ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Μοντέλα εποπτευόμενης μηχανικής μάθησης για αναγνώριση συναισθήματος μέσω διεπαφής εγκεφάλου υπολογιστή σε παιχνίδια σοβαρού σκοπού για την υγεία

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ιωσήφ-Γεώργιος Ι. Παναγιωτόπουλος

Επιβλέπων : Κωνσταντίνα Νικήτα

Καθηγήτρια

Αθήνα, Φεβρουάριος 2022

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ
ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

Μοντέλα εποπτευόμενης μηχανικής μάθησης για αναγνώριση συναισθήματος μέσω διεπαφής εγκεφάλου υπολογιστή σε παιχνίδια σοβαρού σκοπού για την υγεία

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ιωσήφ-Γεώργιος Ι. Παναγιωτόπουλος

Επιβλέπων : Κωνσταντίνα Νικήτα

Καθηγήτρια

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 28η Φεβρουαρίου 2022.

....	....	....
Κωνσταντίνα Νικήτα	Ανδρέας - Γεώργιος Σταφυλοπάτης	Γιώργος Στάμου
Καθηγήτρια	Καθηγητής	Καθηγητής

Αθήνα, Φεβρουάριος 2022

....

Ιωσήφ-Γεώργιος Ι. Παναγιωτόπουλος

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

ΠΕΡΙΛΗΨΗ

Στο πλαίσιο της παρούσας διπλωματικής εργασίας διερευνάται η δυνατότητα αξιοποίησης δεδομένων διεπαφής υπολογιστή-εγκεφάλου για την αναγνώριση συναισθήματος κατά την αλληλεπίδραση με παιχνίδια σοβαρού σκοπού για την υγεία. Για το σκοπό αυτό, πραγματοποιήθηκε εξαγωγή χαρακτηριστικών από ηλεκτροεγκεφαλογραφικές (ΗΕΓ) καταγραφές που προέρχονται από τη βάση δεδομένων DREAMER και συλλέχθηκαν κατά τη διάρκεια διέγερσης των υποκειμένων της βάσης από κινηματογραφικό υλικό. Για την ανάπτυξη μοντέλων ταξινόμησης που βασίζονται σε εποπτευόμενη μηχανική μάθηση, αξιοποιήθηκε το μοντέλο περιγραφής συναισθημάτων VAD, που αξιοποιεί τις κλίμακες σθένους (Valence), διέγερσης (Arousal) και κυριαρχίας (Dominance). Τα δέντρα αποφάσεων ενισχυμένης κλίμακας χαρακτηρίστηκαν από την καλύτερη απόδοση αναγνώρισης συναισθήματος σε συγκριτική μελέτη που πραγματοποιήθηκε ανάμεσα σε μοντέλα εποπτευόμενης μηχανικής μάθησης. Η διπλωματική εργασία ολοκληρώνεται με την παρουσίαση του εννοιολογικού πλαισίου ενός παιχνιδιού σοβαρού σκοπού που ενσωματώνει ως είσοδο το συναίσθημα του χρήστη, όπως προκύπτει από τον ταξινομητή που αναπτύχθηκε στο πλαίσιο της διπλωματικής εργασίας, και έχει ως στόχο την ρύθμιση συναισθημάτων.

Λέξεις Κλειδιά

Συναισθηματική Υπολογιστική, Ταξινόμηση Συναισθημάτων, ΗΕΓ, Διεπαφή υπολογιστή-εγκεφάλου, Παιχνίδια Σοβαρού Σκοπού, Μηχανική Μάθηση.

Abstract

In the present Diploma Thesis, the potential of utilizing brain computer interface data for emotion recognition within the context of serious games for health has been explored. For this purpose, electroencephalogram (EEG) features were extracted from the DREAMER database, which contains data collected during the stimulation of subjects from cinematographic material. Classification models based on supervised machine learning, were developed to classify emotions using Valence, Arousal and Dominance (according to VAD emotion description model). Gradient boosted decision trees were characterized by the best performance of emotion recognition in a comparative study of different supervised machine learning models. The Thesis concludes with the conceptual framework of a serious game that utilizes the exported classification results and aims at providing users with an emotion regulation tool.

Index Keys

Affective Computing, Emotion classification, EEG, Brain-Computer Interface, Serious Games, Machine Learning.

ΕΥΧΑΡΙΣΤΙΕΣ

Η διπλωματική αυτή εργασία σηματοδοτεί την ολοκλήρωση ενός σημαντικού επιτεύγματός στην ζωή μου, την ολοκλήρωση των προπτυχιακών μου σπουδών. Επομένως σε αυτό το σημείο θα ήθελα να ευχαριστήσω μερικούς ανθρώπους, οι οποίοι με βοήθησαν, ο καθένας με τον δικό του τρόπο, για να φτάσω μέχρι εδώ.

Αρχικά θα ήθελα να ευχαριστήσω την επιβλέπουσα καθηγήτρια μου κα. Κωνσταντίνα Νικήτα, η οποία μου έδωσε την ευκαιρία να ασχοληθώ με το συναρπαστικό αυτό θέμα, καθώς και την βοήθεια της κατά ολη την διάρκεια της διπλωματικής μου εργασίας. Επιπλέον θα ήθελα να ευχαριστήσω τον διδάκτορα Κωνσταντίνο Μήτση, ο οποίος ήταν εκεί για να με βοηθήσει να ξεκαθαρίσω το θέμα στο μυαλό μου και να μου λύσει οποιαδήποτε απορία ή δυσκολία εμφανιζόταν.

Ακόμα, θα ήθελα να ευχαριστήσω τους φίλους μου και συμφοιτητές μου, Αντώνη Βάλμα, Παναγιώτη Λουλάκη, Παναγιώτη Μηναΐδη, Παναγιώτη Μαραβίτσα, Νταβίντ Σβέζεντσεβ, Γιώργο Μπαμπίλη, Γεώργιο Γέννη και Αλέξανδρο Παγώνη για την βοήθεια τους και την συμπαράσταση τους σε δυσκολίες που εμφανίστηκαν κατά την διάρκεια των σπουδών μου. Επιπλέον θα ήθελα να τους ευχαριστήσω και για όλες τις καλές στιγμές και όλα τα γέλια που είχαμε μαζί.

Πάνω από όλα, θα ήθελα να ευχαριστήσω την οικογένεια μου. Αρχικά, θα ήθελα να ευχαριστήσω τα αδέρφια μου, Νικήτα-Απόλλων Παναγιωτόπουλο και Ιουλία-Ελένη Παναγιωτοπούλου για τις συμβουλές τους κατά την διάρκεια των σπουδών τους και τις επιστημονικές τους κατευθύνσεις κατά την διάρκεια εκπόνησης αυτής της εργασίας. Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου Ιωάννη Παναγιωτόπουλο και Ελένη Στεφάνου, οι οποίοι ήταν κοντά μου σε όλες μου τις δυσκολίες.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

ΕΚΤΕΤΑΜΕΝΗ ΠΕΡΙΛΗΨΗ

Η συναισθηματική υπολογιστική γίνεται ένας όλο και πιο σημαντικός κλάδος, ενσωματώνοντας τα ανθρώπινα συναισθήματα στο σύνολο της διαθέσιμη επεξεργάσιμης πληροφορίας των υπολογιστικών συστημάτων. Η ανίχνευση και η ταξινόμηση των συναισθηματικών πληροφοριών είναι υποχρεωτικά στοιχεία για την ερμηνεία των συναισθηματικών καταστάσεων των ανθρώπων από τα σύγχρονα συστήματα υπολογιστών. Μια μηχανή που ερμηνεύει τη συναισθηματική κατάσταση των ανθρώπων μπορεί επιπλέον να χρησιμοποιήσει αυτές τις ερμηνείες προκειμένου να προσαρμόσει τη συμπεριφορά της, κάτι που θα επέτρεπε ή θα διευκόλυνε την ανάπτυξη μηχανών με συναισθηματική νοημοσύνη. Τα συστήματα υπολογιστών με την ικανότητα προσομοίωσης της ενσυναίσθησης μπορούν να οδηγήσουν στη δημιουργία εφαρμογών τεχνητής νοημοσύνης που λειτουργούν ως κοινωνικές επαφές ή ως εικονικοί θεραπευτές και θα είχαν τη δυνατότητα να ενισχύσουν τη θεραπεία ψυχολογικών διαταραχών.

Η αναγνώριση συναισθημάτων στον τομέα των Σοβαρών Παιχνιδιών παρέχει τα απαραίτητα εργαλεία για την ενσωμάτωση δυναμικών στοιχείων στο παιχνίδι τα οποία εξελίσσονται δυναμικά γύρω από τα συναισθήματα και τις συναισθηματικές καταστάσεις του χρήστη. Επιπλέον, τα συναισθήματα μπορούν να χρησιμοποιηθούν ως άμεσος τρόπος αναπληροφόρησης σχετικά με το περιεχόμενο του Σοβαρού Παιχνιδιού από τον χρήστη, κάτι που μπορεί να χρησιμοποιηθεί περαιτέρω για την αξιολόγηση του σχεδιασμού του ίδιου του παιχνιδιού και τον προσδιορισμό του κατά πόσο επιτυγχάνει να δημιουργήσει τα στοχευόμενα συναισθήματα στον χρήστη του. Τέλος, τα συναισθήματα μπορούν να χρησιμοποιηθούν ως εργαλείο για τον ενεργό έλεγχο στοιχείων του παιχνιδιού, επιτρέποντας τη χρήση ενός Σοβαρού Παιχνιδιού για την εξοικείωση του χρήστη με συγκεκριμένες νοοτροπίες.

Στο πλαίσιο αυτής της έρευνας, ένας Ταξινομητής Συναισθημάτων υλοποιείται και επικυρώνεται χρησιμοποιώντας ένα υπάρχον σύνολο δεδομένων. Στη συνέχεια συζητείται ο σχεδιασμός μιας ιδέας παιχνιδιού, η οποία χρησιμοποίει την έξοδο του Ταξινομητή Συναισθημάτων.

Η παρούσα διπλωματική εργασία ξεκινά με μια Ενότητα «Background» όπου παρουσιάζονται και εξηγούνται οι θεμελιώδεις όροι και οι τεχνολογίες που χρησιμοποιούνται σε αυτήν την έρευνα.

Συνεχίζει με την Ενότητα «Methods & Materials» όπου επεξηγούνται οι διαδικασίες που ακολουθήθηκαν σε αυτήν την έρευνα. Παρουσιάζει τα τρία βασικά στοιχεία που θα απαιτούνταν για την εφαρμογή της Αναγνώρισης Συναισθημάτων σε ένα Σοβαρό Παιχνίδι. Περιλαμβάνει μεθόδους απόκτησης δεδομένων από τον χρήστη, μεθόδους εξαγωγής χαρακτηριστικών που βασίζονται σε ΗΕΓ και ταξινόμησή τους σε συναισθήματα και, τέλος, μεθόδους για την εφαρμογή αυτών των συναισθημάτων σε ένα σοβαρό παιχνίδι. Ο κύριος στόχος αυτής της έρευνας ήταν να δημιουργήσει και να επικυρώσει την απόδοση ενός Ταξινομητή Συναισθημάτων που χρησιμοποιεί Χαρακτηριστικά Βασισμένα στο ΗΕΓ.

Στην Ενότητα «Results» παρουσιάζονται οι μετρήσεις που αφορούν την απόδοση του ανεπτυγμένου Ταξινομητή. Αυτή η ενότητα ακολουθείται από μια ενότητα «Discussion», όπου αυτά τα αποτελέσματα συζητούνται. Στο πλαίσιο αυτής της ενότητας συζητούνται επίσης οι μελλοντικές κατευθύνσεις αυτής της έρευνας.

ΜΕΘΟΔΟΙ ΚΑΙ ΥΛΙΚΑ

Η λειτουργία ενός ταξινομητή συναισθημάτων δεδομένου σημάτων ΗΕΓ μπορεί να χωριστεί σε τρία κύρια τμήματα. Το πρώτο τμήμα ασχολείται με την απόκτηση σημάτων χρησιμοποιώντας μια συσκευή διεπαφής υπολογιστή εγκεφάλου για την καταγραφή της εγκεφαλικής δραστηριότητας από το κεφάλι ενός χρήστη. Το δεύτερο τμήμα ασχολείται με την απομόνωση των δεδομένων ΗΕΓ από τα καταγεγραμμένα σήματα και την ταξινόμηση τους σε ένα τρισδιάστατο διάνυσμα με βάση το μοντέλο VAD. Το τελευταίο τμήμα ανακτά τα διανύσματα από το τμήμα δύο, χρησιμοποιώντας τα ως παραμέτρους στο πλαίσιο ενός παιχνιδιού σοβαρού σκοπού. Οι διαθέσιμες επιλογές σχεδίασης για αυτές τις παραμέτρους αναλύονται σε σχέση με τις βελτιώσεις και τις αλλαγές που θα παρείχαν στις "σοβαρές" και "gameplay" πτυχές του παιχνιδιού. Ένα διάγραμμα ροής του πρωτοκόλλου του πειράματος φαίνεται στο σχήμα 7.

ΕΠΙΛΟΓΗ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ

Εξετάστηκαν επτά βάσεις δεδομένων, πριν επιλεγεί το DREAMER για το πλαίσιο αυτού του πειράματος. Τα σύνολα δεδομένων αξιολογήθηκαν με βάση το μέγεθός τους, την προσβασιμότητα τους, τον αριθμό των υποκειμένων δοκιμής και τη μέθοδο που χρησιμοποιήθηκε για τη συλλογή δεδομένων.

“EEG Brainwave Dataset: Feeling Emotions”[17] περιέχει προεπεξεργασμένα δεδομένα ΗΕΓ, που αντιστοιχούν σε Θετικές ή Αρνητικές εμπειρίες. Η βάση δεδομένων απορρίφθηκε λόγω της εξαίρεσης ακατέργαστων μη επεξεργασμένων δεδομένων, καθώς και λόγω του μικρού αριθμού υποκειμένων που χρησιμοποιήθηκαν για τη δημιουργία αυτού του συνόλου δεδομένων (1 Αρσενικό Υποκείμενο).

“EEG Brainwave Dataset: Mental State”[18] αυτό το σύνολο δεδομένων ταξινομεί τα δεδομένα EEG σε τρεις ετικέτες, "Χαλαρό", "Ουδέτερο" και "Συγκέντρωση". Το σύνολο δεδομένων εξαιρέθηκε λόγω του μικρού αριθμού υποκειμένων που χρησιμοποιήθηκαν για τη δημιουργία αυτού του συνόλου δεδομένων (1 Άνδρας, 1 Γυναίκα).

