x} | D(x, \hat{x}) \leq d_{(k)}\right\}$$

where $d_{(k)}$ and $D(x, \hat{x})$ are the k_{th} order statistic of $\{D(x, \hat{x})\}_1^N$ and the distance metric respectively. The samples in region $R(x)$ labeled y are denoted by the term $k[y]$. kNN estimates the posterior probability $p(y|x)$ of the observation point x that is

$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \cong \frac{k[y]}{k}$. The evaluation of $k[y]$ values and the selection of the highest $k[y]$ value class leads to the decision function $g(x) = \begin{cases} 1, & k\,[y = 1] \geq k\,[y = -1] \\ -1, & k[y = -1] \geq k[y = 1] \end{cases}$. Finally, kNN aims to maximize the posterior probability and thus, converts the decision function as:

$$g(x) = \operatorname{sign}(\operatorname{ave}_{x_i \in R(x)} y_i)$$

C. Li et al. [78] tested kNN algorithm in order to enhance the accuracy of clinical lymph node metastasis in gastric cancer. Results indicated that kNN is, in general, highly efficient (accuracy = 80.79%). Although, they also showed [78] that data normalization improved the classification accuracy of kNN, namely 83.68% while the employment of dimensionality reduction algorithms, that is combined PCA and LDA, significantly increased accuracy (96.33%).

A. Kataria et al. [79] compared kNN to Bayes algorithm and Euclidean distance. They indicated that kNN maintains its efficiency as far as Bayes algorithm is concerned, although Euclidean distance performed better. Despite its efficiency, kNN has high computational complexity and it is fully dependent on the training set. A. Kataria et al. [79] recommend the use of Genetic Algorithms to overcome these kind of issues.

P. Thanh Noi et al. [80] compared the performances of kNN, SVM and Random Forest (RF) classifiers with different training sample sizes regarding the same remote sensing images. Tuned parameters play an important role for the performance of all classifiers. As for kNN, k value is the key tuning parameter and results demonstrated that the lowest error was achieved when $k = 1$. This value of k is finally chosen as optimal since as k increases, kNN error also increases. Final outcomes indicated that SVM provides the highest accuracy (95.32%) on imbalanced datasets while kNN and RF score 94.59% and 94.7% respectively. Moreover, training sample size is inversely proportional to accuracy. As for the balanced datasets, SVM still outperforms (accuracy=95.29%) whereas kNN and RF score 94.59% and 94.1% respectively. Furthermore, different training sample sizes did not seriously affect the general performance, although kNN resulted in strongly decreased accuracy with small training sample sizes.

Therefore, considering the high accuracy and simplicity of kNN, we select to use it as one of our classification methods and in particular, Fine kNN and Medium kNN. Fine kNN provides finely detailed distinctions between classes and the k tuning parameter is set to 1, while Medium kNN provides medium distinctions between classes and the k tuning parameter is set to 10 [66]. They both provide medium prediction speed and memory usage, as well.

2.6.4 Ensemble methods

Ensemble methods are employed in order to solve a specific problem by training multiple learners, namely they aim to create a set of learners and combine them [81]. Z. Zhou suggests [81] that these methods are commonly preferred in pattern recognition problems since they provide large generalization ability and they transform weak base learners to strong learners which can make accurate predictions. This is also the reason why they are very useful in machine learning community and their low computational cost makes them even more appealing.

Ensemble Bagged Trees

Taking into consideration all the above, we select to use two ensemble methods in the classification process. First, the parallel ensemble method *Ensemble Bagged Trees* is selected, namely Random Forest (RF) classifier and bagging with decision tree learners. Parallel ensemble methods is the combination of base learners generated in parallel with *Bagging*, firstly introduced by Breiman in 1996 [82], as a representative. Bagging is the process when data subsets used to train base learners are obtained through bootstrap sampling [81]. Bagging is suitable for binary classification problems and aggregates its outputs using the technique of voting for classification. Breiman describes the mathematical background of bagging algorithm [82] as follows.

Given a learning set $\mathcal{L} = \{\boldsymbol{x}_n, y_n\}$ $n = 1, \dots N$ and a sequence of learning sets $\{\mathcal{L}_k\}$ that consist of N independent observations from the same $\mathcal{L}$ distribution, we aim to form a predictor $\varphi(\boldsymbol{x}, \mathcal{L})$ in order to predict y given the input vector $\boldsymbol{x}$. In case it is possible to work with a sequence of predictors $\varphi(\boldsymbol{x}, \mathcal{L}_k)$ from the same $\mathcal{L}$ distribution and not the single predictor $\varphi(\boldsymbol{x}, \mathcal{L})$, we prefer to do so in order to result in a more accurate predictor. Nevertheless, we usually have a single learning set $\mathcal{L}$ and thus, the desired predictor is formed as $\varphi(\boldsymbol{x}, \mathcal{L}^{(B)})$, where $\{\mathcal{L}^{(B)}\}$ represents the repeated boostrap samples obtained from $\mathcal{L}$. Now the output is

$$\varphi_B(x) = \mathrm{ave}_{\mathrm{B}} \varphi(\boldsymbol{x}, \mathcal{L}^{(B)})$$

If y stands for a class label, $\varphi(\boldsymbol{x}, \mathcal{L}^{(B)})$ *votes* for $\varphi_B(x)$. $\mathcal{L}^{(B)}$ generate data subsets that consist of N cases drawn with replacement. Each $\{\boldsymbol{x}_n, y_n\}$ can be absent or repeated many times in a specific $\mathcal{L}^{(B)}$. This process is **b**ootstrap **agg**regat**ing**, namely bagging. Stability in the process of φ generation is the dominant factor that seriously affects the algorithm's accuracy. More specifically, unstable procedures which imply that small changes in $\mathcal{L}$ lead to large changes in φ, provide higher accuracy and vice versa.

As far as the RF model is concerned, it was initially introduced by Breiman in 2001 [83] after several years of studies and it consists an extension over bagging. As he stated, a random vector produces the particular tree classifiers which in the process, classify the input vector. RF classifier is defined by M. Pal [84] as a tree-based classifier, namely an ensemble of classification trees [85]. More specifically, features are selected randomly from the whole variable set at each split [85]. In the process, each tree grows on different random subsamples in the training set performing the 'so-called' *bagging* and the splitter is also randomly determined [85]. As for the training, each tree has to grow till the maximum depth on the new data [84].

In particular, Breiman [83] describes thoroughly the procedure over, let's say, the k_{th} tree. At first, a random vector Θ_k is produced which is not related to the previous vectors $\Theta_1 \cdots \Theta_{k-1}$ although it has the same distribution. As it was mentioned, after that, a tree is grown via the training set and the random vector Θ_k. This tree results in a classifier $h(\mathbf{x}, \Theta_k)$ where $\mathbf{x}$ stands for the input vector. If N is the size of the examples in the initial training set, the random vector Θ is produced through bagging by replacing N examples and the new training set is generated. As for the random split selection, the random vector Θ can be defined as the number of random integers ranging from 1 to K. K parameter determines the incorporation of randomness. The tree construction determines the dimensionality of the random vector Θ and, in the end, after the tree generation of a large number, they vote for the corresponding class of belonging.

Zhou underlines [81] that decision boundaries of RF and its base classifiers provide greater flexibility and better generalization combined with Bagging as well as lower test error.

He also states that bag with pruned or unpruned decision tress does not affect the RF performance.