“**K-EmoCon”**[19] είναι ένα σύνολο δεδομένων, το οποίο περιλαμβάνει τη χρονοσειρές ΗΕΓ οι οποίες αντιστοιχίζονται με συναισθήματα. Το σύνολο δεδομένων δημιουργήθηκε από ένα δείγμα 32 Υποκειμένων, που τους ζητήθηκε να συζητήσουν σε τυχαία ζεύγη.

“Seed”[20][21] είναι ένα προεπεξεργασμένο σύνολο δεδομένων που δημιουργήθηκε από ένα δείγμα 15 υποκειμένων (7 Άνδρες, 8 Γυναίκες). Τα δεδομένα συγκεντρώθηκαν καθώς τα υποκείμενα βάλθηκαν να παρακολουθήσουν 15 διαφορετικές ταινίες μικρού μήκους. Λόγω της φύσης των ταινιών και των παιχνιδιών σοβαρού σκοπού που βασίζονται αντίστοιχα στην αλληλεπίδραση του χρήστη με τα οπτικοακουστικά ερεθίσματα, αυτό το σύνολο δεδομένων ήταν μεταξύ των προτιμώμενων, ωστόσο δεν συμπεριλήφθηκε λόγω της χαμηλής προσβασιμότητας αυτού του συνόλου δεδομένων και της έλλειψης ακατέργαστων μη επεξεργασμένων δεδομένα. Αυτό το σύνολο δεδομένων, συνδυάζει χαρακτηριστικά ΗΕΓ με τρεις ετικέτες: Happy, Neutral, Sad.

“Seed-IV”[22] μια επέκταση του SEED αλλά χρησιμοποιείται για την ταξινόμηση μεταξύ: Happy, Neutral, Sad και Fear. Χρησιμοποιεί επιπλέον χαρακτηριστικά, που δημιουργούνται μέσω συσκευών παρακολούθησης ματιών. Αυτό το σύνολο δεδομένων εξαιρέθηκε για τον ίδιο λόγο με το SEED.

“DREAMER”[23] είναι μια βάση δεδομένων που αποτελείται από σήματα ηλεκτροεγκεφαλογραφήματος (ΗΕΓ) και ηλεκτροκαρδιογραφήματος (ΗΚΓ) που καταγράφονται κατά την συναισθηματική διέγερση μέσω οπτικοακουστικών ερεθισμάτων. Καταγράφηκαν σήματα από 23 συμμετέχοντες μαζί με την αυτοαξιολόγηση των συμμετεχόντων της συναισθηματικής τους κατάστασης μετά από κάθε ερέθισμα, ως προς το σθένος (Valence), τη διέγερση(Arousal) και την κυριαρχία (Dominance). Όλα τα σήματα καταγράφηκαν χρησιμοποιώντας φορητό, φορετό, ασύρματο, χαμηλού κόστους εξοπλισμό διαθέσιμο στην αγορά, που έχει τη δυνατότητα να επιτρέπει τη χρήση σε καθημερινές εφαρμογές. Αυτό το σύνολο δεδομένων επιλέχθηκε για τα πειράματα σε αυτό το έργο.

“DEAP”[24] είναι ένα σύνολο δεδομένων για την ανάλυση των ανθρώπινων συναισθηματικών καταστάσεων. Περιλαμβάνει το ηλεκτροεγκεφαλογράφημα (ΗΕΓ) και τα περιφερειακά φυσιολογικά σήματα 32 συμμετεχόντων, που καταγράφηκαν καθώς ο καθένας παρακολουθούσε 40 αποσπάσματα μουσικών βίντεο διάρκειας ενός λεπτού. Οι Συμμετέχοντες βαθμολόγησαν κάθε βίντεο ως προς τα επίπεδα διέγερσης (Arousal), σθένους(Valence), μου αρέσει/δεν μου αρέσει, κυριαρχία (Dominance) και εξοικείωση. Αυτό το σύνολο δεδομένων είναι υποψήφιο για μελλοντικό πειραματισμό. Το DREAMER προτιμήθηκε έναντι αυτού του συνόλου δεδομένων, λόγω της ταχύτερης απόκρισης του κατόχου του συνόλου δεδομένων, σχετικά με τη χρήση του συνόλου δεδομένων για ακαδημαϊκή έρευνα.

Περιγραφη της βασης DREAMER

Για την κατασκευή του συνόλου δεδομένων DREAMER, χρησιμοποιήθηκαν ακουστικά και οπτικά ερεθίσματα με τη μορφή ταινιών για να προκαλέσουν συναισθηματικές αντιδράσεις στους συμμετέχοντες της μελέτη. Ένα σύνολο δεδομένων που αποτελείται από 18 αποσπάσματα ταινιών, που επιλέχθηκαν και αξιολογήθηκαν από τους Gabert Quillen et al.[25] χρησιμοποιήθηκε για την πρόκληση συναισθημάτων. Αυτά τα αποσπάσματα ταινιών περιέχουν αποκομμένες σκηνές από διαφορετικές ταινίες που έχουν αποδειχθεί ότι προκαλούν ένα ευρύ φάσμα συναισθημάτων. Από αυτά τα 18 αποσπάσματα ταινιών, δύο από το καθένα στόχευαν ένα από τα ακόλουθα εννέα συναισθήματα: διασκέδαση, ενθουσιασμό, ευτυχία, ηρεμία, θυμό, αηδία, φόβο, λύπη και έκπληξη. Η διάρκεια των κλιπ ταινιών ήταν μεταξύ 65 και 393 δευτερολέπτων (M = 199s), η οποία θεωρείται επαρκής αφού, σύμφωνα με τους ψυχολόγους, τα ερεθίσματα βίντεο από 1 έως 10 λεπτά είναι ικανά να προκαλούν μεμονωμένα συναισθήματα[26],[27]. Ωστόσο, η συναισθηματική κατάσταση ενός ατόμου μπορεί να αλλάξει με την πάροδο του χρόνου, ειδικά όταν χρησιμοποιούνται ερεθίσματα βίντεο μεγαλύτερης διάρκειας. Για να αποφευχθεί η μόλυνση των δεδομένων με πολλαπλά συναισθήματα, μόνο οι εγγραφές που καταγράφηκαν κατά τα τελευταία 60 δευτερόλεπτα κάθε κλιπ ταινίας χρησιμοποιήθηκαν για περαιτέρω ανάλυση.

Για τη λήψη σημάτων ΗΕΓ, τοποθετήθηκαν ηλεκτρόδια στις ακόλουθες δεκαέξι θέσεις σύμφωνα με το Διεθνές σύστημα 10-20: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, M1 και M2. Ο αισθητήρας μαστοειδούς στο M1 λειτούργησε ως σημείο αναφοράς γείωσης για τη σύγκριση της τάσης όλων των άλλων αισθητήρων, ενώ ο αισθητήρας μαστοειδούς στο M2 ήταν μια αναφορά τροφοδοσίας για τη μείωση των εξωτερικών ηλεκτρικών παρεμβολών, και επομένως τα δεδομένα για τα M1 και M2 δεν αποθηκεύονται στο βάση δεδομένων. Ο ρυθμός δειγματοληψίας ορίστηκε στα 128 Hz.

Τα δεδομένα ΗΚΓ που περιέχονται στο σύνολο δεδομένων δεν χρησιμοποιήθηκαν στο πλαίσιο αυτού του πειράματος.

Στα υποκείμενα της δοκιμής δόθηκαν τρεις κλίμακες αξιολόγησης, για Διέγερση (Arousal), Σθένος (Valence) και Κυριαρχία (Dominance) και τους δόθηκε η οδηγία να βαθμολογήσουν κάθε κλιπ ταινίας σε κάθε μία από αυτές τις κλίμακες, με τιμή από 1 έως 5.

Η παραπάνω διαδικασία έχει οδηγήσει στη δημιουργία ενός συνόλου δεδομένων που έχει μορφοποιηθεί όπως εξηγείται παρακάτω. Για καθέναν από τους 23 έγκυρους συμμετέχοντες, υπάρχουν 18 διαφορετικές δομές δεδομένων, μία για κάθε ταινία. Αυτές οι δομές δεδομένων περιλαμβάνουν τις αξιολογήσεις του συμμετέχοντα και για τις τρεις κλίμακες αξιολόγησης, καθώς και 14 συστοιχίες που περιέχουν τα δείγματα ΗΕΓ για τη διάρκεια κάθε φιλμ, μία συστοιχία για κάθε διαφορετικό κανάλι ηλεκτροδίων. Τα δεδομένα ΗΕΓ για κάθε συμμετέχοντα στο σύνολο δεδομένων διαχωρίζονται σε βασικά χαρακτηριστικά και δοκιμαστικά χαρακτηριστικά. Τα χαρακτηριστικά της γραμμής βάσης υπολογίζονται από τις εγγραφές ΗΕΓ των τελευταίων 4 δευτερολέπτων ενός ουδέτερου κλιπ φιλμ που εμφανίζεται πριν από κάθε κλιπ ταινίας που προκαλεί επίδραση. Συμπεριλήφθηκε επίσης μια πρόσθετη δομή δεδομένων που περιείχε τα δεδομένα σημάτων ΗΚΓ καθώς και πληροφορίες σχετικά με το φύλο και την ηλικία κάθε συμμετέχοντα.

ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ

Τα δεδομένα του DREAMER υποβάλλονται σε προεπεξεργασία και χρησιμοποιούνται για τη δημιουργία μιας σειράς χαρακτηριστικών, τα οποία στη συνέχεια χρησιμοποιούνται ως δεδομένα από τους ταξινομητές συναισθημάτων που κατασκευάστηκαν για αυτό το πείραμα. Στη συνέχεια, τα αποτελέσματα της ταξινόμησης συγκρίνονται με τα αποτελέσματα παρόμοιων ερευνών.

ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ

Προτού τροφοδοτηθούν τα δεδομένα στον ταξινομητή, είναι απαραίτητο να προεπεξεργαστούν τα ακατέργαστα δεδομένα ΗΕΓ προκειμένου να απαλλαχτούν από τον θόρυβο και τα τεχνουργήματα που ανιχνεύονται αλλά δεν προέρχονται από τον εγκέφαλο. Τα σήματα που προκαλούνται από την καρδιακή δραστηριότητα, την κίνηση των ματιών και τη μυϊκή δραστηριότητα, καθώς και ο θόρυβος της γραμμής τροφοδοσίας συλλαμβάνονται από τη συσκευή ΗΕΓ και έτσι υποβαθμίζουν την ποιότητα των καταγεγραμμένων δεδομένων, καθιστώντας τη χρήση μεθόδων απενεργοποίησης θορύβου απαραίτητη.

Στα σήματα ΗΕΓ, τα περισσότερα οφθαλμικά τεχνουργήματα (ανοιγόκλειμα των ματιών, κίνηση των ματιών, καρδιακές παρεμβολές, κ.λπ.) κυριαρχούν κάτω από 4 Hz, οι κινήσεις των μυών παράγουν τεχνουργήματα πάνω από 30 Hz [28] και ο θόρυβος της γραμμής ρεύματος συνήθως βρίσκεται στα 50 ή 60 Hz, ενώ οι ζώνες συχνοτήτων που περιέχουν πληροφορίες σχετικά με την αναγνώρισης συναισθημάτων βρίσκονται στην περιοχή από 4 – 30 Hz. Αυτό το εύρος συχνοτήτων χωρίζεται συνήθως σε ζώνες θήτα, άλφα και βήτα. Εφαρμόζονται τρία ξεχωριστά ζωνοπεράτα φίλτρα Butterworth χρησιμοποιώντας συναρτήσεις python από τη βιβλιοθήκη scipy.signal προκειμένου να εξαχθούν μόνο οι συχνότητες εντός των περιοχών ενδιαφέροντος. Τα τεχνουργήματα που μπορεί να έχουν εισαχθεί στην αρχή ή στο τέλος των δεδομένων ΗΕΓ με τη μορφή μετατοπίσεων DC, αντιμετωπίζονται με συμπλήρωση των δεδομένων με μια σταθερά DC στην αρχή και στο τέλος, πριν από την εκ νέου δειγματοληψία. Τέλος, τα φιλτραρισμένα δεδομένα ΗΕΓ μετατοπίζονται κατά την καθυστέρηση ομάδας του φίλτρου. Το τελευταίο βήμα για την προετοιμασία των δεδομένων ΗΕΓ για περαιτέρω ανάλυση είναι η εφαρμογή της μεθόδου Common Average Reference (CAR), όπως προτείνεται από τον Cohen[29], η οποία υπολογίζει τη μέση τιμή σε όλα τα ηλεκτρόδια και την αφαιρεί από κάθε δείγμα κάθε ηλεκτροδίου. Με αυτόν τον τρόπο έχουμε αναλύσει τα δεδομένα ΗΕΓ σε έξι διαφορετικές συστοιχίες δεδομένων, τρεις από τις οποίες περιέχουν δεδομένα για τις ζώνες Άλφα, Βήτα και Θήτα για τα χαρακτηριστικά της βάσης σύγκρισης Οι άλλες τρείς περιέχουν δεδομένα για τις ζώνες Alpha, Beta και Theta χαρακτηριστικών δοκιμής.

ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

Είναι καλά τεκμηριωμένο ότι οι φασματικές πυκνότητες ισχύος (PSD) των σημάτων ΗΕΓ σε διαφορετικές ζώνες συσχετίζονται με τη συναισθηματική κατάσταση ενός ανθρώπου [31].

Soleymani et al.[26] έδειξαν ότι οι συνιστώσες υψηλότερης συχνότητας των σημάτων ΗΕΓ μεταφέρουν πιο σημαντικές πληροφορίες σχετικά με τα θετικά συναισθήματα σε σύγκριση με τα αρνητικά (υψηλού και χαμηλού σθένους αντίστοιχα), ενώ έχει επίσης αναφερθεί συσχέτιση μεταξύ ισχύος βήτα και θετικής συναισθηματικής αυτεπαγωγής [32] . Οι Koelstra et al. [30] βρήκαν επίσης ισχυρές συσχετίσεις μεταξύ σθένους και σημάτων ΗΕΓ σε όλες τις ζώνες συχνοτήτων. Επιπλέον, η μελέτη τους βρήκε επίσης αρνητικές συσχετίσεις μεταξύ της διέγερσης και των ζωνών θήτα, άλφα και γάμμα των σημάτων ΗΕΓ, ενώ προηγούμενες μελέτες [33] [34] ανέφεραν μια αντίστροφη σχέση μεταξύ της ισχύος της ζώνης άλφα και του γενικού επιπέδου διέγερσης.