Ensemble subspace k-Nearest Neighbors

Random space (RS) method was firstly introduced by Ho in 1998 [86] and it implies the coexistence of various classifiers in a subspace of data feature space. The outputs of individual classifiers generate the final classification results through *majority voting* [87]. RS classifier is capable of reducing original data size while maintaining the training samples size. This leads to effective classification, mainly when a large number of features is processed. RS is considered to be a simplified and easy to interpret ensemble model which provides good generalization and avoids the risk of overfitting [87]. Therefore, we select to apply this RS method combined with k-Nearest Neighbors (kNN) classifier since this combination provides higher accuracy compared to the conventional kNN classifiers [86]. In particular, given a feature vector, RS method is the stochastic process which implies the random selection of components so as to generate each classifier. As far as kNN classifiers are concerned, they are produced using the projected distances, namely the distances computed by the projection of all the points in a selected subspace. For each random subspace, k nearest neighbors are generated and then, majority voting determines the corresponding class of the test sample. The mathematical background of Ensemble Subspace kNN is thoroughly described by T. Ho as follows [86].

Given a set $S = \{(x_1, x_2, \dots x_n) | x_i \text{ is real for all } 1 \leq i \leq n\}$ of N points in a feature space of n dimensions, Random Subspace defines m-dimensional subspaces described as

$$(x_1, x_2, \dots x_n)|x_i = \begin{cases} 1, & i \in I \\ 0, & i \notin I \end{cases}$$

Where I represents a particular m-element subset of $\{1,2,\dots,n\}$ $(m < n)$. Every iteration leads to the selection of a random subspace and the projection of all points onto this particular subspace. Each testing point results into k nearest neighbors $(1 \leq k \leq N)$ among the projected training points using Euclidean distance and a list C which includes $\{c_1, c_2, \dots, c_k\}$ class labels of the k nearest neighbors. The final class assignment of the instance x_i is the class label from the list C and it is based on majority voting.

T. Ho suggests that this technique is suitable for signal processing tasks and large-scale data mining applications. In general, high-dimensional feature spaces are followed by higher complexity compared to smaller feature spaces, an effect known as “the curse of dimensionality” [88]. Although, P. Mewada et al. [88] underline that the RS ensemble divide-and-conquer methodology is capable of dealing with this matter since it breaks down the initial high-dimensional problem to lower dimensional sub-problems and reduces the computational complexity.

2.6.5 5-*fold* Cross Validation

The accuracy estimation of a classifier is more than necessary in order to make the appropriate classifier selection and/or combination and predict its future accuracy. The classifier's performance is usually measured in terms of prediction error. The estimation method should be characterized of low variance and low bias (the expected value minus the estimated value) [89]. We select to apply one of the most common accuracy estimation methods, namely the K-fold cross validation.

T. Fushiki et al. [90] describes the mathematical background of K-fold cross validation as follows. Given a dataset $D = \{x_1, x_2, \dots x_N\}$ with N independent and identically obtained observations derive from a distribution F. The aim of cross validation is to construct a prediction model based on D and estimate its prediction accuracy or equivalently the prediction error. At first, we define $Z = (X, Y)$ and we use a set of $\{h(x;\theta)|\theta \in \Theta\}$ to explain Y by X in order to predict Y at X by $h(X;\hat{\theta})$. $\hat{\theta}$ is computed when the term $N^{-1}\sum_i (y_i - h(X;\hat{\theta}))^2$. Thus, the estimator can be considered as $\hat{\theta}(F_N) = \mathrm{argmin}_{\theta\in\Theta}\{\int \Psi(z;\theta)\mathrm{d}F_N(z)\} = \mathrm{argmin}_{\theta\in\Theta}\frac{1}{N}\sum_{i=1}^{N}\Psi(z_i;\theta)$, where F_N represents the empirical distribution of D and the prediction error is finally given by

$$E = \frac{1}{N}\sum_{i=1}^{N}\Psi\left(z_i;\hat{\theta}(F_N)\right)$$

The estimation of the prediction error is commonly performed by K-fold cross validation. This method splits the dataset into K equal sub-datasets $D^{(1)}, D^{(2)}, \dots, D^{(K)}$. Then, subset $D^{(a)}$ is removed from D and the subset acquired is $D^{(-a)} = D \backslash D^{(a)}$. We consider $m_a = |D^{(a)}|$ so that $\sum_a m_a = N$ and $p_a = m_a/N$. Thus, the prediction error estimate using K-fold cross validation is

$$\mathrm{CV}_{N,K} = \sum_{a=1}^{K} p_a \int \Psi\left(z;\hat{\theta}\left(F_{N,K}^{(-a)}\right)\right) \mathrm{d}\, F_{N,K}^{(a)}(z)$$

Where $F_{N,K}^{(a)}$ and $F_{N,K}^{(-a)}$ denote the empirical distributions of $D^{(a)}$ and $D^{(-a)}$ respectively.

This method implies the estimation of each θ based on a subset of D providing an upward bias. If $K = N$, the estimation method is known as leave-out-cross validation which may be unbiased however, it requires great computational cost regarding time processing. Generally speaking, K value is the parameter which plays the dominant role regarding the algorithm performance. In particular, moderate k values reduce the variance and increase the bias, while smaller k values and thus, smaller sample sizes lead to an increase in variance [91]. J. Rodriguez et al. [91] indicated that different numbers of folds did not result in significant differences regarding total variance. Moreover, as far as bias is concerned, they underline that k value equal to 2 leads to the largest bias and the lowest variance for classification compared to bigger values of k. It is, therefore, necessary to select the k value that provides the optimal trade-off between variance and bias concerning the given problem. Since we are interested in prediction error, we select $k = 5$, a k value that is less biased and provides lower computational cost.

Chapter 3

Results and Discussion

3.1 Dominant Features

In Chapter 2, we also described thoroughly the feature extraction procedure, underlining all the features we selected so as to gather valuable information for the brain's functional connectivity. Although, not all of them provide the same significant value and the feature selection leads to the most important features concerning the three scenarios mentioned before. In general, *weight values* of the edges that connect nodes in the graph appear to be the most dominant features which finally determine the functional connectivity concerning each of the five dimensions. As O. Sporns explained [92], weights in graph theory represent the density or efficacy of a connection and thus, it is clear why they consist the dominant characteristics. Brain connectivity features appear to be less important, although some of them tend to be valuable in specific traits recognition (**Figure 3**)[5]. An extensive table which presents all the best features for each scenario and personality dimension can be found in the appendix. The results analyzed below are based on this particular table.

As far as the HVHA scenario is concerned, the openness trait is highly correlated with betweeness centrality feature. In particular, *betweeness centrality* in the smr band dominates openness (8 over 10 best features). Moreover, within-*module degree z-score* in the gamma band slightly affects the conscientiousness trait (3/10 features). The same traits are affected in the LVHA scenario. In particular, nodal *eccentricity* in beta band is present in openness trait (1/10) while *participation coefficient* in alpha_1 band plays a significant role for conscientiousness trait (4/10). R. Guimera et al. [60] underline the significance of nodal participation coefficient and nodal strength. Specifically, they denote that participation coefficient is capable of facilitating the global integration between modules of a system and our work appears to agree with this finding. The third and most significant scenario regarding the prediction accuracy, that is the fusion scenario, results in the dominance of betweeness centrality in the smr band (7/10) concerning the openness trait. The latter is, in general, thought to be a trait hard to understand and interpret due to its controversy. Although, research was led to outcomes that add a significant value in the general understanding of openness. In particular, R. E. Beaty et al. [93] determined the role of openness through Default Network topology analysis. Results declared a strong correlation between openness and network *global efficiency* (β=0.25, P=0.03) while the other four personality traits did not affect the efficiency measure. Moreover, Q. Gao et al. [94] indicated a strong correlation between extraversion and

5 https://www.researchgate.net/figure/Emotiv-EPOC-headset-14-channel-placement-with-two-reference-channels_fig2_309427804

neuroticism and the AUC of nodal betweeness centrality as well as a significant correlation between extraversion and normalized clustering coefficient. In this work, we suggest a significant relation between betweeness centrality and openness trait detection. Besides, as far as the brain features are concerned, *betweeness centrality* is considered as a dominant feature. In general, increased nodal betweeness centrality implies enhanced coordination of brain networks. Brain regions characterized by high betweeness centrality play a dominant role in information transition and control the information flow [95].