Μετά το στάδιο της προεπεξεργασίας, τα ληφθέντα σήματα ΗΕΓ διαχωρίστηκαν στις ζώνες συχνοτήτων θήτα (4 Hz - 8 Hz), άλφα (8 Hz - 13 Hz) και βήτα (13 Hz - 20 Hz). Στη συνέχεια, ο εκτιμητής μέσου όρου επικαλυπτόμενων τμημάτων του Welch χρησιμοποιείται για την εκτίμηση του PSD κάθε ζώνης EEG, χρησιμοποιώντας ένα παράθυρο 256 δειγμάτων με επικάλυψη 128 δειγμάτων. Οι λογάριθμοι του PSD από καθεμία από τις προαναφερθείσες ζώνες εξάγονται από το σήμα καθενός από τα 14 ηλεκτρόδια προκειμένου να χρησιμοποιηθούν ως χαρακτηριστικά, όπως προτείνεται επίσης στα [22], [26], [30] , οδηγώντας στη δημιουργία ενός συνόλου από 42 χαρακτηριστικών (3 από καθένα από τα 14 ηλεκτρόδια). Ο προαναφερθείς υπολογισμός ολοκληρώνεται τόσο στα χαρακτηριστικά βάσης σύγκρισης όσο και στα δοκιμαστικά χαρακτηριστικά.

Πριν της δημιουργίας των διανυσμάτων χαρακτηριστικών, είναι απαραίτητο να γίνουν τα δεδομένα συγκρίσιμα, καθώς το μέγεθος και το εύρος τους εξαρτώνται σε μεγάλο βαθμό από συγκεκριμένες συνθήκες, χαρακτηριστικά σήματος πηγής και διαφέρουν ανάμεσα στα υποκείμενα της βάσης δεδομένων. Το μέγεθος των χαρακτηριστικών που εξάγονται από το ΗΕΓ ποικίλλει σημαντικά ανάλογα με τον τύπο του χαρακτηριστικού και την πηγή του. Για παράδειγμα, κατά την εξαγωγή χαρακτηριστικών από σήματα ΗΕΓ, το PSD για υψηλότερες συχνότητες έχει πολύ μικρότερο μέγεθος από το PSD για χαμηλότερες συχνότητες [35]. Τα φυσιολογικά σήματα τείνουν να έχουν υψηλή διακύμανση μεταξύ διαφορετικών υποκειμένων, καθώς και μεταξύ των ίδιων χαρακτηριστικών που μετρώνται σε διαφορετικές στιγμές για κάθε άτομο [36]. Χρησιμοποιείται κανονικοποίηση χαρακτηριστικών για την αντιμετώπιση αυτών των ζητημάτων. Μετά τον υπολογισμό των τιμών PSD τόσο για τα δοκιμαστικές χαρακτηριστικά όσο και για τα χαρακτηριστικά της βάσης σύγκρισης, καθένα από τα δοκιμαστικά χαρακτηριστικά διαιρείται με τα αντίστοιχα χαρακτηριστικά της βάσης σύγκρισης. Λόγω της διαίρεσης των εξαγόμενων χαρακτηριστικών με τα χαρακτηριστικά βάσης συγκρίσης, τα προκύπτον κανονικοποιημένα χαρακτηριστικά περιέχουν πληροφορίες σχετικά με τη σχέση των αρχικών χαρακτηριστικών με τη δραστηριότητα υποβάθρου, δηλαδή δραστηριότητα που υπάρχει στα δεδομένα αλλά δεν διαμορφώνεται από τα πραγματικά συναισθηματικά ερεθίσματα [37]. Ως αποτέλεσμα, η μέθοδος κανονικοποίησης χαρακτηριστικών χρησιμοποιείται για την αφαίρεση ή την έντονη άμβλυνση της δραστηριότητας του υποβάθρου προκειμένου να ληφθούν μόνο αλλαγές που σχετίζονται με ερεθίσματα στις εγγραφές ΗΕΓ.

Τέλος, όλα τα χαρακτηριστικά ενώνονται στο τελικό διάνυσμα χαρακτηριστικών F_{EEG} ως εξής: Έστω $F_{i\theta}$, $F_{i\alpha}$ και $F_{i\beta}$, ο κανονικοποιημένος λογάριθμος του PSD για το σήμα του i-στου ηλεκτροδίου, i= 1,2,…,14 , για τις ζώνες θήτα, άλφα και βήτα αντίστοιχα. Το τελικό διάνυσμα χαρακτηριστικών ορίζεται ως $F_{EEG} = [F_{1\theta}\ F_{1\alpha}\ F_{1\beta} \dots F_{14\theta}\ F_{14\alpha}\ F_{14\beta}]$

Πειραματα Ταξινομησησ

Ορίστηκαν τρία διαφορετικά σχήματα δυαδικής ταξινόμησης: η ταξινόμηση μεταξύ χαμηλής/υψηλής διέγερσης (ήρεμη/διεγερμένη), χαμηλού/υψηλού σθένους (δυσάρεστη/ευχάριστη) και χαμηλή/υψηλή κυριαρχία (χωρίς έλεγχο/ενδυνάμωση). Με βάση αυτό το σχήμα ταξινόμησης, η κλίμακα βαθμολογίας 5 βαθμών που χρησιμοποιήθηκε από τους συμμετέχοντες χωρίστηκε σε δύο κατηγορίες (χαμηλή και υψηλή).

Τρεις μέθοδοι ταξινόμησης δοκιμάστηκαν στα δεδομένα, χρησιμοποιώντας δύο διαφορετικές προσεγγίσεις για την επικύρωση των αποτελεσμάτων. Οι μέθοδοι ταξινόμησης που συγκρίθηκαν περιελάμβαναν τον αλγόριθμο Gaussian Naïve Bayes, χρήση SVM και την χρήση δέντρων απόφασης ενισχυμένης κλίσης υλοποιημένα με XGBoost. Τα πειράματα ταξινόμησης πραγματοποιήθηκαν ανεξάρτητα για κάθε διαφορετικό συμμετέχοντα. Κατά την πρώτη προσέγγιση, τα δεδομένα χωρίστηκαν σε δεδομένα εκπαίδευσης και σε δεδομένα δοκιμής, χρησιμοποιώντας μια αναλογία 2:1. Το μοντέλο ταξινόμησης εκπαιδεύτηκε χρησιμοποιώντας το 67% των Δεδομένων πριν δοκιμαστεί στο υπόλοιπο 33% των δεδομένων. Για κάθε συμμετέχοντα μετρήθηκαν οι Ορθότητα, Ακρίβεια, Ανάκληση και βαθμολογία F1. Τα αποτελέσματα του ταξινομητή συγκρίνονται επίσης με των Katsigiannis et al[22], λαμβάνοντας υπόψη ότι όλα τα Πειράματα Ταξινόμησης εκτελούνται χρησιμοποιώντας το ίδιο σύνολο δεδομένων. Η μέθοδος ταξινόμησης που δείχνει τα καλύτερα αποτελέσματα επικυρώνεται στη συνέχεια χρησιμοποιώντας τη δεύτερη προσέγγιση, ακολουθώντας ένα πρωτόκολλο διασταυρούμενης επικύρωσης 10 τμημάτων. Αυτό επιτρέπει τη σύγκριση της απόδοσης του Ταξινομητή χρησιμοποιώντας διαφορετικά σετ εκπαίδευσης και έτσι παρέχει καλύτερη αναπαράσταση της απόδοσης του ταξινομητή εκτός των πλαισίων αυτού του πειράματος.

Για περαιτέρω βελτίωση της απόδοσης του ταξινομητή, επιχειρείται η εκτέλεση Grid Search πάνω από τις υπερ παραμέτρους του ταξινομητή για τα δεδομένα κάθε συμμετέχοντα. Στη συνέχεια, ο ταξινομητής επικυρώνεται ξανά χρησιμοποιώντας διασταυρούμενη επικύρωση 10 τμημάτων, συγκρίνοντας τα αποτελέσματα με τα προηγούμενα.

Αφού ο ταξινομητής έχει δοκιμαστεί για την απόδοσή του στην περίπτωση που ο εκπαιδεύεται ανεξάρτητα για κάθε διαφορετικό συμμετέχοντα, το πείραμα επιχειρεί επιπλέον να μετρήσει την απόδοση του ταξινομητή για την αντίθετη περίπτωση. Χρησιμοποιώντας μια προσέγγιση Leave-one-out, όλοι οι συμμετέχοντες εκτός από έναν, προστίθενται ως δεδομένα εκπαίδευσης στον ταξινομητή. Στη συνέχεια, ο ταξινομητής ελέγχεται στα δεδομένα εκείνου του ενός συμμετέχοντα που έμεινε εκτός των δεδομένων εκπαίδευσης. Αυτή η διαδικασία επαναλαμβάνεται, με τον συμμετέχοντα της βάσης δεδομένων, έως ότου όλοι οι συμμετέχοντες χρησιμοποιηθούν ως δεδομένα εκπαίδευσης και ως δεδομένα δοκιμών.

Σχεδιαση παιχνιδιου σοβαρου σκοπου

Για να αξιοποιηθεί η έξοδος του ταξινομητή και να μπορέσουν να πραγματοποιηθούν περαιτέρω πειράματα σε αυτόν, αυτό το έγγραφο προχωρά στην πρόταση ενός σχεδιασμού παιχνιδιού σοβαρού σκοπού, ο οποίος επιτρέπει την προσαρμογή της συναισθηματικής κατάστασης του χρήστη στη μηχανική του παιχνιδιού.

Κατά κύριο λόγο, είναι υποχρεωτικό να συμπεριληφθούν χαρακτηριστικά που λειτουργούν ως δείκτες της καταγεγραμμένης συναισθηματικής κατάστασης. Οι ενδείξεις μπορεί να είναι οπτικές, ακουστικές ή και τα δύο. Για να υπεδείχθη η καταγεγραμμένη τιμή του σθένους (Valence), μπορεί να εφαρμοστεί ένα οπτικό φίλτρο για τη ρύθμιση της ζεστασιάς των ορατών χρωμάτων, δημιουργώντας μια πιο κρύα μπλε εικόνα κατά την εισαγωγή χαμηλής βαθμολογίας σθένους ή μια πιο ζεστή κοκκινωπή απόχρωση όταν καταγράφεται μια υψηλή βαθμολογία σθένους. Οι τιμές κορεσμού της οθόνης μπορούν να αντιστοιχιστούν με την καταγεγραμμένη διέγερση(Arousal), με τη χαμηλή διέγερση να οδηγεί σε χαμηλότερο κορεσμό. Αντίστοιχα, η καταγραφή της κυριαρχίας (Dominance) μπορεί να εκχωρηθεί στη φωτεινότητα των γραφικών του παιχνιδιού, με τις σκηνές να γίνονται πιο σκοτεινές όταν καταγράφεται υψηλή βαθμολογία κυριαρχίας. Επιπλέον, μπορεί να προστεθεί μια επικάλυψη οπτικών εφέ προκειμένου να ενισχυθεί η οπτική επίδραση των μηχανισμών ένδειξης που προτάθηκαν προηγουμένως. Η μπλε απόχρωση που παράγεται όταν το καταγεγραμμένο σθένος είναι χαμηλό, μπορεί να συμπληρωθεί με το οπτικό αποτέλεσμα των γωνιών της οθόνης που παγώνουν, η πτώση του κορεσμού που παράγεται για να υποδηλώνει χαμηλή διέγερση μπορεί να συμπληρωθεί με μια επικάλυψη θορύβου RGB που προστίθεται στα άκρα οθόνη του παίκτη. Τα ακουστικά χαρακτηριστικά που μπορούν να συμπεριληφθούν για τη βελτίωση αυτών των ενδείξεων περιλαμβάνουν τη συμπερίληψη ηχητικών εφέ που ταιριάζουν με τυχόν εμφανιζόμενα οπτικά εφέ, καθώς και τροποποιήσεις του τόνου της μουσικής υπόκρουσης, μειώνοντάς τον για χαμηλότερες τιμές κυριαρχίας. Η εγγραφή του ίδιου κομματιού μουσικής υπόκρουσης με διαφορετικά όργανα και η μίξη των εντάσεων τους με βάση τις τιμές του διανύσματος VAD είναι επίσης μια έγκυρη επιλογή.

Επιπλέον, σχεδιάζονται λειτουργίες που ελέγχουν την κατάσταση του παιχνιδιού και χαρακτηριστικά που προωθούν έναν συγκεκριμένο στόχο για το παιχνίδι. Στο πλαίσιο αυτού του εγγράφου, ο σχεδιασμός του παιχνιδιού υποστηρίζει λειτουργίες που του επιτρέπουν να χρησιμοποιηθεί ως εφαρμογή ρύθμισης συναισθημάτων. Αυτές οι λειτουργίες προσαρμόζονται απευθείας στους μηχανισμούς και στο gameplay, επομένως θα συζητούνται παράλληλα με την εξήγηση των βασικών στοιχείων του παιχνιδιού.

Αποτελεσματα

Στο κεφάλαιο των αποτελεσμάτων παραθέτονται διαγράμματα και πινάκες με τις μετρήσεις που λαμβάνονται για την επικύρωση της λειτουργικής ποιότητας του ταξινομητή συναισθημάτων.

Αρχικά συγκρίνονται οι ορθότητες πρόβλεψης και τα F1-Score των τριών μεθόδων ταξινόμησης που χρησιμοποιούνται με την μελέτη αναφοράς. Για αυτές τις συγκρίσεις η επικύρωση γίνεται χωρίζοντας τα δεδομένα από κάθε υποκείμενο της βάσης σε δεδομένα εκπαίδευσης και σε δοκιμαστικά δεδομένα με αναλογία δύο προς ένα (2:1). Η μέθοδος δέντρων απόφασης ενισχυμένης κλίσης υλοποιημένα με XGBoost παρουσιάζει ορθότητα εκτιμημένη στο 85% και F1-Score στο 80%, υπερτερώντας των υπολοίπων μεθόδων. Στη μελέτη αναφοράς τα αντίστοιχα μεγέθη εκτιμούνται στο 62% και 52%, επομένως παρατηρούμε μια βελτίωση της κλίμακας +23% . Με τη χρήση Γκαουσιανού Ταξινομητή (GNB Classifier), η ορθότητα εκτιμάται στο 60% και το F1-Score στο 54%. Τέλος με τη χρήση SVM η ορθότητα εκτιμάται στο 68% και το F1-Score στο 58%.

Στη συνέχεια η μέθοδος δέντρων απόφασης ενισχυμένης κλίσης υλοποιημένα με XGBoost, επικυρώνεται με χρήση ενός πρωτόκολλου διασταυρούμενης επικύρωσης 10 τμημάτων. Η ταξινομήσεις του σθένους (Valence), διέγερσης(Arousal) και κυριαρχίας(Dominance), παρουσιάζονται σε ξεχωριστούς πινάκες. Η μετρικές της ορθότητας ανά τα υποκείμενα της βάσης δεδομένων για την ταξινόμηση του σθένους υπολογίζουν κατά μέσο όρο 88.12%. Αντίστοιχα καταγράφονται μετρικές για την ακρίβεια στο 87.14%, ανάκληση στο 81.65%, F1-Score στο 82.65% και ROC AUC στο 94.48%. Οι μετρικές της διέγερσης και της κυριαρχίας παρουσιάζουν παρόμοια αποτελέσματα.