Another important part concerning the features selected is that only local features are chosen by ReliefF algorithm while global measures do not appear to play an important role. The efficiency and the crucial role of local functional connectivity measures were lately confirmed by J. Xu et al. [96] who examined resting-state brain activity using fMRI. They provide evidence that task performance is strongly associated with nodal and not global efficiency as they expected. In general, it is suggested that intrinsic brain activity is capable of predicting human behavior. Apart from this, P. Taylor et al. [97] emphasized on the limitations of global network properties analysis and underlined the necessity of analyzing local regions such as within brain areas, namely modular organization. They denote the importance of modularity within and between brain areas in the further understanding of mental disorders. Therefore, we can now better understand why within-module degree z-score was one of the dominant features in the HVHA scenario.

Characteristic Brain Regions

It is easy to notice (**Figure 3**) that each personality dimension is associated with a particular brain region. Starting with agreeableness, L. Nummenmaa et al. [98] stated, that this trait can be detected in the posterior cingulate cortex, namely the upper part of the limbic lobe and the temporal lobe (superior temporal gyrus). In this work, we observe that this trait is related to brain activity in the frontal and the occipital lobe. Research suggests that neuroticism is associated with activity in the middle frontal gyrus [99], medial prefrontal cortex [100], anterior cingulated [101], temporal pole [102], amygdala [103] and the basal ganglia [104]. Furthermore, it has been documented that increased right-sided activity is related to neuroticism [105]. Apart from this, Zuckerman [106] has documented, through EEG study that neuroticism is associated with higher activation of right frontal lobe compared to the left. This study confirms the association between neuroticism and activity in the temporal lobe. It also suggests an increased theta activity in the occipital lobe while it contradicts the right frontal lobe activation that Zuckerman [106] and J. Spielberg et al. [105] emphasized on since left side is now dominating. As far as conscientiousness is concerned, J. Tangi et al. [107] underline that it is covaried with volume in lateral prefrontal cortex. This statement is confirmed by our study which also denotes an activity in both parietal lobes.

The largest amount of information is provided for the extraversion trait. In particular, study conducted by H. Cremers et al. [108] resulted in positive correlation ($p_{\mathrm{FWE}} < .05$ corrected for extent of ROI) between both right medial orbitofrontal cortex and right centromedial amygdala and extraversion trait. Furthermore, J. Spielberg underlined that extraversion which is associated with left-sided activity [105]. As for the frequency bands, Tran et al. [109] resulted in greater amplitude of alpha wave (8-13 Hz) for extroverts in contrast to low levels associated with introverts and greater delta and theta activities associated with high extraversion in males during resting state EEG [110]. T. Johannisson [111] stated that alpha frequency resulted in significant negative correlation with extraversion (-0.16, $p < 0.05$). High degree of extraversion is present in the center of the 8 Hz frequency group but not in the surrounding zones. Apart from this, Hagemann et al. [112] used EEG alpha activity and MRI method to prove that skull thickness and external factors do not affect the positive correlation between extraversion and alpha frequency, as T. Johannisson has already stated [111]. In this work, we confirm the increased brain activity related to the frontal and left temporal lobe. Although, except for the significant alpha band frequency, we notice an interesting theta band activity which can be considered equally significant.

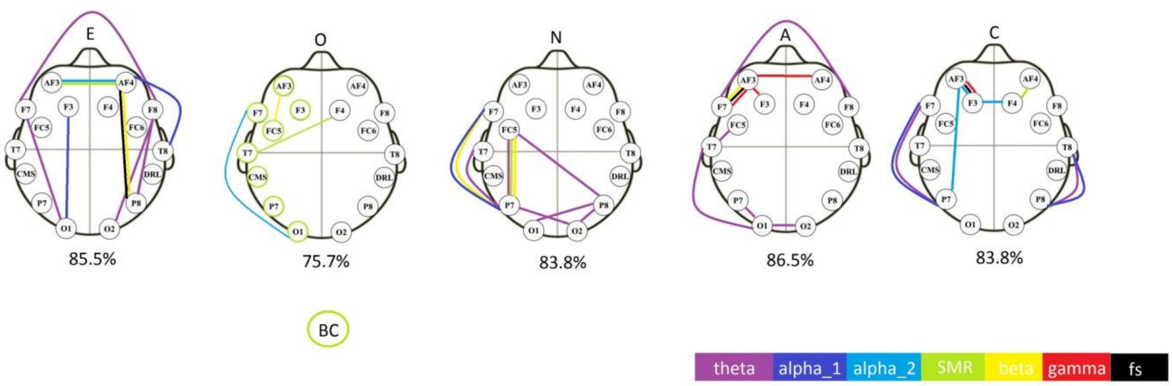

Figure 3

This figure presents the dominant features concerning each dimension of the Big-Five. Links denote only the weight values that connect specific edges while colors represent the band a particular feature corresponds to.

3.2 Best Aggregated Results

In Chapter 2, we realized a thorough analysis of the algorithms tested in the classification procedure. Taking into consideration the suitability and the performance of SVMs, this algorithm dominates the final results. Apart from this, results concern three different scenarios as it was stated in the proposed methodology, namely HVHA scenario, LVHA scenario and fusion of them. Assessment of results is based on four main parameters, that is the *Accuracy*, *Sensitivity*, *Specificity* and the *Area Under the Curve* (AUC). In particular, the prediction of each trait is binarized to low and high, thus we consider the successful prediction of a low and high trait as True Positive (TP) and True Negative (TN) respectively. If the classification fails for the low trait, we consider the sample as a False Negative (FN) and in case of high trait failure, we result into a False Positive (FP) sample. Therefore, accuracy is defined as

$$acc = \frac{\text{TP} + \text{TN}}{\text{Total Population}}$$

while sensitivity is the TP rate given by : $\frac{\text{TP}}{\text{TP+FN}}$ and specificity is the TN rate: $\frac{\text{TN}}{\text{TN+FP}}$. AUC denotes the prediction or not concerning the binarized classes generated by k -means clustering and it is important as far as non-homogeneous classes are concerned. Tables 3.2.1, 3.2.2 and 3.2.3 present the best aggregated results regarding the three scenarios. The dominant classifier is SVM, although in some cases other classifiers provide higher accuracy.

R. McCrae et al. [113] examined the positive and negative valence dimensions from the perspective of the Big-Five model. They stated that high valence is a definer of extroversion factor. Indeed, in the HVHA scenario, as it can be noticed from Table 3.2.1, extroversion is one of the strongest predictors resulting in 83.8% accuracy while conscientiousness scores the highest accuracy (86.5%). Openness and neuroticism score lower in accuracy, although openness is also described by a low AUC value and we could consider it as the weakest Big-Five predictor. What's more, openness sensitivity is equal to zero which means that the method fails completely in the detection and classification of the low openness trait. This could make sense if we consider that, in general, researchers have described openness as the most controversial trait among the five dimensions. Besides, R. McCrae thoroughly examined the openness trait and suggests that it is a factor characterized by unconventionality, thin mental boundaries and intuition [114].

Table 3.2.1

Best Aggregated Results in HVHA Scenario

Functional Connectivity		HVHA				
Meta-trait	Dimension/Trait	Accuracy	Sensitivity	Specificity	AUC	Algorithm
Plasticity	Extroversion	83.8%	0.82	0.85	0.86	Medium Gaussian SVM
	Openness	73%	0	1	0.49	Coarse Gaussian SVM
Stability	Neuroticism	78.4%	0.92	0.54	0.84	Medium Gaussian SVM
	Agreeableness	75.7%	0.9	0.56	0.82	Medium Gaussian SVM
	Conscientiousness	86.5%	0.77	0.92	0.83	Fine Tree

Table 3.2.2 indicates that the LVHA scenario could be characterized as the "weakest" one since it presents the lowest accuracy scores regarding the three scenarios. We could say that this is a reasonable fact as long as low valence videos are expected to generate less spontaneous affective responses, or let's say more neutral. In fact, as C. Lithari et al. [52] have observed, negative (low) valence is difficult to be perceived by human brain and it implies the activity of multiple brain regions increasing the process complexity. Although, we manage to result in encouraging accuracy rates which confirm that high arousing stimuli can lead us to

credible and important outcomes [52]. More specifically, agreeableness and openness are the strongest predictors with accuracy rate equal to 73% for both of them. R. McCrae et al. [113] stated that negative valence is associated with agreeableness and conscientiousness and this is confirmed in the LVHA scenario as long as, apart from agreeableness, conscientiousness provides the second highest accuracy (70.3%). As for the openness trait, it performs better in the LVHA scenario compared to the HVHA scenario if we consider the current sensitivity and AUC parameter.

Table 3.2.2

Best Aggregated Results in LVHA Scenario

Functional Connectivity		LVHA				
Meta-trait	Dimension/Trait	Accuracy	Sensitivity	Specificity	AUC	Algorithm
Plasticity	Extroversion	59.5%	0.59	0.6	0.57	Medium Gaussian SVM
	Openness	73%	0.5	0.81	0.64	Fine Tree
Stability	Neuroticism	67.6%	0.75	0.54	0.73	Cubic SVM
	Agreeableness	73%	0.76	0.69	0.75	Linear SVM
	Conscientiousness	70.3%	0.54	0.79	0.69	Ensemble Subspace kNN

The two scenarios fusion clearly improves the results regarding all the parameters we examine. As we can observe from Table 3.2.3, the fusion provides the highest accuracy rates for four of the Big-Five dimensions except for conscientiousness which results in a slightly decreased accuracy rate (83.8% vs 86.5%). The lowest sensitivity parameters are detected in openness and conscientiousness traits while the respective AUC values could be considered satisfying.

Table 3.2.3

Best Aggregated Results in Fusion Scenario

Functional Connectivity		Fusion				
Meta-trait	Dimension/Trait	Accuracy	Sensitivity	Specificity	AUC	Algorithm
Plasticity	Extroversion	85.5%	0.82	0.9	0.86	Fine kNN
	Openness	75.7%	0.5	0.85	0.76	Ensemble Bagged Trees
Stability	Neuroticism	83.8%	0.91	0.69	0.79	Medium Gaussian SVM
	Agreeableness	86.5%	0.9	0.81	0.92	Medium Gaussian SVM
	Conscientiousness	83.8%	0.62	0.88	0.77	Medium Gaussian SVM

3.3 Best Results for the Dominant Classifier

In the previous section, we presented the best results provided by the performance of different classifiers. However, it is important to select one specific classifier and check its performance in order to result in homogeneous outcomes. Therefore, we select the Medium Gaussian SVM, which outperformed in most cases as far as the aggregated results are concerned.

If we observe the following tables, we can state that the classifier outcomes reveal a similar behavior compared to the aggregated results. In particular, accuracy scores are also high regarding almost every personality dimension, a fact that makes this classifier a promising choice. As for the HVHA scenario, the least accurate predictor concerns the openness trait, as we expected while a decrease in accuracy can be observed for the conscientiousness trait (67.6% vs 86.5%). In the LVHA scenario, a significant decrease in accuracy rate concerning the conscientiousness trait can be again detected, whereas openness is now the trait that consists the strongest predictor. Although, the TP rate is equal to zero, namely the sensitivity factor is not satisfying. Last but not least, the fusion scenario provides again the highest accuracy rates revealing slight differences compared to the best aggregated results. We observe an overall high performance, with 4 out of the 5 dimensions resulting in accuracy rate above 80%. Openness is the trait which scores lower in accuracy (73%).

Table 3.3.1

Results in HVHA Scenario with Medium Gaussian SVM

Functional Connectivity		HVHA			
Meta-trait	Dimension/Trait	Accuracy	Sensitivity	Specificity	AUC
Plasticity	Extroversion	83.8%	0.82	0.85	0.86
	Openness	64.9%	0	0.89	0.67
Stability	Neuroticism	78.4%	0.92	0.54	0.84
	Agreeableness	75.7%	0.90	0.56	0.82
	Conscientiousness	67.6%	0.15	0.96	0.73

Table 3.3.2

Results in LVHA Scenario with Medium Gaussian SVM

Functional Connectivity		LVHA			
Meta-trait	Dimension/Trait	Accuracy	Sensitivity	Specificity	AUC
Plasticity	Extroversion	56.8%	0.47	0.65	0.55
	Openness	73%	0	1	0.31
Stability	Neuroticism	64.9%	0.96	0.08	0.64
	Agreeableness	70.3%	0.81	0.56	0.71
	Conscientiousness	62.2%	0.31	0.79	0.72

Table 3.3.3

Results in Fusion Scenario with Medium Gaussian SVM

Functional Connectivity		Fusion			
Meta-trait	Dimension/Trait	Accuracy	Sensitivity	Specificity	AUC
Plasticity	Extroversion	83.8%	0.82	0.85	0.90
	Openness	73%	0.1	0.96	0.74
Stability	Neuroticism	83.8%	0.92	0.69	0.79
	Agreeableness	86.5%	0.90	0.81	0.92
	Conscientiousness	83.8%	0.6	0.88	0.77

3.4 High Valence Scenario Dominance

Valence is a term commonly used in psychology and especially in emotion discussion. The emotions are characterized by positive or negative valence, namely a pleasant or unpleasant feeling, regarding the events or situations they are generated by. Furthermore, valence describes the hedonic tone[6] of feelings and affect. Same valence emotions may similarly influence peoples' judgments and choices and in general, affect the human behavior. If we take a closer look at the Fusion scenario in the appendix, we observe that the high valence scenario is dominating in every dimension (it is denoted by number (1)). This can be explained taking into consideration an extensive research conducted by L. Barrett [115] who considers valence an important factor, if not the most important, in the determination of emotional states.

In particular, L. Barrett has stated [115] that valence consists an invariant part of the emotional experience as well as a fundamental measure of emotional responses. She also underlines that an affective system based on valence and its intensity can be regarded as the main corpus of emotional life. As for individuals, a significant variant of valence is the valence focus that defines the aspect of experience a person emphasizes on. For instance, some people may be highly valence focused, namely they focus on the valence that an emotional experience produces or they can focus on other properties such as arousal. L. Barrett [115] also suggests that neuroticism and extraversion are the two dimensions more related to valence since she considers them as the main indicators of sensitivity to valence information. In this work, we present an overall good performance concerning the Big-Five traits since high sensitivity and accuracy are provided in the whole valence scale, especially in high valence. An exception could be the openness trait, which remains a strong predictor although, in some cases it is characterized by negligible sensitivity. What's more, in her study, she indicates that high valence which is strongly associated with the focus on the hedonic

[6] https://en.wikipaideia.org/wiki/Valence_(psychology)

content leads to increased psychological responses to positive and negative information and stimuli. This is confirmed in our work since we suggest that all dimensions are strongly related to high valence with openness and neuroticism to be the dimensions exclusively dominated by the high valence scenario.

In general, neutral is considered as the baseline emotional state while other more pronounced emotional states (i.e. joy, fear) are characterized by higher valence [116] and lead to clear and intense emotional responses. High valence appeared to be associated with positive mood as Wadlinger et al. [117] demonstrated. Specifically, individuals induced into positive mood responded spontaneously to peripheral stimuli compared to neutral individuals. Therefore, it becomes obvious that valence is a useful measure in the scientific study of emotional processing and more specifically, high valence constitutes a significant emotion stimulator which leads to accurate and strong predictions.

Chapter 4

4.1 Conclusion and Future Work

We have presented an extensive and innovative research which has as main purpose the personality traits detection and prediction using EEG physiological signals. We focus on the EEG modality since it suits the requirements of an effective and affordable Human-Computer Interaction system especially when the signals are recorded using a low cost device such as the Emotiv Epoc.