Για έναν πιο εκτενή έλεγχο της λειτουργικής ποιότητας τους ταξινομητή, απεικονίζονται τα Box Plots για όλες τις μετρικές και τις κατηγόριες ταξινόμησης. Μέσα από αυτά μπορεί να εκτιμηθούν οι διακυμάνσεις της ποιότητας του ταξινομητή και η αξιοπιστία του.

Έπειτα παρουσιάζονται οι μετρήσεις που έγιναν κατά την διαδικασία εύρεσης υπερπαραμέτρων για τον ταξινομητή XGBoost, μέσα από τις οποίες επιχειρήθηκε η περαιτέρω βελτίωση της λειτουργία ςτου ταξινομήτη.

Τέλος, παρουσιάζονται σε μορφή πίνακα, οι μετρικές της ορθότητας του ταξινομητή, στην περίπτωση που τα δεδομένα πολλαπλών χρηστών αναμειχθούν για την εκπαίδευση του και στη δοκιμαστούν στα δεδομένα αγνώστου υποκείμενου, εμφανίζοντας μια μέση ορθότητα της κλίμακας του 56.25%.

INTRODUCTION

Affective computing is becoming increasingly important by extending the possibilities of computing technologies through recognition of emotions in real time during human computer interactions. Detecting and classifying emotional information are mandatory components towards the interpretation of the emotional states of humans by modern computer systems. A machine that interprets the emotional state of humans can furthermore make use of these interpretations to adapt its behavior, something which would enable or facilitate the development of machines with emotional intelligence. Computer systems with the capability to simulate empathy can lead to the creation of AI-Based applications that acting as a social contact or as a virtual therapist they could reinforce the treatment of psychological conditions.

Emotion recognition in the field of Serious Game provides the necessary tools to implement gameplay elements that would dynamically evolve around the emotions and the emotional states of the user. In additional to that, emotions can be used as a direct way to receive feedback concerning the content of the Serious Game from the user, something which can be further utilized to evaluate the design of the game itself and determine whether it achieves to deliver the targeted emotions to its user. Finally, emotions can be used as a tool to actively control elements of the game, allowing a Serious Game to be used in order to train the user into using specific mentalities.

Within the context of this research, an Emotions Classifier is implemented and validated using an existing dataset. Then the design of a game concept is discussed, which utilized the output of the Emotions Classifier.

This document begins with a "**Background" Section** where the fundamental terms and the technologies used within this research are presented and explained.

It continues with the **"Methods & Materials" Section** where the procedures followed for this research are explained. It presents the three core components which would be required to implement Emotion Recognition in a Serious Game. It includes methods to acquire data from the user, methods to extract EEG-Based features and classify them into emotions and finally methods to implement those emotions into a Serious Game. The main focus of this research was to create and validate the performance an Emotions Classifier which makes use of EEG-Based Features.

In the **"Results" Section** the measurements concerning the performance of the developed Classifier are presented. This section is followed by a **"Discussion" Section**, where these results are being discussed. The future directions of this research are also being discussed within the context of this section.

Background

Affective Computing

Affective Computing is an interdisciplinary field of study spanning computer science, psychology and cognitive science. [1] Affective Computing revolves around the research and development of computing systems capable of recognizing, interpreting, processing and simulating human affects.

Emotion recognition is one of the tasks, which affective computing is striving to resolve. Existing technologies built to tackle this task, either use information deriving from the alterations detected to a person’s speech, their facial expressions, their body gestures or other features extracted from the various physiological signs of the individual. Data from Electroencephalograms (EEG) can also be used for the extraction of features that prove to be useful for the task of emotion recognition.

VAD Model

The VAD Model (Valence, Arousal, Dominance Model), is a Spatial projection of human emotions to three axis. As such, each emotion is represented by a three-dimensional vector, for the three values of Valence, Arousal and Dominance correlated to the recorded emotion. Valence is the level of pleasantness detected in the emotion and is defined along a continuum from negative to positive. Arousal ranges from relaxation at low values, to excitement at high values. Dominance the level of control felt by the individual within a situation. High dominance values tend to be recorded in emotions, where the individual is feeling in stronger, such as anger, while low dominance values are detected in submissive emotions such as fear. The VAD Model can also be referred to as PAD Model.[2]

Brain-computer Interface

A brain-computer interface (BCI) is a computer-based system that acquires brain signals, analyzes them, and translates them into commands that are relayed to an output device to carry out a desired action. Any type of brain signal could be used to control a BCI system, albeit the most commonly used are electrical signals measured from electrodes on the scalp, on the cortical surface or in the cortex.[3]

A BCI system can be divided in 4 sequential components. The first component regards the task of signal acquisition. The second component deals with features extraction, meaning the task of processing the acquired signals and selecting features deemed valuable. The third component, feature translation, converts the features into the appropriate input for the output device. The final component of a BCI system is none other than the output device itself. The output device acquires data from the feature translation component, which might either provide it with a command, something that allows the user to actively operate the output device, or it could be provided with data used to parameterize the output device's passive operation.

There are multiple techniques used for the signal acquisition of a BCI system. These techniques can be sorted into Non-invasive, Semi-invasive and Invasive. Non-invasive techniques hold the lowest clinical risk, allowing signals to be acquired without requiring a surgical invasion to the user's body. Electroencephalography (EEG) is the most commonly used non-invasive method for studying brain signals, due to its lower cost and hardware portability. Other non-invasive methods that can be used included: Magnetoencephalography (MEG), Positron emission tomography (PET), functional Magnetic resonance imaging (fMRI) and near-infrared spectroscopy (fNIRS).

Electroencephalography (EEG)

EEG provides the recording of electrical activity of the brain from the surface of the scalp. Electrodes are placed on the scalp to pick up the electrical current generated by the brain.

When a neuron fires it forms a dipole, with a lower voltage at synapses and a higher voltage at the axon. If it's an inhibitory neuron, the dipole is flipped, with lower voltage at the axon and higher at the synapses. This voltage shift is caused when Na+ channels open on along the dendrite, causing a flood of positive electrons, this positive charge carry down the axon, opening more sodium channels, and causing an electric charge to carry down the axon, discharging at the synapse and releasing neurotransmitters along with it. When groups of neurons fire together their signals become measurable from the scalp, where an electrode is positioned, making it possible to measure the electrical activity of that cluster of neurons. By placing multiple electrodes to several different positions across the scalp, it becomes possible to fetch data from multiple clusters of neurons.

EEG data contains rhythmic activity, which reflects neural oscillations. Oscillations are described by frequency, power and phase. Each electrode records data to a separate channel, while the data of each EEG channel is described as a time-series, sampling the value of voltage with the passage of time. Using appropriate mathematical transformations, such as Fourier Transform or Wavelet transformations, the recorded data is analyzed to its spectral components. With the further use of band-pass filters, it's possible to analyze the data into alpha, beta, gamma, delta and theta brainwaves.

International 10-20 Electrode System

The 10–20 system or International 10–20 system[4] is an internationally recognized method to describe and apply the location of scalp electrodes in the context of an EEG exam, polysomnograph sleep study, or voluntary lab research. This method was developed to maintain standardized testing methods ensuring that a subject's study outcomes (clinical or research) could be compiled, reproduced, and effectively analyzed and compared using the scientific method. The system is based on the relationship between the location of an electrode and the underlying area of the brain, specifically the cerebral cortex.

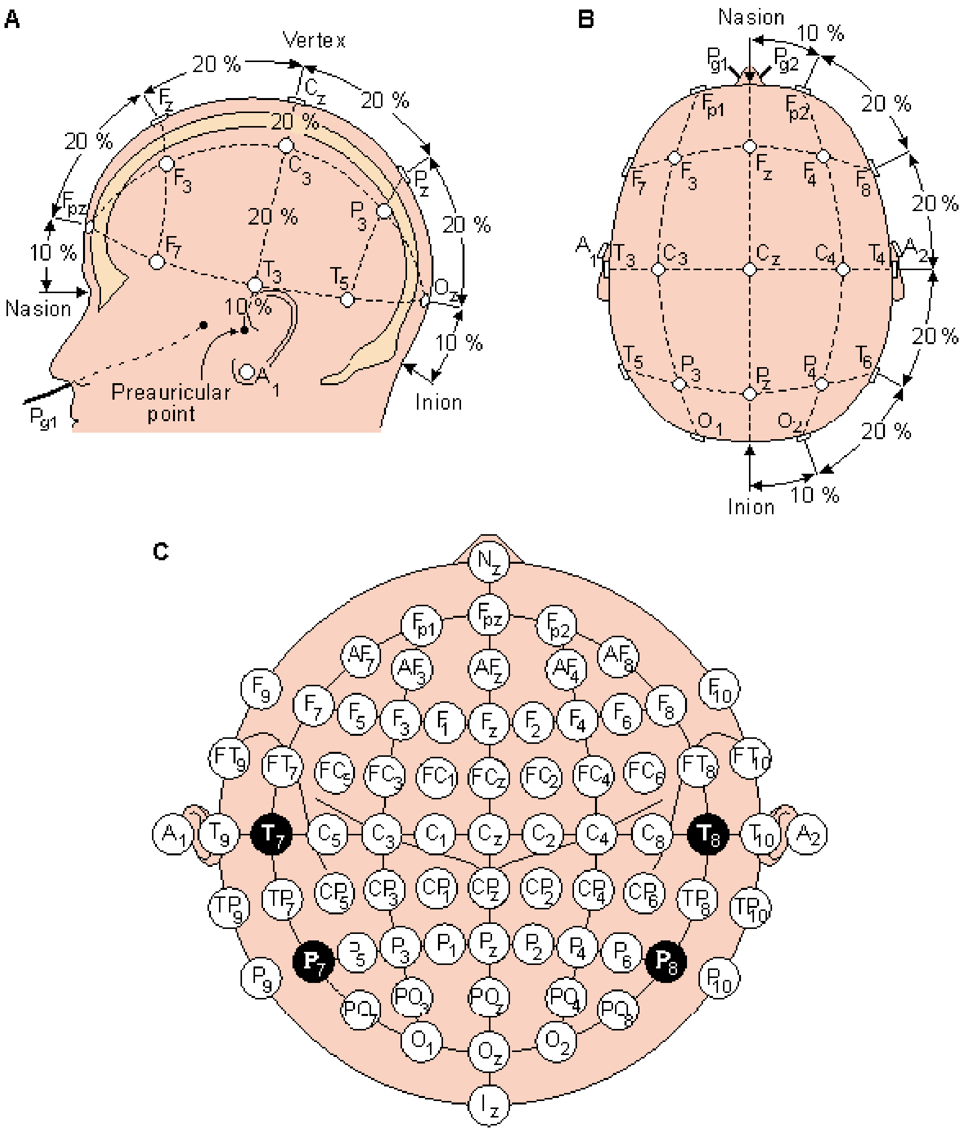


FIGURE 1 – INTERNATIONAL 10-20 ELECTRODE SYSTEM[4]

CLASSIFICATION PREDICTIVE MODELING

As part of the features translation component of a BCI system, it is common to make use of machine learning classification models in order to predict and assign class labels to the input data. The task of approximating a mapping function from input variables X to discrete output variables Y is referred to as Classification predictive modeling.

For example, recognizing whether a cell is cancerous or not based on a picture of that cell can be identified as a classification problem. This is a binary classification since there are only 2 classes as cancerous and non-cancerous. A classifier utilizes some training data to understand how given input variables relate to the class. In this case, known cancerous and non-cancerous cell images have to be used as the training data. When the classifier is trained accurately, it can be used to predict the class of an unknown email.

Classification algorithms are various and can be divided in two categories, lazy learners and eager learners. Lazy learners store the training data and classify testing data based on the most related data in their storage. Compared to eager learners, lazy learners have less training time but require more time for the actual predictions. Eager learners construct a classification model based on the training data before receiving data for classification. It must be able to commit to a single hypothesis that covers the entire instance space. Due to the model construction, eager learners take a long time for training and less time to predict.

Classification Algorithms

There is rich variety of classification algorithms[5][6] which can be used to solve classification problems. Choosing which of those algorithms to use depends on the application and the nature of the available data set. For instance, if the classes are linearly separable, the linear classifiers like logistic regression, Fisher's linear discriminant can outperform sophisticated models and vice versa.

Logistic Regression

Logistic Regression estimates discrete values based on a given set of independent variables. Simply put, it basically predicts the probability of occurrences of an even by fitting data to a logit function. A logistic regression algorithm chooses parameters that maximize the likelihood of observing sample values rather than minimize the sum of squared errors, like in ordinary regression.

Decision Trees

This algorithm splits the data into two or more homogenous sets based on the most significant attributes making the groups as distinct as possible.

Naive Bayes Classifier

This classification technique is based on an assumption of independence between predictors or what's known as Bayes' theorem. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

Building a Bayesian model is simple and particularly functional in case of enormous data sets. Along with simplicity, Naive Bayes is known to outperform sophisticated classification methods as well.

Bayes theorem provides a way of calculating posterior probability $P(c|x)$ from $P(c)$, $P(x)$ and $P(x|c)$. The expression for Posterior Probability is as follows (Figure 2).

Likelihood

Class Prior Probability

$$P(c \mid x) = \frac{P(x \mid c)\,P(c)}{P(x)}$$

Posterior Probability

Predictor Prior Probability

$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

FIGURE 2 – BAYES THEOREM

Where:

P(c|x) is the posterior probability of class (target) given predictor (attribute)
P(c) is the prior probability of class.
P(x|c) is the likelihood which is the probability of predictor given class.
P(x) is the prior probability of predictor.

SUPPORT VECTOR MACHINE (SVM)

In this algorithm, each data item is plotted as a point in a n-dimensional space, where n is the number of features available, with the value of each feature being the value of a particular coordinate.

The algorithm calculates a line that splits the data between two differently classified groups of data. This line is constructed such that the distances from the closest point in each of the two groups will be the farthest away. This line becomes the classifier, determining the class of any given data, depending on where the data lands on either side of the line.

When dealing with more than two dimensions, the “constructed line” is instead defined as a hyper-plane, giving form to a frontier that best segregates the two classes.

GRADIENT BOOSTING

Machine learning models can be fitted to data individually, or combined in an ensemble. An ensemble is a combination of simple individual models that together create a more powerful new model.

Machine learning boosting is a method for creating an ensemble. It starts by fitting an initial model to the data. Then a second model is built that focuses on accurately predicting the cases where the first model performs poorly. The combination of these two models is expected to be better than either model alone. Then this process of boosting is repeated multiple times, each successive model attempts to correct the shortcomings of combined boosted ensemble of all the previous models.

Gradient boosting[7] is a type of machine learning boosting. It relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error. The key idea is to set the target outcomes for this next model in order to minimize the error.

The target outcome for each case of data depends on how much changing that case's prediction impacts the overall prediction error:

If a small change in the prediction for a case causes a large drop in error, then the next target outcome of the case is a high value. Predictions from the new model that are close to its targets will reduce the error.
If a small change in the prediction for a case causes no change in error, then next target outcome of the case is zero. Changing this prediction does not decrease the error.

The name gradient boosting arises because target outcomes for each case are set based on the gradient of the error with respect to the prediction. Each new model takes a step in the direction that minimizes prediction error, in the space of possible predictions for each training case.

XGBoost Algorithm

XGBoost[8][9] stands for extreme gradient boosting. It is an extension to gradient boosted decision trees, specially designed to improve speed and performance.

XGBoost is used for supervised learning problems, where training data with multiple features $\mathbf{x_i}$ is used to predict a target variable $\mathbf{y_i}$. In the following paragraphs, the elements of Supervised Learning are explained, before proceeding to explaining the principles of XGBoost.

Model and Parameters

The **model** in supervised learning usually refers to the mathematical structure of by which the prediction $\mathbf{y_i}$ is made from the input $\mathbf{x_i}$. A common example is a linear model, where the prediction is given as a linear combination of weighted input features.

$$\hat{y}_i = \sum_j \theta_j x_{ij},$$

The prediction value can have different interpretations, depending on the task, i.e., regression or classification. For example, it can be logistic transformed to get the probability of positive class in logistic regression, and it can also be used as a ranking score when required to rank the outputs.

The **parameters** are the undetermined part that is calculated from data. In linear regression problems, the parameters are the coefficients ϑ.

Objective Function: Training Loss and Regularization

With judicious choices for $\mathbf{y_i}$, a variety of tasks may be expressed, such as regression, classification, and ranking. The task of training the model amounts to finding the best parameters ϑ that best fit the training data $\mathbf{x_i}$ and labels $\mathbf{y_i}$. In order to train the model, it's a prerequisite to define the **objective function** in order to measure how well the model fit the training data.

A salient characteristic of objective functions is that they consist of two parts: **training loss** and **regularization term**:

$$\text{obj}(\vartheta) = L(\vartheta) + \Omega(\vartheta)$$

Where **L** is the training loss function and **Ω** is a regularization term.

The training loss measures how *predictive* the model is with respect to the training data. A common choice of L is the mean squared error, which is given by

$$L(\theta) = \sum_i (y_i - \hat{y}_i)^2$$

Another commonly used loss function is logistic loss, to be used for logistic regression:

$$L(\theta) = \sum_i [y_i \ln (1 + e^{-\hat{y}_i}) + (1 - y_i) \ln (1 + e^{\hat{y}_i})]$$

The **regularization term** controls the complexity of the model, which is used to avoid overfitting.

Decision Tree Ensembles

The model of supervised learning chosen for the implementation of XGBoost is Decision tree ensembles. The tree ensemble mode consists of a set of classification and regression trees (CART). In figure 3, there is a simple example of a CART that classifies whether someone will like a hypothetical computer game X.

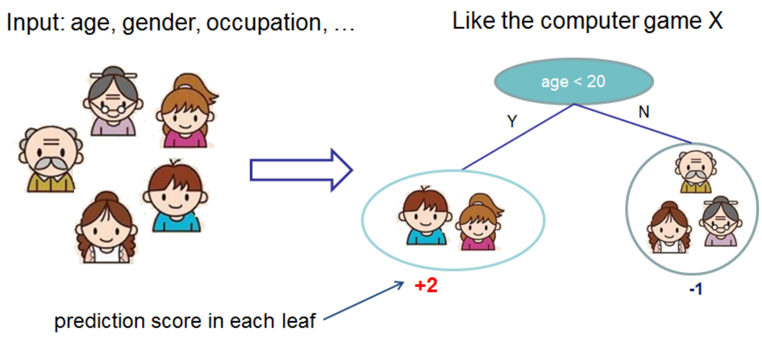


Figure 3 – Example of CART that classifies whether someone will like a computer game X

The members of the family in Figure 3 are being classified into different leaves, before they are assigned the score on the corresponding leaf. A CART is a bit different from decision trees, in which the leaf only contains decision values. In CART, a real score is associated with each of the leaves, which creates richer interpretations that go beyond classification. This also allows for a principled, unified approach to optimization.

Usually, a single tree is not strong enough to be used in practice. What is actually used is the ensemble model, which sums the prediction of multiple trees together.

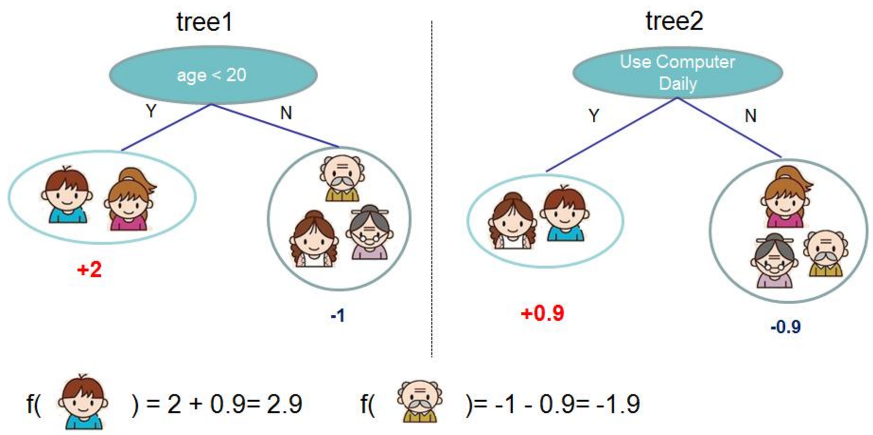


Figure 4 – Example Tree Ensemble of Two Trees

Figure 4 presents an example tree ensemble of two trees. The prediction scores of each individual tree are summed up to get the final score. As showcased by the example, it's important that the two trees endeavor to complement each other. Mathematically, the model gets the following form:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in \mathcal{F}$$

Where **K** is the number of trees, **f** is a function in the functional space **F**, and **F** is the set of all possible CARTs. The objective function to be optimized is given by

$$\text{obj}(\theta) = \sum_{i}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

Tree Boosting

With the model being introduced, it's important to train the algorithm. Something like that is achieved for supervised learning models by defining and optimizing and objective function.

Let the following be the objective function

$$\text{obj} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^{t} \Omega(f_i)$$

Primarily, it is required to calculate the parameters of trees. This requirement is translated to require learning the functions $\mathbf{f_i}$, each containing the structure of the tree and the leaf scores. Learning tree structure is harder than traditional optimization problems where it's possible to simply take the gradient. It is intractable to learn all the trees at once. Instead, an additive strategy is used: What's already learned is fixed before adding one new tree at a time. Writing the prediction value at step **t** as $\hat{\boldsymbol{y}}_{\boldsymbol{i}}^{(\boldsymbol{t})}$:

$$
\begin{aligned}
\hat{y}_i^{(0)} &= 0 \\
\hat{y}_i^{(1)} &= f_1(x_i) = \hat{y}_i^{(0)} + f_1(x_i) \\
\hat{y}_i^{(2)} &= f_1(x_i) + f_2(x_i) = \hat{y}_i^{(1)} + f_2(x_i) \\
&\dots \\
\hat{y}_i^{(t)} &= \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i)
\end{aligned}
$$

For each step, the tree that optimizes the objective function is added.

$$
\begin{aligned}
\text{obj}^{(t)} &= \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t)}) + \sum_{i=1}^{t} \Omega(f_i) \\
&= \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) + \text{constant}
\end{aligned}
$$

By further using mean squared error (MSE) as the loss function, the objective becomes

$$
\begin{aligned}
\text{obj}^{(t)} &= \sum_{i=1}^{n} (y_i - (\hat{y}_i^{(t-1)} + f_t(x_i)))^2 + \sum_{i=1}^{t} \Omega(f_i) \\
&= \sum_{i=1}^{n} [2(\hat{y}_i^{(t-1)} - y_i) f_t(x_i) + f_t(x_i)^2] + \Omega(f_t) + \text{constant}
\end{aligned}
$$

The form of MSE is friendly, with a first order term, usually called the residual, and a quadratic term. For other losses of interest, it is not so easy to get such a nice form. So in the general case, the Taylor expansion of the loss function up to the second order is used:

$$
\text{obj}^{(t)} = \sum_{i=1}^{n} [l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) + \text{constant}
$$

Where $\mathbf{g_i}$ and $\mathbf{h_i}$ are defined as

$$
\begin{aligned}
g_i &= \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)}) \\
h_i &= \partial^2_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})
\end{aligned}
$$

After the constants are removed, the specific objective at step **t** becomes

$$\sum_{i=1}^{n}[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)] + \Omega(f_t)$$

This becomes the optimization goal for the new tree. One important advantage of this definition is that the value of the objective function only depends on $\mathbf{g_i}$ and $\mathbf{h_i}$. This is how XGBoost supports custom loss functions. Every loss function can be optimized, including logistic regression and pairwise ranking, using exactly the same solver that takes $\mathbf{g_i}$ and $\mathbf{h_i}$ as input.

To define the regularization term, it's important to define the complexity of the tree **Ω(f)**. In order to do so, the definition of the tree **f(x)** is first refined as

$$f_t(x) = w_{q(x)}, w \in R^T, q: R^d \to \{1,2,\cdots,T\}.$$

Here **w** is the vector of scores on leaves, **q** is a function assigning each data point to the corresponding leaf, and **T** is the number of leaves. In XGBoost, complexity is defined as

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$$

After re-formulating the tree model, it becomes possible to write the objective value with the **t**-th tree as:

$$\begin{aligned}\text{obj}^{(t)} &\approx \sum_{i=1}^{n}[g_i w_{q(x_i)} + \frac{1}{2}h_i w_{q(x_i)}^2] + \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2 \\ &= \sum_{j=1}^{T}[(\sum_{i\in I_j} g_i)w_j + \frac{1}{2}(\sum_{i\in I_j} h_i + \lambda)w_j^2] + \gamma T\end{aligned}$$

Where $I_j = \{i|q(x_i) = j\}$ is the set of indices of data points assigned to the **j**-th leaf. Notice that in the second line, the index of summation was changed, because all the data points on the same leaf get the same score. The expression could be compressed by defining $G_j = \sum_{i\in I_j} g_i$ and $H_j = \sum_{i\in I_j} h_i$:

$$\text{obj}^{(t)} = \sum_{j=1}^{T}[G_j w_j + \frac{1}{2}(H_j + \lambda)w_j^2] + \gamma T$$

In this equation, $\mathbf{w_j}$ are independent with respect to each other, the form

$G_j w_j + \frac{1}{2}(H_j + \lambda)w_j^2$ is quadratic and the best $\mathbf{w_j}$ for given structure **q(x)** and the best objective reduction which can be acquired is as follows

$$w_j^* = -\frac{G_j}{H_j + \lambda}$$

$$\text{obj}^* = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j + \lambda} + \gamma T$$

The last equation measures how good a tree structure q(x) is.

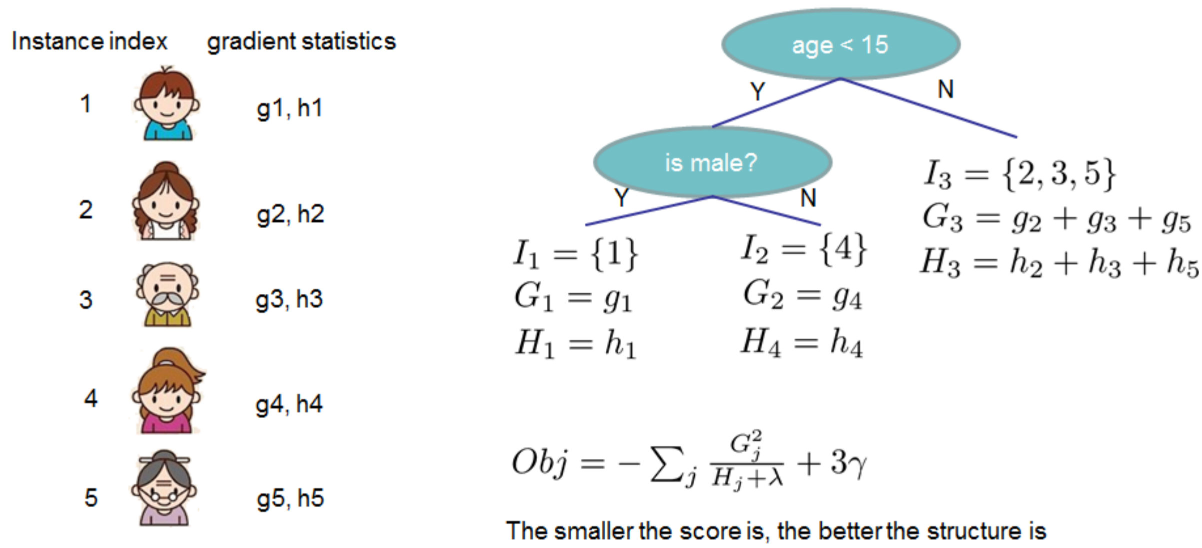


FIGURE 5 – DECISION TREE SCORE CALCULATION

Figure 5 explains how the scores can be calculated. Basically, for a given tree structure, the statistics $\mathbf{g}_i$ and $\mathbf{h}_i$ are pushed to the leaves they belong to. Then the statistics are summed together and the formula is used to calculate how good the tree is. This score is like the impurity measure in a decision tree, except that it also takes the model complexity into account.

Having defined a measure for determining the quality of a tree, it would have been ideal to enumerate all possible trees and pick the best one. In practice this is intractable, so it gets optimized one level of the tree at a time. Specifically, each leaf is split into two leaves, and the score it gains is

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}\right] - \gamma$$

This formula can be decomposed as

- The score on the new left leaf
- The score on the new right leaf
- The score on the original leaf
- Regularization on the additional leaf.

It is important to note that if the gain is smaller than γ, it would do better to not add a branch.

For real valued data, it continues by searching for an optimal split. To efficiently do so, it places all the instances in sorted order, like the following picture.