In particular, we perform EEG signal processing and detection of connectivity patterns through EEG functional connectivity, a significant technique related to the field of Personality Neuroscience which is attracting more and more interest. The real innovation in this project is the successful personality detection through the concept of emotional processing since it has been documented that resting-state EEG cannot lead to personality detection. The AMIGOS dataset is the most suitable in this work including multiple physiological signals recorded during the display of affective videos as well as completed personality questionnaires and personality scores. We focus on the EEG signals recorded by the low-cost Emotiv Epoc (14 channels) and the mean personality scores. K-means clustering uses the mean personality scores to produce binarized (high/low) clusters for each personality dimension. EEG signals concern only the videos characterized by high arousal which is considered to enhance the efficiency of brain networks and thus, we create 3 scenarios of interest, namely High Valence-High Arousal (HVHA), Low Valence-High Arousal (LVHA) and Fusion scenario. We extract edge weights and brain connectivity features using the Brain Connectivity Toolbox implementation in Matlab Software. ReliefF algorithm selects the 10 best features regarding each personality dimension and scenario separately (30 features/ dimension). Based on the features selected, the classification process follows which results in significant outcomes.

First of all, taking into consideration the best aggregated results, we are led to strong and accurate personality predictors since 4 out of 5 Big-five traits provide accuracy above 80% in the Fusion scenario while openness is the trait with the lowest though important accuracy rate (75.7%). Furthermore, HVHA scenario performs better than LVHA scenario and thus, it is the dominant in the Fusion scenario. This is reasonable if we consider that high valence leads to more spontaneous and intense affective responses regarding positive or negative stimuli compared to low valence. Apart from this, we observe Support Vector Machine classifier dominance in all three scenarios and this is why we also present the best results using Medium Gaussian SVM, except for the best aggregated results. These outcomes indicate a similar behavior regarding the successful or not personality prediction and they preserve homogeneity. Therefore, they are considered equally important. Last but not least, best features concern mostly edge weights, which describe efficacy or density of connections, and specific local brain features such as betweeness centrality, participation coefficient and within-module degree z-score. Openness is strongly related to betweeness centrality and as for the brain regions, we observe an increased left parietal and frontal activity.

Despite the overall good performance, there are some points in our work which need further improvement. Namely, future work may concern the enhancement of the ambiguous openness trait prediction or the attempt to increase low valence stimulation. This can be achieved through different stimuli display under the promising concept of emotional processing or the fusion of multiple modalities and neuroscientific methods. Finally, the results can be improved using alternative classifiers, different software tools.

Appendix

This appendix presents extensively the features selected using the Relief-F algorithm concerning each scenario and personality dimension. The first indicator is the frequency band corresponding to a particular feature while the second indicator represents the edge connecting two nodes. The first table concerns the HVHA scenario, the second table describes the LVHA scenario. The third and last table concerns the Fusion scenario where the number in parenthesis denotes the scenario that gives each feature and indicates the dominance of High Valence-Scenario (1).

HVHA (1)				
Extroversion	**Openness**	**Neuroticism**	**Agreeableness**	**Conscientiousness**
Alpha_1 edge 1_8	Smr edge 5_12	Beta edge 4_6	Gamma edge 1_14	Gamma edge 5_6
Alpha_2 edge 1_14	Beta edge 1_4	Theta edge 4_9	Beta edge 1_2	Gamma edge 5_7
Alpha_1 edge 10_14	Smr BC node 1	Smr edge 4_6	Smr edge 2_5	Beta edge 5_7
Theta edge 8_13	Smr BC node 2	Theta edge 4_6	Fs edge 1_2	Theta edge 5_6
Alpha_1 edge 8_9	Smr BC node 3	Alpha_1 edge 2_6	Gamma edge 1_3	Alpha_2 edge 3_7
Gamma edge 3_13	Smr BC node 4	Theta edge 7_9	Theta edge 5_6	Smr edge 3_7
Theta edge 8_14	Smr BC node 5	Theta edge 6_9	Theta edge 9_10	Gamma edge 9_14
Alpha_2 edge 5_7	Smr BC node 6	Theta edge 6_8	Theta edge 7_8	Gamma Z_node 1
Alpha_2 edge 6_7	Smr BC node 7	Beta edge 2_6	Theta edge 7_12	Gamma Z_node 2
Gamma edge 7_9	Smr BC node 8	Theta edge 8_9	Gamma edge 1_2	Gamma Z_node 3
LVHA (2)				
Extroversion	**Openness**	**Neuroticism**	**Agreeableness**	**Conscientiousness**
Alpha_ edge 2_9	Alpha_1 edge 9_11	Gamma edge 1_14	Theta edge 4_5	Fs edge 6_7
Theta edge 1_11	Beta edge 5_11	Alpha_2 edge 6_14	Alpha_2 edge 8_13	Fs edge 3_4
Alpha_2 edge 1_11	Beta edge 11_13	Theta edge 7_8	Alpha_2 edge 2_9	Gamma edge 1_3
Theta edge 2_10	Fs edge 9_11	Theta edge 6_8	Alpha_1 edge 1_3	Fs edge 10_13
Theta edge 11_13	Alpha_2 edge 11_14	Alpha_2 edge 12_14	Alpha_2 edge 5_8	Alpha_1 edge 2_6
Smr edge 1_2	Smr edge 7_11	Alpha_2 edge 4_12	Beta edge 1_14	Gamma edge 2_3
Smr edge 2_11	Alpha_2 edge 2_9	Alpha_1 edge 12_14	Alpha_1 edge 3_6	Alpha_1 PC_node 1
Gamma edge 7_11	Theta edge 7_11	Theta edge 1_10	Alpha_2 edge 4_8	Alpha_1 PC_node 2
Alpha_1 edge 2_11	Theta edge 10_12	Alpha_2 edge 2_14	Theta edge 3_4	Alpha_1 PC_node 3
Gamma edge 11_14	Beta ECC_node 1	Alpha_2 edge 6_12	Theta edge 2_13	Alpha_1 PC_node 4
Fusion				
Extroversion	**Openness**	**Neuroticism**	**Agreeableness**	**Conscientiousness**
(1)Alpha_2 edge 1_14	(1)Beta edge 1_4	(1)Beta edge 4_6	(1)Gamma edge 1_14	(1)Theta edge 9_10
(1)Alpha_1 edge 10_14	(1)Alpha_2 edge 2_7	(1)Theta edge 4_6	(1)Gamma edge 1_3	(2)Alpha_1 edge 2_6
(1)Theta edge 8_13	(1)Smr edge 5_12	(1)Theta edge 6_8	(1)Beta edge 1_2	(1)Theta edge 2_6
(2)Theta edge 2_7	(1)Smr BC_node 1	(1)Beta edge 2_6	(2)Theta edge 5_7	(1)Alpha_2 edge 1_6
(2)Alpha_1 edge 3_7	(1)Smr BC_node 2	(1)Theta edge 8_9	(2)Theta edge 4_5	(2)Alpha_2 edge 3_14
(1)Beta edge 9_14	(1)Smr BC_node 3	(1)Theta edge 7_9	(1)Fs edge 1_2	(1)Smr edge 12_14
(2)Theta edge 2_10	(1)Smr BC_node 4	(1)Alpha_1 edge 2_6	(1)Gamma edge 1_2	(2)Gamma edge 1_3
(1)Fs edge 9_14	(1)Smr BC_node 5	(1)Smr edge 4_6	(1)Theta edge 7_8	(1)Fs edge 1_3
(1)Smr edge 1_14	(1)Smr BC_node 6	(1)Theta edge 4_9	(2)Theta edge 2_13	(2)Alpha_2 edge 1_3
(1)Theta edge 9_13	(1)Smr BC_node 7	(1)Theta edge 5_6	(2)Theta edge 6_7	(1)Alpha_1 edge 9_10

REFERENCES

[1] A. Vinciarelli, “Personality Computing: How Machines Can Deal With Personality Traits,”MAPTRAITS ’14 Proceedings of the 2014 Workshop on Mapping Personality Traits Challenge and Workshop, 2014.