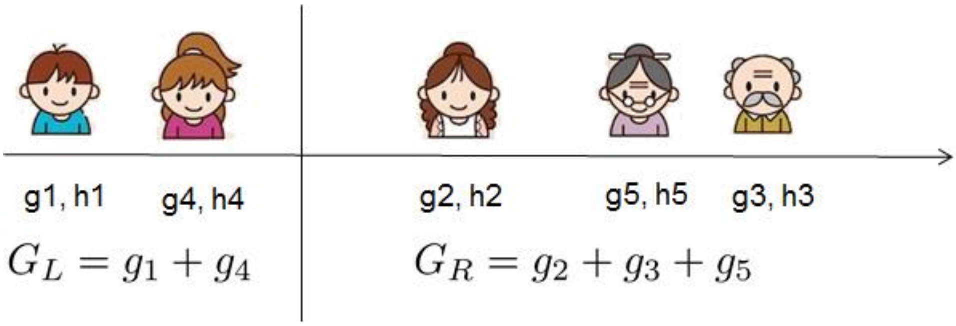


FIGURE 6 – INSTANCES OF THE EXAMPLE SORTED IN ORDER

A left to right scan is sufficient to calculate the structure score of all possible split solutions, which allows finding the best split efficiently.

SERIOUS GAMES FOR HEALTH AND MEDICINE

SERIOUS GAMES

A Serious game is defined as any piece of software that merges a non-entertaining purpose (serious) with a video game structure (game).[10]

SERIOUS GAMES CORE COMPONENTS

To better understand the concept of Serious Games[11][12][13] and the methods of developing a Serious Game for any application, including health and medicine, it is necessary to define some basic terms that are widely used for video game development and to explain the development process. Before anything, the developer team has to determine the tools, the technologies and the contents which have to be used in the game.

The tools can be separated into three main groups[14]: The game engine, the database and the design software. A game engine is a software framework, which provides a suite of visual development tools, used to bring the games logic and code into a playable state. The game engine is what is used to bring all the components together, allowing the game to be created. The database maintains all the data and information required by platform including player information, score, image files, animations, sounds, etc. The game engine imports and uses the data stored in the Database, while it can also record information within it for future reference. The design software includes all the tools used for engineering the data and mechanics of the game. It includes illustration software for the creation of image data and drafting the Graphical User Interface (GUI). It also includes text document editing tools for preparing the content or explaining the game’s mechanics. Design software might also include spreadsheets and any other tools used to organize work between the members of the developer's team and help break down the game construction circle into manageable portions.

The content can be defined as significant information which will be delivered to the players when the serious game is played. Content is provided by experts and converted to useful information according to the objective of the serious game.

Technology is the branch of knowledge that deals with the creation and the use of technical means and their interrelation with life, society, and the environment, drawing upon such subjects as industrial arts, engineering, applied science, etc. In this context, technologies might refer to the usage of special equipment and the branch of knowledge related to it. Two noticeable technologies in this field include Virtual and Augmented Reality.

For the Developers team to make the appropriate choices about their tools, content and technologies, they first need to identify the game objective and the game genre.

Game objective, in the case of serious games, is mainly focused on education, training and informing in an effective and incisive manner. Other objectives can be practice, testing, simulation, diagnosis and treatment. A serious game can contain more than one objective. For example, serious games which are related to health purpose are not only focused on edutainment but they also show many good results on diagnostic and treatment purposes as well.

Game genre is used to categorize video games based on their gameplay. Some of the proposed game genres are: Adventure games which involve exploration of, and interaction with, the environment as a main facet of gameplay; Sport games, like the name implies, are games based on sports which emulate traditional physical sports such as basketball, golf, football, etc.; Strategy games in which the players often autonomous decision-making skills have a high significance in determining the outcome; Exergames which is a term combining exercise and gaming, by merging exercise equipment with video games, aims to encourage people to exercise more by making the activity more fun; Puzzle games are short but addictive graphical games that usually require the player to solve a puzzle such as a maze, logical problem or positioning different pieces together; Action games which players are required to have good reflexes, hand-eye coordination and quick reaction times in order to overcome challenges such as combat, avoiding traps, jumping, running, completing tasks within a pressing time limit, etc; Role-playing games in which the player is given the role of a fictional character and is required to tend to that character's responsibilities. The player's character is often given skills and abilities represented by statistics. Gameplay involves the character exploring and completing quests that build up their statistics and possessions. The game can be single or multi-player; and Simulation games that attempt to realistically mimic the conditions of a particular environment or activity.

Review of serious games for health

In the following section a review of serious games in health is presented. The review includes games from review articles, commercial and online sources. A short summary is provided of each game, indicating their most relevant characteristic by year of publication.

Table 1. Review of Serious games in health[14]

Developer/Author	Game name	Description
Johnston & Duskin (2004)	Ben's Game	A Serious game designed to help kids fight against cancer. The game was initiated by Ben Duskin, nine years old, who is in remission from Leukemia.
MIT Teacher Education Program (2004)	Outbreak @ MIT	A handheld game occurring within the premises of MIT that aims at exploring an epidemic of avian influenza being triggered in college.
Montreal Science Centre (2004)	Sleep: A to Z	A serous game which aims to present the operation and the importance of the different stages of sleep by a set of mini games.
Respondesign (2004)	MayaFit	An exergame dedicated to physical training. Many fitness exercises are offered by virtual coach.
Nintendo (2005)	Brain Age 2	A serious game that evaluates the brain age between 20 to 80, by approximating the range of brain responsiveness.
Archimage, Inc. (2006)	Escape from Diab	An adventure game in healthy eating and exercise that focuses on obesity and type 2 diabetes prevention. The game guides the player through goal setting, problem solving, energy balance, and other game-play activities.
Believe in Tomorrow National children's Foundation, Ltd. (2006)	FreeDive	A serious game that helps chronically ill children cope with pain and anxiety by distracting them while they undergo painful medical treatments.
Gameloft (2006)	Brain Challenge	A collection of serious games aiming to exercise the brain and keep intellectual muscles in shape. It also checks the performance and monitors the overall results of the brain activity with the daily follow-up. The game is based on four subjects: logic, math, memory and visual.

TABLE 1. REVIEW OF SERIOUS GAMES IN HEALTH(CONTINUE)

Developer/Author	Game name	Description
Janomedia (2006)	Terveelinen Ateria	A serious game aiding practical nurses (PN) and school staff train in the practical aspects of preparing meals for people with different nutritional requirements.
McGIll University (2006)	Grow Your Chi	A serious game designed to increase self-esteem by displaying a positive ability or function of the player.
BreakAway, Ltd(2007)	Pulse!! The virtual clinical learning lab	An immersive virtual learning space where health care professionals can train the clinical skills by dealing with injured patients, bioterrorism or other catastrophes.
Edheads(2007)	Virtual Knee Replacement Surgery	A science game for patients and their caregivers to take on the role of a surgeon and complete a knee replacement surgery while learning about the procedure, the technology, and health risks and benefits.
Fatworld.org(2007)	Fatworld	A videogame that explores the relationships between nutrition, obesity and socioeconomic factors in the contemporary U.S. Budgets, subsidies, regulations and physical world characteristics are taken into account.
Intelligent System Co., Ltd.(2007)	Otona no DS Kao Training	A brand game released only in Japan. The training software instructs the user on how to perform several facial exercises that have a goal of keeping your face healthy looking and wrinkle free.
Nintendo EAD (2007)	WiiFit	An exergame consisting of activities utilizing the Wii Balance Board peripheral. Wii Fit Plus, an enhanced version of the original Wii Fit, was launched in 2009 by adding new exercises and tools to personalize the exercise routine.
Nordic Innovation Centre (2007)	Valion Energiasummaaja	A nutrition game aimed to build a healthy and balanced meal. The effects of the meal on blood sugar are shown, and possible improvements in the meal are suggested.
	MC Urho	A game that contains information regarding lifestyle effects on health and aims to teach young people about the effects of smoking, high blood pressure and cholesterol.
	MoFun Circus	An exergame in which players cycle to capture falling objects in the game. A camera is used to follow the users and display the activity on-screen.
SEGA Corporation (2007)	Mind Quiz (Nounenrei)	Series of 16 mini-games for mental training, aiming to measure and improve particular parts of the player’s brain, such as one’s brain age and its brain stress degree.

TABLE 1. REVIEW OF SERIOUS GAMES IN HEALTH(CONTINUE)

Developer/Author	Game name	Description
Anderson (2008)	-	A simulation game that was created to help nursing students to identify the roles and responsibilities of multidisciplinary teams of professionals when caring for a Maternal-Child Health (MCH) patient and family in crisis.
Glasgow Caledonian University (2008)	Nurse Education	A virtual learning environment in Second Life, a massively multiplayer online role-playing game (MMORPG), for use in nurse education.
Hatfield D. (2008)	My Stop Smoking Coach	An educational game for smokers to quit immediately and permanently. The game is run on several platforms including iPhone, Mobile (Jave ME) and Nintendo DS.
Imperial College London Faculty of Medicine (2008)	Game-based Learning for Virtual Patients	A region in Second Life provides a learning space where virtual patients suffering five different respiratory illnesses(such as lung cancer or pneumonia) can be diagnosed, investigated and treated by players wanting to perform roleplaying learning activities under the feedback and guidance of medical staff.
Mili, Barr, Harris & Pittiglio (2008)	VI-MED	A Virtual training to be used as a precursor and as a supplement to real practical training.
Sliney & Murphy (2008)	Medical Simulation Training Program (JDoc)	A computer-aided junior simulator used for training and teaching junior doctors their interpersonal communication and decision making skills, and to ease the transference of the medical information available to them.
The Partnership for Food Safety Education (2008)	The Food Detectives Fight BAC game	A web base game for 8-12 year old kids to learn about foodborne illness.
TruSim (2008)	Triage Trainer	A serious game to train in triage, the process of prioritizing the treatment of multiple casualties based on the severity of their injuries.
Vermont department of health (2008)	Khemia	A serious game designed to help people quit smoking. The game provides both a distraction from cigarette cravings and personalized support for quitting through integrated MyQuitKit tool.
Warner Bros. Entertainment, Inc. (2008)	Pamoja Mtaani	A video game that simulates real life settings in Kenyas capital Nairobi aims to teaching Kenyan youth how to avoid contracting HIV.
BBG Entertainment GmbH (2009)	Train Your Brain with Dr. Kawashima	A brain-exercising game with 30 specifically designed and scientifically tested exercises. Along with the goal of the game, it explains how the brain will be activated and developed by the training.

TABLE 1. REVIEW OF SERIOUS GAMES IN HEALTH (CONTINUED)

Developer/Author	Game name	Description
Blitz Games Studios, Ltd. (2009)	The Biggest Loser	A health and lifestyle game. Based on the hit NBC reality TV show, USA. The game mirrors the formal of the show by featuring intense training routines, weekly challenges, nutritional goals and information and the iconic weekly weigh-in and elimination from the show.
Burke, McNeill, Charles, Morrow, & Crosbie (2009)	Arrow Attack	A serious game developed for bimanual rehabilitation (both arms).
	Catch task	A serious game for upper limb stroke rehabilitation (focused on bilateral rehabilitation)
	Rabbit Chase	A serious game developed for single arm rehabilitation (either right or left arm).
	Whack a Mouse	A serious game designed to encourage movement and to improve the accuracy and speed of the upper limb movement.
	Virtual Vibraphone	A serious game that uses Nintendo Wii remote controllers for wrist and arm rehabilitation.
Collision Studios (2009)	Daisy Fuentes	A fitness game of Pilates exercise, a system of exercise created in the 1920s by Joseph H., features a 3D Avatar of Daisy Fuentes who performs the exercises with the player.
Deponti, Maggiorini, & Palazzi (2009)	DroidGlove	A ubiquitous game therapy for wrist rehabilitation. The exercise has to be done while holding the smartphone in the hand. The performance will be automatically recorded for the doctor supervision.
Hopelab (2009)	Re-Mission	A video game with 20 levels that takes the player on a journey through the body of young patients with different kinds of cancer. The main aim is to engage young cancer patients through entertaining game play while impacting specific psychological and behavioral outcomes associated with successful cancer treatment.
Keele University (2009)	-	A System which aims to train pharmacists by using a virtual patient. Traits such as race, age and gender are taken into account in the treatment of patients to let learners understand the clinical significance.
Kim JA, Kang, Yang & Kim D (2009)	A Sensory Gate-Ball Game	PC-based 3D graphics game designed for aged people; It uses a realistic gate-ball stick and balls as interfaces. In the game, players use the same stick and ball as the real gate-ball.
Laikari (2009)	Fitness Adventure	A location-aware fitness game which takes advantage of a variety of mobile phones, location information and Bluetooth GPS receivers; combines mobile games with exercising outdoors.

TABLE 1. REVIEW OF SERIOUS GAMES IN HEALTH (CONTINUED)

Developer/Author	Game name	Description
Learning Games Lab (2009)	Science Pirates: The Curse of Brownbeard	An adventure game allowing the child to learn about food safety and the underlying scientific principles. The adventure is made up of different challenges: problem solving, scavenger hunt, etc.
Lightning Fish Games (2009)	NewU Fitness First Personal Trainer	A fitness game featuring both structured exercise programs and nutrition programs. The fitness programs are designed by the Fitness First gym chain, the nutrition programs are associated with the You Are What You Eat television series.
McKanna, Jimison, & Pavel (2009)	21 Tally	A collection of 2D games used to detect divided attention unobtrusively, by using performance on a computer game designed to force players to attend to different dimensions of attention simultaneously in order to succed.
Persuasive Games LLC (2009)	Killer Flu	A game about seasonal and pandemic flu attempting to explain how flu really mutates and spreads, and how challenging it can be for a deadly strain to affect a large population geographically. The player takes the role of the flu itself, trying to mutate and spread in a variety of conditions.
QOVEO (2009)	Prevenir la gripe A H1N1	A serious game to raise awareness about H1N1 virus prevention. It is a shooting game where the player has to destroy viruses. However, in order to reload their weapon, players have to answer questions related to the virus.
RANJ Serious Games (2009)	The Great Flu	A serious game aims to raise the awareness of similar outbreaks by having the player control the deadly Gamers Flu. The goal of the game is to control a possible pandemic by select options to apply actions or assign research teams in order to stop the flu.
Raylight S.r.l (2009)	Train your Sense	A serious game to train the visual and aural senses through 22 exercises. The game lets the player try first on his weaknesses and improve his skills through targeted training.
Succubus Interactive (2009)	Happy Night Club	A serious game which aims to sensitize teenagers about the risk of the over-alcohol and binge drinking. Player performs as a secret agent sent on a mission in a night club he needs to investigate without drinking too much, in order to stay clear enough to fulfill the mission.
Virtual Heroes, Inc (2009)	Zero Hour: Americas Medic	A 4.8 million dollar serious game, designed by George Washington University's office of homeland security and Virtual Heroes, Inc., which aims to train and exercise first responders to respond to mass casualty incidents such as earthquakes or terrorist attacks.

TABLE 1. REVIEW OF SERIOUS GAMES IN HEALTH (CONTINUED)

Developer/Author	Game name	Description
Vitnen & Leikas (2009)	Virku – Virtual Fitness Centre	A system that allows users to exercise in a virtual environment. The game is controlled by a user interface based on an exercise cycle, and users may practice individually or in a group.
Anchor Bay Entertainment (2010)	10 Minute Solution	An exergame which allows players to construct their own workout regimens based on 10 minute exercise sessions. Exercises are organized into 3 major categories: Cardio Boxing, Mixed Games, and Aerobics, each featuring several different workout games to choose from.
Atkinson & Narasimhan (2010)	-	A medical diagnostic gaming environment that is used to gather Parkinson's patient information in a casual environment. The system employs the novint falcon human interface device (Novint Technologies, Inc., 2011) to guide a patient within the game.
Bartolome, Zorilla, & Zaripain (2010)	-	A serious game to analyze the behavior and promote certain social skills (conversation, negotiation, etc.) of people with Neurological development Disabilities
Break Away, Ltd (2010) & Medical College of Georgia school of Dentistry faculty and students (2010)	Virtual Dental Implant Training Simulation Program	A 3D virtual environment for students to train in the correct decision-making protocol to determine patient preparation (both physical and mental) for dental implant surgery.
Clawson, Patel, & Starner (2010)	DITS	A mobile phone game similar to Dance Dance Revolution (DDR). Instead of using a dance pad, DITS uses wireless 3-axis accelerometers that are worn around the player's ankles and uses a mobile phone to control the game and to display graphics.
Electronic Arts, Inc. (2010)	EA Sports Active	An exergame focused on cardio exercise. Players can choose from three week or nine week programs which are rigidly guided systems that track players through the range of weeks selected.
Finkelstein, Nickel, Barnes, & Suma (2010)	Astrojumper	A stereoscopic virtual reality exergame for children with autism. During the game, virtual space-themed objects fly forward toward the user who must use their own physical movements to avoid collisions.

TABLE 1. REVIEW OF SERIOUS GAMES IN HEALTH (CONTINUED)

Developer/Author	Game name	Description
Fishing Cactus (2010)	R.O.G.E.R	A serious game for patients who has a lack of logic and organizational skills (typically post-stroke patients, Alzheimer, hemi-negligent patients, etc.).
Gago, Barreira, Carrascosa, & Segovia (2010)	Nutri-Trainer	A collection of serious games about nutritional hearth following professional recommendations, cooperating with doctors and nutritionists to give coherence to the information collected in the nutritional databases.
Grau, Tost, Campeny, Moya, & Ruiz (2010)	-	A neuropsychological rehabilitation game that allows patients to navigate through the virtual environment and perform cognitive tasks.
HopeLab (2010)	Zamzee	An online rewards program and game-like experience powered by your physical activity. Players wear a device with an accelerometer that monitors their movement and translates it to the points that can be then redeemed with both digital and real-world rewards.
Innovation in Learning, Inc. (2010)	CliniSpace	A medical training game for healthcare professionals focusing on clinical diagnosis and patient management. Players in the role of a doctor, who may consult medical records to a patient, plan an operation or provide a clinical consultation.
KTM Advance (2010)	Alphega Game	An education game to train pharmacists focus on patients observation by counseling with the virtual patients in game.
Miller (2010)	Market Virtual Patient Care Simulation (MUVE)	Patents simulations for students and professionals (nurses, pharmacists, paramedics, emergency medical technicians, social workers, etc.) training.
Sabri et al (2010)	-	A serious game designed to train orthopedic surgical procedures to orthopedic surgical residents.
Skills2Learn, Ltd. (2010)	Nursing and Midwifery	A program that helps nurses and midwives increase their ability to assess patients. The interactive scenario is based on the simulation of the 36 weeks of pregnancy realistically.
TruSim (2010)	Patient Rescue	A serious game which supports health professionals to recognize the signs of patient deterioration, use set protocols to assess a patient's condition and intervene effectively.
Ubisoft Divertissements, Inc. (2010)	Your Shape: Fitness Evolved	This console game focuses entirely on fitness routines led by virtual trainers, and is divided into structured personal training, pick-up fitness classes, and active gym games.
Vidan, Chittaro, & Carchietti (2010)	EMSAVE	A serious game for training in emergency medical procedures concerning disabled patients. It allows users to experience emergency situations involving disabled persons.

TABLE 1. REVIEW OF SERIOUS GAMES IN HEALTH (CONTINUED)

Developer/Author	Game name	Description
Visual Imagination Software (2010)	Chirurgie Simulator	A surgery simulation in which players take on the role of a surgeon at a hospital. The surgical procedures include operating on fractures, removing an inflamed appendix or tonsils, treating infected gall bladders, attending to varicose veins, repairing hernias, restoring vision in cataract procedures and dealing with the injuries of a road traffic accident.
Wang, Sourina, & Nguyen (2010)	Brain Chi and Dancing Robot	The EEG-based concentration games named Brain Chi (2D) and Dancing Robot (3D) are developed for concentration level control.
Zumba Fitness LLC (2010)	Zumba Fitness	An interactive exercise program helps to perform high calorie-burning workouts. The game concept is based on calorie-burning dance fitness-party. Zumba Fitness is available in PlayStation 3, Wii and Xbox 360 platform.
Association RMC / BFM (2011)	Staying Alive	A 3D simulation aiming to teach how to deal with emergency situation cardiac arrest.
Botella et al. (2011)	-	A mobile phone game for the treatment of cockroach phobia. The objective is to reduce the level of fear and avoidance.
CCCP (2011)	LudoMedic	An educational game aims to teach children to prepare for an MRI, have surgery, or undergo chemotherapy.
De Bortoli & Gaggi (2011)	PlayWithEyes	This serious game aims to test children eyes while they are having fun playing with Lea symbols and images taken from popular cartoons, using a touch interface.
Fuchslocher, Niesenhaus, & Krmer (2011)	Balance	A health game developed to optimize the self-management of teenagers with diabetes mellitus type-I.
Imbeault, Bouchard, & Bouzouane (2011)	-	A serious game created specifically for patients suffering from Alzheimer; advances in the field of artificial intelligence such as activity recognition and guidance to offer optimal experience through the training sessions.
Lakeside Center for Autism (Microsoft) (2011)	Kinetix Academy	A series of several games for autism by using Kinect in order to stimulate: motor development, use of language and comprehension, use of different cognitive processes and social interactions.
Lin (2011)	-	An augmented reality serious game for facilitating the patients to execute rehabilitative activities without any geographical or time limitations.
Moya, Grau, Tost, Campeny, & Ruiz (2011)	-	A 3D virtual environment for neuro-rehabilitation of the upper limb. Patient wears a special suit with sensors integrated to move one of their arms trying to simulate concrete daily actions, such as grasping a bottle, opening a door or putting a book on a shelf.

TABLE 1. REVIEW OF SERIOUS GAMES IN HEALTH (CONTINUE)

Developer/Author	Game name	Description
Nauta & Spil (2011)	-	A educational diabetes game which aims to enhance a healthy lifestyle by educating and coaching self monitoring reinforcements and observational learning.
Public Health Agency of Canada (2011)	Buffet busters	A game developed for grade 5 educators and students to promote infectious disease awareness and to introduce concepts related to food and waterborne infectious diseases as well as basic principles of epidemiology.
Queiros et al. (2011)	-	A low-cost laparoscopy simulator, for novice surgeons training, which is able to monitor and assist the trainee's laparoscopy surgical movements.
Red Hill Studios (2011)	-	A collection of therapeutic games for Parkinson's disease patients to increase their balance, designed by Red Hill Studios and the School of Nursing, University of California San Francisco. The games are played by performing movements which are known to be beneficial for balance control and the movements are captured and processed by the system.
Scarle et al. (2011)	Match-3	A serious game designed to combat childhood obesity. The Wii-mote is being used for a rowing action which propels a vehicle forward, while direction is altered by leaning left and right on the Wii Balance board.
Schnauer, Pintaric, Kosterink, & Vollenbroek (2011)	-	A chronic pain rehabilitation game provides multimodal interaction including full body motion capture by the use of Kinect, and other bio-signal capture devices. Patients can manage their state and train physically on their own.
The Diablotines (2011)	L' Affaire BIRMAN	This educational game aims to help patients treated by insulin to practice the technique of functional insulin therapy.
Urturi (2011)	-	A serious game for Autism Spectrum Disorder (ASD), oriented to first aid education: what to do in certain situations, basic knowledge about healthcare, medical specialties, etc.
Van E, Peper, van A, & Salverda (2011)	-	A set of computer games to help children with spastic cerebral palsy (CP) to loosen the coupling between their hands. Patients were challenged to move both hands simultaneously in various phase relations.
Vazquez (2011)	-	A serious game to promote hand hygien among health care professionals and citizens. The correct hand hygien practice and its indications depends on the health care typology, the environment, and the clinical task, focusing specially on five movements for hand hygiene.
Applied Research Associates, Inc. (2012)	HumanSim	An immersive world where doctors and nurses train to learn the nuances of complex, unusual or other error-prone tasks until they become experts.
Cagatay, Ege, Tokdemir, & Cagiltay (2012)	-	A 3D game developed for speech and language disordered children. The game is used during the treatment process of Turkish children with language disorders.

TABLE 1. REVIEW OF SERIOUS GAMES IN HEALTH (CONTINUE)

Developer/Author	Game name	Description
Chan, Qin, Chui, & Heng (2012)	-	A computer based surgical simulator which aims to train an ultrasound-guided needle placement which is a key step in a lot of radiological intervention procedures such as a biopsy, local anesthesia and fluid drainage.
e-Learning Studios (2012)	The iSpectrum	This serious game aims to improve the work based social skills and relevant work skills of people with high functioning Autistic Spectrum Disorders (ASD) and Asperger's disorder.
EMCO3 (2012)	MD Advisor	A serious game that allows medical students to test their skills as future doctors such as performing the differential diagnosis in a virtual doctor's office.
GENIOUS Interactive (2012)	Voracy Fish	A serious game for the upper limb rehabilitation. Movements are captured by Kinect. The player dives into a sea universe, looking for some treasures and evolves while devouring other fish to become the strongest.
IKARE (2012)	MUCOPla y	A learning game to help caregivers, sufferers and families of cystic fibrosis. The game provides information about the right gestures and the means to validate their knowledge about cystic fibrosis and its care required.
Milo Foundation (2012)	Miloland	A serious game for children with a language delay, aged 9-11 years old by with a mental age of 5-6 years, i.e. starting literacy.
MIRROR project (2012)	CLinIC	The serious games focused on difficulty communication between nursing career staff and patients/residents. These games aim to foster reflection around difficult dialogues and to maximize learners ability to self-regulate their training.
Nike + Kinect Training (2012)	Nike+ Kinect Training	An exergame that combines fitness with gaming elements focusing on personalizing training and providing statistics.
SAIC, Inc. (2012)	OLIVE	This serious game aims to train the medical professionals with a set of training data with different scenarios in a virtual hospital.
University Medical Center Utrecht (2012)	Air Medic Sky 1	The interactive bio-feedback game consists of mini-games and lectures which describe the basic concepts required for efficient communication and teamwork resulting in patient safety.
Verduin et al. (2012)	-	A computer simulation designed for alcohol use disorders (AUDs) to practice relapse prevention skills.
University of Auckland (2013)	Sparx	A free online computer game intended to help young people with mild to moderate depression stress or anxiety. Through the game, this e-therapy will teach them how to resolve their issues on their own, according to a talking psychotherapeutic approach called cognitive behavioral therapy.
University of Almería (2017)	Stigma-Stop	A serious game that presents four characters who suffer from different mental disorders, namely depression, schizophrenia, bipolar disorder and panic disorder with agoraphobia. The objective of the player is to convince the characters toward a common goal, which is to participate in a video game design contest.

Methods & Materials

The subject of this research can be separated in three major segments. The first segment deals with signal acquisition[15] using a brain computer interface device to record brain activity from the head of a user. The second segment deals with isolating the EEG Data from the recorded signals and classifying it to a three dimensional vector based on the VAD Model. The last segment fetches the vectors from segment two, using them as parameters within the context of a serious game. The available design options for these parameters are analyzed with regards to the improvements and changes they would deliver to the "serious" and "gameplay" aspects of the game. A flowchart of the classifier's usage environment can be seen in figure 7.

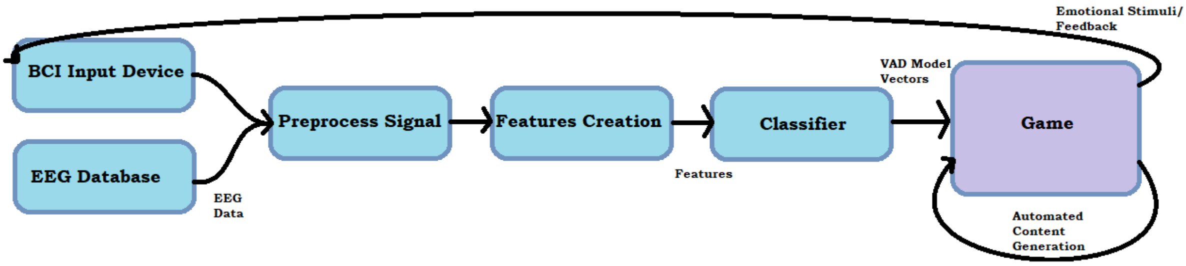


Figure 7 – Classifier's Usage Environment Flowchart

Data Acquisition

For the purpose of signal acquisition, an OpenBCI 8-channel Cyton Biosensing Board was designed[16] to be used within the context of this project, using 5 mm Comb Reusable Dry EEG Electrodes. However, due to technical difficulties regarding the shipment of the mentioned hardware, the signal acquisition has not been attempted in a practical level. Instead, it was deemed necessary to make use of an existing database in order to facilitate experimentations respecting the classification of EEG Data to Emotions.

Database Selection

Seven datasets were considered, before DREAMER was selected for the context of this experiment. The datasets were evaluated based on their size, their accessibility, their number of test subjects and their method used for gathering data.

"EEG Brainwave Dataset: Feeling Emotions"[17] contains preprocessed EEG Data, matched to Positive or Negative experiences. The dataset was dismissed due to the exclusion of raw unprocessed data, as well as due to the small number of subjects used to generate this dataset (1 Male Subject).

"EEG Brainwave Dataset: Mental State"[18] this dataset classifies EEG Data to three labels, "Relaxed", "Neutral" and "Concentrating". The dataset was excluded due to the small number of subjects used to generate this dataset (1 Male Subject, 1 Female Subject).

“**K-EmoCon”**[19] is a multimodal dataset, which includes EEG Series matched to emotions. The dataset was generated from a sample of 32 Subjects, asked to debate in random pairs.