[2] V. Ozbek, U. Alniacik, F. Koc, M. E. Akkilic, E. Kas,”The impact of personality on Technology Acceptance : A Study on Smart Phone Users,”Procedia – Social and Behavioral Sciences, vol. 150, pp. 541-551, 2014.

[3] D. Ozer and V. Benet-Martinez, ”Personality and the prediction of consequential outcomes,” Annual Reviews of Psychology, vol. 57, pp. 401-421, 2006.

[4] A. Vinciarelli and G. Mohammadi,”More Personality in Personality Computing,”IEEE Transactions on Affective Computing, vol. 5, no. 3, 2014.

[5] A. Vinciarelli and G. Mohammadi,”A Survey of Personality Computing,”IEEE Transactions on Affective Computing, 2014.

[6] R. R. McCrae, O. P. John,”An introduction to the Five-Factor Model and its applications,”Journal of Personality, vol. 60, no.2, pp. 175-215, 1992.

[7] Rosalind W. Picard: Affective Computing. Cambridge, MA: MIT Press, 1997, xii + 292 pp.

[8] N. Al Moubayed, Y. Vazquez-Alvarez, A. McKay, and A. Vinciarelli,”Face-Based Automatic Personality Perception,”Proceedings of the ACM International Conference on Multimedia – MM ’14,vol. 60, no. 2, pp. 1153-1156, 2014.

[9] S. E. Kahou, X. Bouthillier, P. Lamblin et al.,”EmoNets:Multimodal deep learning approaches for emotion recognition in video,”Journal on Multimodal User Interfaces, vol. 10, no. 2, pp. 99-111, 2016.

[10] A. Patwardhan and G. Knapp,”Affect Intensity Estimation using Multiple Modalities, pp. 130-133, 2013.

[11] M. Pantic, N. Sebe, J. F. Cohn, T. Huang,”Affective multimodal human-computer interaction,”MULTIMEDIA ’05 Proceedings of the 13th annual ACM International conference on multimedia, pp. 669-676, 2005.

[12] J. M. Carroll,”Human-computer interaction: psychology as a science of design,”Annual Review of Psychology, vol. 48, no. 1, pp. 61-83, 1997.

[13] H. Ronen and D. Teeni,” The Impact of HCI Design on Health Behavior: The Case for Visual, Interactive, Personalized-content (VIP) Feedback,”ICIS 2013 Proceedings, pp. 1-18, 2013.

[14] P. Ekhman,”Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life, New York, Henry Holt, 2007.

[15] E. B. Foa and M. J. Kozak,”Emotional processing of Fear: Exposure to Corrective Information,”Psychological Bulletin, vol. 99, pp. 20-35, 1986.

[16] J. D. Teasdale,”Emotional processing, three modes of mind and the prevention of relapse in depression,” Behaviour Research and Therapy, vol. 37, pp. 53-77, 1999.

[17] D. Szczygiel, J. Buczny, R. Bazinska,”Emotion regulation and emotional information processing: The moderating effect of emotional awareness,”Personality and Individual Differences, vol. 52, pp.433-437, 2012.

[18] M. Vuorela and L. Nummenmaa,”Experienced emotions, emotion regulation and student activity in a web-based learning environment,”European Journal of Psychology of Education, vol. 19, no. 4, pp. 423-436, 2004.

[19] A.Vinciarelli, M.Pantic, and H.Bourlard, “Social Signal Processing: Survey of an Emerging Domain“, Image and Vision Computing Journal, Vol. 27, no. 12, pp. 1743-1759, November 2009.

[20] N. Jadhav, R. Manthalkar and Y. Joshi,”Effect of meditation on emotional response: An EEG-based study”, Biomedical Signal Processing and Control, vol. 34, pp. 101-113, 2017.

[21] L. Qiu, H. Lin, J. Ramsay, and F. Yang,” You are what you tweet : Personality expression and perception on Twitter,”Journal of Research in Personality, vol. 46, no. 6, pp. 710-718, 2012.

[22] M. Skowron, M. Tkalcic, B. Ferwerda and M. Schedl,”Fusing social media cues: Personality Prediction from Twitter and Instagram,”Proceedings of the 25^{th} International Conference Companion on World Wide Web,pp. 107-108, 2016.

[23] G. M. Chen,”Revisiting the social enhancement hypothesis : Extroversion indirectly predicts number of Facebook friends operating through Facebook usage,”Computers in Human Behavior, vol. 39, pp.263-269, 2014.

[24] D. Quercia, D. Las Casas, J. Pesce, D. Stillwell, M. Kosinski, V. Almeida, and J. Crowcroft,” Facebook and Privacy : The Balancing Act of Personality, Gender, and Relationship Currency,”Procceedings of the 6^{th} International AAAI Conference on Weblogs and Social Media, pp. 367-370, 2010.

[25]G. Chittaranjan, J. Blom and D. Gatica-Perez,”Mining large-scale smartphone data for personality studies,”Personal and Ubiquitous Computing, vol. 17, no. 3, pp. 433-450, 2013.

[26] J. Atkinson and D. Campos,”Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers,”Expert Systems With Applications : An International Journal, vol. 47, pp. 35-41, 2016.

[27] M. Abadi, J. Correa, J. Wache, H. Yang, I. Patras, and N. Sebe,"Inference of personality traits and affect schedule by analysis of spontaneous reactions to affective videos,"2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition,2015.

[28] J. Wache, R. Subramanian, M. Abadi, R. Vieriu, N. Sebe, and S. Winkler," Implicit User-centric Personality Recognition Based on Physiological Responses to Emotional Videos," *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15, pp. 239-246, 2015.*

[29] A. V. Bocharov, G. G. Knyazev, and A. N. Savostyanov,"Depression and implicit emotion processing: An EEG study,"Clinical Neurophysiology, vol. 47, no. 3, pp. 225-230, 2017.

[30] B. M. Herbert, O. Pollatos, and R. Schandry,"Interoceptive sensitivity and emotion processing : An EEG study,"International Journal of Psychophysiology,vol. 65, pp. 224-227, 2007.

[31] N. Zollinger,"The emotional processing of music : a high-density EEG study, 2013.

[32] K. E. Guiseffi,"Processing Facial Emotions: An EEG Study of the Differences between Conservatives and Liberals and Across Political Participation,"Political Science Department – Theses,Dissertations, and Student Scholarship, 2012.

[33] V. C. Pezoulas, M. Zervakis, S. Michelogiannis,and M. A. Klados,"Resting-State Functional Connectivity and Network Analysis of Cerebellum with Respect to Crystallized IQ and Gender,"Frontiers in Human Neuroscience, 2017.

[34] H. Park and K. Friston,"Structural and functional brain networks: From connections to cognition,"Science, vol. 342, no. 6158, 2013.

[35]L. Pessoa,"Understanding brain networks and brain organization,"Physics of Life Reviews, vol. 11, no. 3, pp. 400-435, 2014.

[36] C. Frantzidis, A. Vivas, A. Tsolaki, M. Klados, M. Tsolaki, and P. Bamidis,"Functional disorganization of small-world brain networks in mild Alzheimer's disease and amnestic Mild cognitive impairment: An EEG study using Relative Wavelet Entropy (RWE),"Frontiers in Aging Neuroscience, vol. 6, pp. 1-11, 2014.

[37] C. G. DeYoung and J. R. Gray,"Personality Neuroscience : Explaining Individual Differences in Affect, Behavior and Cognition,"In Corr, P. J., & Matthews, G. (Eds.) The Cambridge Handbook of personality psychology, pp. 323-346, New York : Cambridge University Press, 2009.

[38] N. Birmauer, C. Weber, C. Neuper, E. Buch, K. Haapen, and L. Cohen,"Chapter 24 Physiological regulation of thinking: brain-computer interface (BCI) research,"Progress in Brain Research, vol. 159, pp. 369-391, 2006.

[39] K. Korjus, A. Uusberg, H. Uusberg, N. Kuldkepp, K. Kreegipuu, J. Allik, R. Vicente, and J. Aru,"Personality cannot be predicted from the power of resting state EEG,"Human Neuroscience, 2015.

[40] A. Vinciarelli,"Social Perception in Machines: The Case of Personality and the Big-Five Traits, pp. 151-164, 2016.

[41] C. DeYoung,"High-order factors of the big five in a multi-informant sample,"Journal of Personality and Social Psychology, vol. 91, no. 6, pp. 1138-1151, 2006.

[42] A. Wright,"Current directions in personality science and the potential for advances through computing,"IEEE Transactions on Affective Computing, vol. 5, no. 3, pp. 292-296, 2014.

[43] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups," ArXiv e-prints, 2017.

[44] F.Citron, M. Gray, H. Critchley, B. Weekes, and E. Ferstl," Emotional valence and arousal affect reading in an interactive way: Neuroimaging evidence for an approach-withdrawal framework,"Neuropsychologia, vol. 56, pp. 79-89, 2014.

[45] C. Schonfeldt-lecuona, U. Herwig, and P. Satrapi,"Using the International 10-20 EEG system for Positioning of Transcranial Magnetic Stimulation,"Brain Topography, vol. 16, no. 2, pp. 95-99, 2003.

[46] S. Al-Harbi, and V. Rayward-Smith,"Adapting *k*-means for supervised clustering,"Applied Intelligence, vol. 24, no. 3, pp. 219-226, 2006.

[47] S. Lee, J. Kim, K. Kim, S. Park,and W. Moon,"K-means clustering approach for kinetic pattern analysis of dynamic contrast enhancement breast MRI ,"Proceedings of Asia Pacific Association for Medical Informatics (APAMI) Conference 2006, pp. 761-764, 2006.

[48] A. Gronlund, K. Larsen, A. Mathiasen, J. Nielsen, S. Schneider, and M. Song," Fast Exact k-Means, k-Medians and Bregman Divergence Clustering in 1D, pp. 1-16, 2017.

[49] B. Anderson, D. Gross, D. Musicant, A. Ritz, T. Smith, and L. Steinberg," Adapting k-medians to generate normalized cluster centers," Proceedings of the Sixth SIAM International Conference on Data Mining, pp. 165-175, 2006.

[50] P. Cherdchu, and E. Chambers,"Personality Classification of Consumers: A Comparison of Variables, Standardization and Clustering Methods,"Journal of Sensory Studies, vol. 28, no. 6, pp. 504-512, 2013.

[51] A. Jain,"Data Clustering: 50 years beyond K-means,"19th International Conference in Pattern Recognition (ICPR), pp. 651-666, 2010.

[52] C. Lithari, M. Klados, P. Bamidis,”Graph Analysis on Functional Connectivity Networks during an Emotional Paradigm,” XII Mediterranean Conference on Medical and Biological Engineering and Computing 2010, vol. 29, no. January, 2010.

[53] M. Rubinov and O. Sporns,”Complex network measures of brain connectivity: Uses and interpretations,”NeuroImage, vol. 52, no. 3, pp. 1059-1069, 2010.

[54] A. Babarasi and R. Albert,”Emergence of Scaling in Random Networks,”Science, vol. 286, pp. 509-512, 1999.

[55] J. Onnela, J. Saramaki, J. Kertesz, and K. Kaski,”Intensity and coherence of motifs in weighted complex networks,”Physical Review E – Statistical, Nonlinear, and Soft Matter Physics, vo. 71, no. 6, pp. 1-4, 2005.

[56] M. Newman,”The Structure and Function of Complex Networks,”SIAM Review, vol. 45, no. 2, pp. 167-256, 2003.

[57] M. Newman,”Analysis of weighted networks,”Physical Review E, vol. 70, no. 5, 2004.

[58] D. Watts and S. Strogatz,”Collective dynamics of ‘small-world’ networks,”Nature, vol. 393, no. 6684, pp. 440-442, 1998.

[59] V. Latora and M. Marchiori,”Efficient Behavior of Small-World Networks,”Physical Review Letters, vol. 87, no. 19, 2001.

[60] R. Guimera and L. Amaral,”Cartography of complex networks: modules and universal roles,”Journal of Statistical Mechanics: Theory and Experiment, P02001, 2005.

[61] L. Freeman,”Centrality in social networks conceptual clarification,”Social Networks, vol. 1, no. 3, pp. 215-239, 1978.

[62] C. Leung and H. Chau,”Weighted assortative and disassortative networks model,”Physica A: Statistical Mechanics and its Applications, vol. 378, no. 2, pp. 591-602, 2007.

[63] R. Durgabai,”Feature Selection using ReliefF Algorithm,”International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, no. 10, pp. 8215-8218, 2014.

[64] Z. Wang, Y. Zhang, Z. Chen, H. Yang, Y. Sun, J. Kang, and Y. Yang,”Application of ReliefF Algorithm To Selecting Feature Sets for Classification of High Resolution Remote Sensing Image, pp. 755-758, 2016.

[65] S. Nancy, and S. Balamururgan,”A comparative study of feature selection methods for cancer classification using gene expression dataset,”Journal of Computer Applications (JCA), vol. 6, no. 3, 2013.

[66] Mathworks," Machine Learning Challenges : Choosing the Best Model and Avoiding Overfitting, pp. 1-8, 2016.

[67] M. Awad, and R. Khanna,"Support Vector Machines for Classification,"Efficient Learning Machines, vol. 71, no. 6, pp. 39-66, 2015.

[68] L. Auria and R. Moro,"Support Vector Machines as a Technique for Solvency Analysis, 1, 2008.

[69] W. Wei and Q. Jia,"Weighted Feature Gaussian Kernel SVM for Emotion Recognition,"Computational Intelligence and Neuroscience, pp. 1-7, 2016.

[70] M. Fischetti" Fast training of Support Vector Machines with Gaussian kernel,"Discrete Optimization, vol. 22, pp. 183-194, 2016.

[71] B. Patel, and S.Prajapati,"Efficient Classification of Data Using Decision Tree," *Bonfring International Journal of Data Mining*, vol. 2, no. 1, pp. 6-12,2012.

[72] F. Alam, F. Bappee, R. Rabbani, and M. Islam," An Optimized Formulation of Decision Tree,"ICAC3 2013: Advances in Computing, Communication, and Control, vol. 361, pp. 105-118, 2013.

[73] M. Magnani, and D. Montesi," Uncertainty in decision tree classifiers," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 6379 LNAI, pp. 250-263, 2010.

[74] R. Barros, M. Basgalupp, A. C.P.L.F. de Carvalho, and A. Freitas," Automatic Design of Decision-Tree Algorithms with Evolutionary Algorithms," Evolutionary computation, vol. 21, no. 4, pp. 659-684, 2013.

[75] PN Tan, M. Steinbach, and V. Kumar,"Introduction to Data Mining," Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 2005.

[76] E. Hunt, J. Marin, and P. Stone,"Experiments in Induction,"New York : Academic Press, 1966.

[77] J. Quinlan," C4.5: programs for machine learning, Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 1993.

[78] C. Li, S. Zhang, H. Zhang, L. Pang, K. Lam, C. Hui, and S. Zhang," Using the K-Nearest Neighbor Algorithm for the Classification of Lymph Node Metastasis in Gastric Cancer," Computational and Mathematical Methods in Medicine, vol. 2012, pp. 1-11, 2012.

[79] A. Kataria, and M. Singh," A Review of Data Classification Using K-Nearest Neighbour Algorithm," International Journal of Emerging Technology and Advanced Engineering, vol. 3, no. 6, pp. 354-360, 2013.

[80] P. Thanh Noi, and M. Kappas," Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery,"Sensors, vol. 18, no. 2, 2017.

[81] Z. Zhou,"Ensemble methods : Foundations and Algorithms,"Chapman & Hall/CRC Machine Learning & Pattern Recognition, CRC Press book, 2012.

[82] L. Breiman,"Bagging predictors,"Machine Learning, vol. 24, no. 2, pp. 123-140, 1996.

[83] L. Breiman,"Random Forests,"Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[84] M. Pal," Random forest classifier for remote sensing classification," International Journal of Remote Sensing, vol. 26, no. 1, pp. 217-222, 2005.

[85] Y. Wang, W. Chen, K. Huang, and Q. Gu," Classification of neonatal amplitude-integrated EEG using random forest model with combined feature," Proceedings - 2013 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2013, pp. 285-290, 2013.

[86] T. Ho," Nearest neighbors in random subspaces,"Advances in Pattern Recognition, vol. 1451, pp. 640-648, 1998.

[87] P. Kayal, and S. Kannan," An Ensemble Classifier Adopting Random Subspace Method based on Fuzzy Partial Mining,"Indian Journal of Science and Technology, vol. 10, no. 12, pp. 1-8, 2017.

[88] P. Mewada, and J. Patil,"Performance Analysis of *k*-NN on High Dimensional Datasets,"International Journal of Computer Applications, vol. 16, no. 2, 2011.

[89] R. Kohavi," A study of cross validation and bootstrap for accuracy estimation and model selection," 14th International Joint Conference on Artificial Intelligence (IJCAI), vol. 2, pp. 1137-1143, 1995.

[90] T. Fushiki," Estimation of prediction error by using K -fold cross-validation,"Statistics and Computing, vol. 21, no. 2, pp. 137-146, 2011.

[91] J. Rodriguez, A. Perez, and J. Lozano," Sensitivity analysis of k-fold cross validation in prediction error estimation," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 32, no. 3, pp. 569-575, 2010.

[92] O. Sporns,"Brain Connectivity,"Scholarpedia, vol. 2, no. 10, 2007.

[93] R. E. Beaty, S. Kaufman, M. Benedek, R. E. Jung, Y. N. Kenett, E. Jauk, A. C. Neubauer, and P. J. Silvia," Personality and Complex Brain Networks : The Role of Openness to Experience in Default Network Efficiency,"Human Brain Mapping, vol. 37, pp. 773-779, 2016.

[94] Q. Gao, Q. Xu, X. Duan, W. Liao, J. Ding, Z. Zhang, Y. Li, G. Lu, and C. Huafu," Extraversion and neuroticism relate to topological properties of resting-state brain networks,"Frontiers in Human Neuroscience, vol. 7, no. June, 2013.

[95] J. Zhang, J. Wang, Q. Wu, W. Kuang, X. Huang, Y. He, and Q. Kong,"Disrupted brain connectivity networks in drug-naive, first-episode major depressive disorder,"*Biol. Psychiatry,* vol. 70, pp. 334–342, 2011.

[96] J. Xu, X. Yin, H. Ge, Y. Han, Z. Pang, Y. Tang, B. Liu, and S. Liu," Attentional performance is correlated with the local regional efficiency of intrinsic brain networks,"Frontiers in Behavioral Neuroscience, vol. 9, no. July, pp. 1-11, 2015.

[97] P. Taylor, Y. Wang, and M. Kaiser," Within brain area tractography suggests local modularity using high resolution connectomics," Scientific Reports, vol. 7, no. January, pp. 1-9, 2017.

[98] L. Nummenmaa, and A. J. Calder," Neural mechanisms of social attention,"Trends in Cognitive Sciences, vol. 13, pp. 135–143, 2009.

[99] T. Canli,Z. Zhao,J. E. Desmond,E. Kang,J. Gross, and J. D. Gabrieli,"An fMRI study of personality influences on brain reactivity to emotional stimuli,"Behavioral Neuroscience, vol. 115, pp. 33–42, 2001.

[100] B. W. Haas,R. T. Constable,and T. Canli,"Stop the sadness: Neuroticismis associated with sustained medial prefrontal cortex response to emotional facial expressions,"NeuroImage, vol. 42, pp. 385–392, 2008.

[101] B. W. Haas, K. Omura, R. T. Constable, and T. Canli," Emotional conflict and neuroticism: Personality-dependent activation in the amygdala and subgenual anterior cingulated,"Behavioral Neuroscience, vol. 121, pp. 249–256, 2007.

[102] K. Jimura, S. Konishi,and Y. Miyashita,"Temporal pole activity during perception of sad faces, but not happy faces, correlates with neuroticism trait,"Neuroscience Letters, vol. 453, pp. 45–48, 2009.

[103] W. A. Cunningham, N. L. Arbuckle, A. Jahn, S. M. Mowrer,and A. M. Abduljalil,"Aspects of neuroticism and the amygdala: Chronic tuning from motivational styles,"Neuropsychologia, vol. 48, pp. 3399–3404, 2010.

[104] A. B. Bruhl, M. C. Viebke,T. Baumgartner,T. Kaffenberger, and U. Herwig,"Neural correlates of personality dimensions and affective measures during the anticipation of emotional stimuli,"Brain Imaging and Behavior, vol. 5, pp. 86–96, 2011.

[105] J. M. Spielberg,J. L. Stewart,R. L. Levin,G. A. Miller,and W. Heller,"Prefrontal cortex, emotion, and approach/withdrawal notivation,"Social and Personality Psychology Compass, vol. 2, pp. 135–153, 2008.

[106] M. Zuckerman,”*Psychobiology of Personality”,*New York, NY: Cambridge University Press, 2005.

[107] J. Tanji,and E. Hoshi,”Role of the lateral prefrontal cortex in executive behavioral control,”Physiology Review, vol. 88, pp. 37–57, 2008.

[108] H. Cremers, M. van Tol, K. Roelofs, A. Aleman, F. G. Zitman, M. A. van Buchem, D. J. Veltman, and N. van der Wee,” Extraversion is linked to volume of the orbitofrontal cortex and Amygdala,” PLoS ONE, vol. 6, no. 12, pp. 1-6, 2011.

[109] Y. Tran, A. Craig, and P. McIsaac,” Extraversion–introversion and 8–13 Hz waves in frontal cortical regions,”Personality and Individual Differences, vol. 30, no. 2, pp. 205–215, 2001.

[110] Y. Tran, A. Craig, P. Boord, K. Connell, N. Cooper, and E. Gordon,” Personality traits and its association with resting regional brain activity,”International Journal of Psychophysiology, vol. 60, no. 3, pp. 215-224, 2006.

[111] T. Johannisson,”Correlations between personality traits and specific groups of alpha waves in the human EEG,”PeerJ, vol.4, pp. 22-45, 2016.

[112] D. Hagemann, J.Hewig, C. Walter, A. Schankin, D. Danner, and E. Naumann,”Positive evidence for Eysenck’s arousal hypothesis: A combined EEG and MRI study with multiple measurement occasions,” Personality and Individual Differences, vol. 47, no. 7, pp. 717-721, 2009.

[113] R. McCrae, and P. Costa,”Positive and Negative Valence within the Five-Factor Model,”Journal of Research in Personality, vol. 29, pp. 443-460, 1995.

[114] R. McCrae,” Openness to Experience: Expanding the boundaries of Factor V,” European Journal of Personality, vol. 8, no. 4 , pp. 251-272, 1994.

[115] L. Barrett,”Valence is a basic building block of emotional life,”Journal of Research in Personality, vol. 40, pp. 35-55, 2006.

[116] K. Kaspar, and P. Konig,”Emotions and personality traits as high-level factors in visual attention,”Frontiers in Human Neuroscience, vol. 6, no. 321, 2012.

[117] H. Wadlinger,and D. Isaacowitz,” Positive mood broadens visual attention to positive stimuli,”Motiv Emot, vol. 30, pp. 89-101, 2006.