“Seed”[20][21] is a preprocessed dataset generated from a sample of 15 Subjects (7 Males, 8 Females). The data was gathered while the subjects were required to watch 15 different short films. Due to the nature of films and serious games both being based on the interaction of the subject with Audio-Visual stimuli, this dataset was among preferred, however it wasn’t included due to the low accessibility of this dataset and the lack of raw unprocessed data. This dataset, pairs EEG features with three labels: Happy, Neutral, Sad.

“Seed-IV”[22] an extension to SEED but used to classify between: Happy, Neutral, Sad and Fear. It makes use of additional features, generated through Eye-tracking devices. This dataset was excluded for the same reason as SEED.

“DREAMER”[23] is a multimodal database consisting of electroencephalogram (EEG) and electrocardiogram (ECG) signals recorded during affect elicitation by means of audio-visual stimuli. Signals from 23 participants were recorded along with the participants self-assessment of their affective state after each stimuli, in terms of valence, arousal and dominance. All the signals were captured using portable, wearable, wireless, low-cost and off-the-shelf equipment that has the potential to allow the use in everyday applications. This dataset was selected for the experiments in this project.

“DEAP”[24] is a multimodal dataset for the analysis of human affective states. It includes the electroencephalogram (EEG) and peripheral physiological signals of 32 participants, recorded as each watched 40 one-minute long excerpts of music videos. The Participants rated each video in terms of levels of arousal, valence, like/dislike, dominance and familiarity. This dataset is a candidate for future experimentation; DREAMER was preferred against this dataset, due to the faster response of the dataset’s author, regarding the utilization of the dataset for academic research.

Database Description

For the construction of DREAMER dataset, Audio and visual stimuli in the form of films was employed to elicit emotional reactions to the participants of the study. A dataset consisting of 18 film clips, selected and evaluated by Gabert Quillen et al.[25] was utilized for eliciting emotions. These film clips contain cut out scenes from different films that have been shown to evoke a wide range of emotions. From these 18 film clips, two of each targeted one of the following nine emotions: amusement, excitement, happiness, calmness, anger, disgust, fear, sadness and surprise. The length of the film clips was between 65 to 393 seconds (M = 199s), which is considered as sufficient since, according to psychologists, video stimuli between 1 to 10 minutes is capable of eliciting single emotions[26],[27]. Nevertheless, the emotional state of a person may change over time, especially when video stimuli of larger length is used. To avoid contaminating data recordings with multiple emotions, only the recordings captured during the last 60 seconds of each film clip were used for further analysis.

For the acquisition of EEG signals, electrodes were placed to the following sixteen locations according to the International 10-20 system: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, M1 and M2. The mastoid sensor at M1 acted as a ground reference point for comparing the voltage of all other sensors, while the mastoid sensor at M2 was a feed-forward reference for reducing external electrical interference, and thus data for M1 and M2 are not stored within the database. The sampling rate was set to 128 Hz.

The ECG data contained within the dataset was not used within the context of this experiment. The test subjects, were given three rating scales, for Arousal, Valence and Dominance and they were instructed to rate each film clip in each of those scales, with a value from 1 to 5.

The process above has lead to the creation of a dataset that's formatted as explained below. For each of the 23 valid participants, there are 18 different data structures, one for each film. These data structures include the participant's ratings for all three rating scales, as well as 14 arrays containing the EEG Samples for the duration of each film, one array for each different electrode channel. The EEG Data for each participant in the dataset is separated in baseline features and trial features. The Baseline features are computed from the EEG recordings of the last 4s of a neutral film clip shown before each affect eliciting film clip. An additional data structure containing the ECG signals data was also included as well as information regarding the gender and age of each participant.

Three different binary classification schemes were defined: the classification between low/high arousal (calm/excited), low/high valence (unpleasant/pleasant), and low/high dominance (without control/empowered). Based on that classification scheme, the 5-point rating scale that was used by the participants was threshold into two classes (low and high). Before proceeding to the classification, it was deemed appropriate to scrutinize the class distributions for all participants after conversion to this two-class rating score system. The balance of the class distributions is a significant factor for selecting the methods of classification and validation.

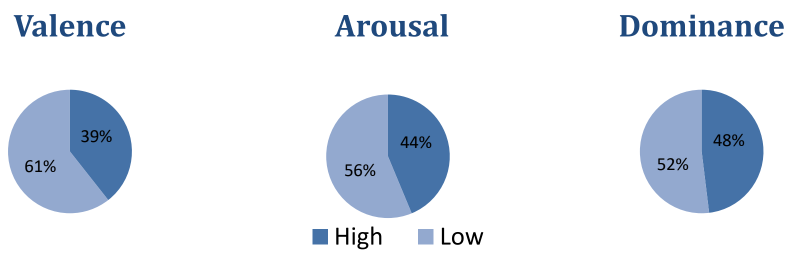
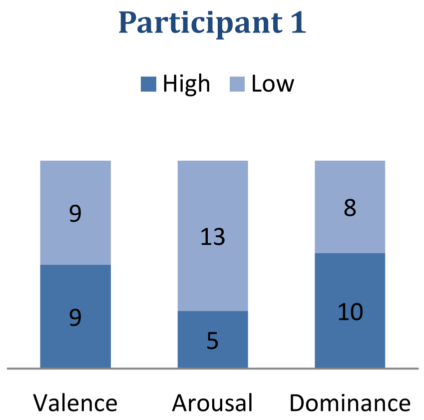
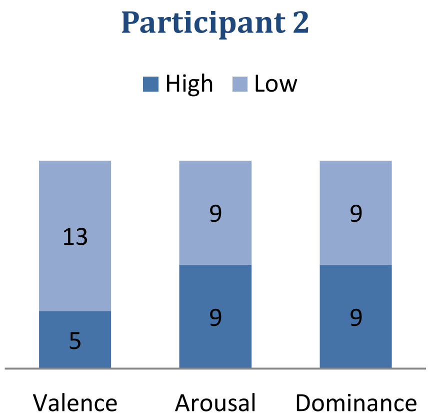
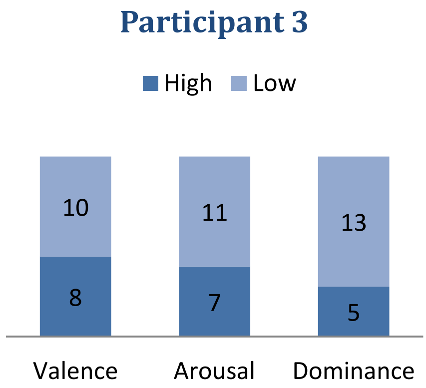
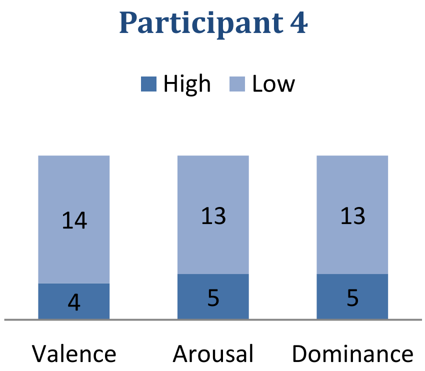


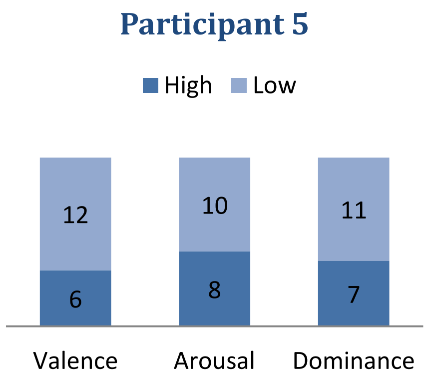
FIGURE 8. OVERALL CLASS DISTRIBUTION ACROSS ALL PARTICIPANTS AFTER CONVERSION TO A TWO-CLASS RATING SCALE

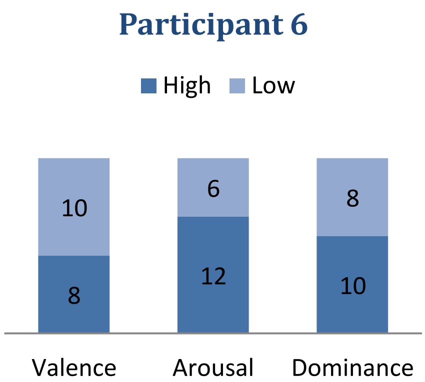

Participant 1
High
Low
9
9
13
5
8
10
Valence
Arousal
Dominance

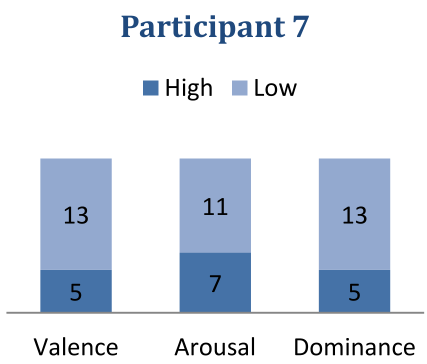

Participant 2
High
Low
13
5
9
9
9
9
Valence
Arousal
Dominance

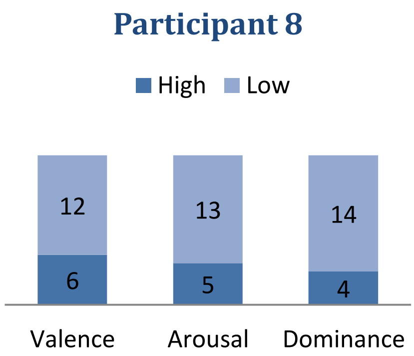

Participant 3
High
Low
10
8
11
7
13
5
Valence
Arousal
Dominance

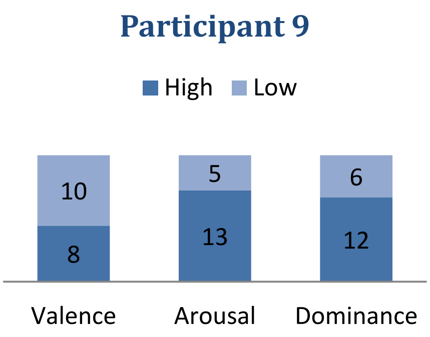

Participant 4
High
Low
14
4
13
5
13
5
Valence
Arousal
Dominance

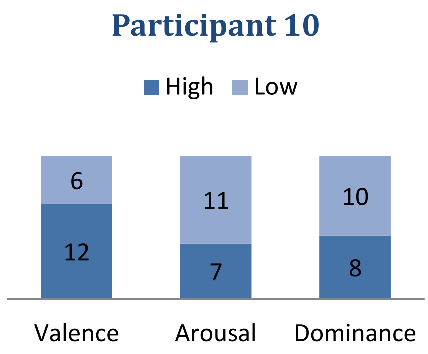

Participant 5
High
Low
12
6
10
8
11
7
Valence
Arousal
Dominance

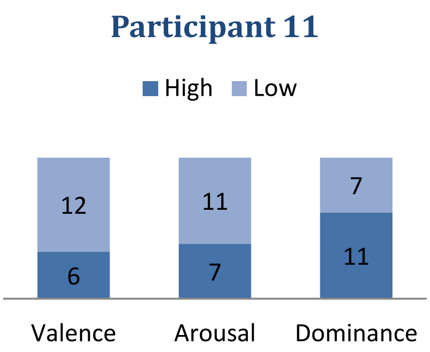

Participant 6
High
Low
10
8
6
12
8
10
Valence
Arousal
Dominance

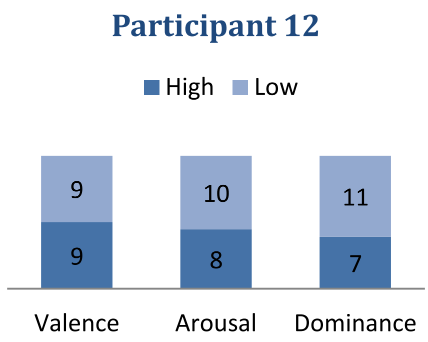

Participant 7
High
Low
13
5
11
7
13
5
Valence
Arousal
Dominance

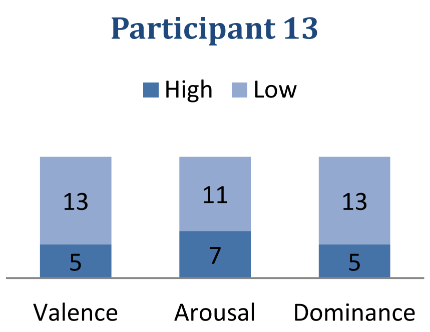

Participant 8
High
Low
12
6
13
5
14
4
Valence
Arousal
Dominance

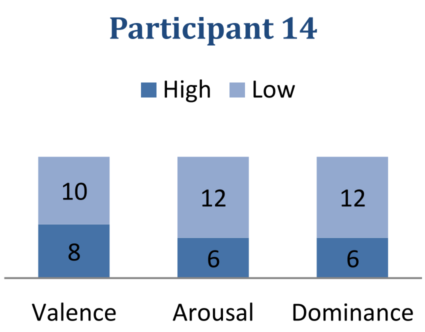

Participant 9
High
Low
10
5
6
8
13
12
Valence
Arousal
Dominance

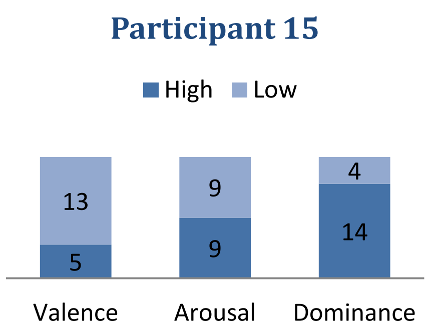

Participant 10
High
Low
6
11
10
12
7
8
Valence
Arousal
Dominance

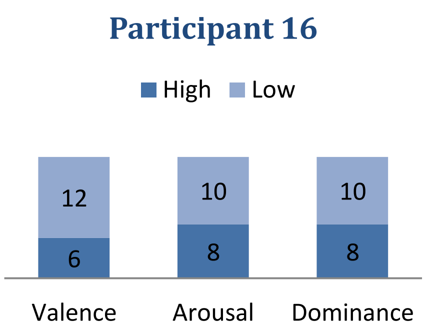

Participant 11
High
Low
12
11
7
6
7
11
Valence
Arousal
Dominance


Participant 12
High
Low
9
10
11
9
8
7
Valence
Arousal
Dominance


Participant 13
High
Low
13
11
13
5
7
5
Valence
Arousal
Dominance


Participant 14
High
Low
10
12
12
8
6
6
Valence
Arousal
Dominance


Participant 15
High
Low
13
9
4
5
9
14
Valence
Arousal
Dominance


Participant 16
High
Low
12
10
10
6
8
8
Valence
Arousal
Dominance

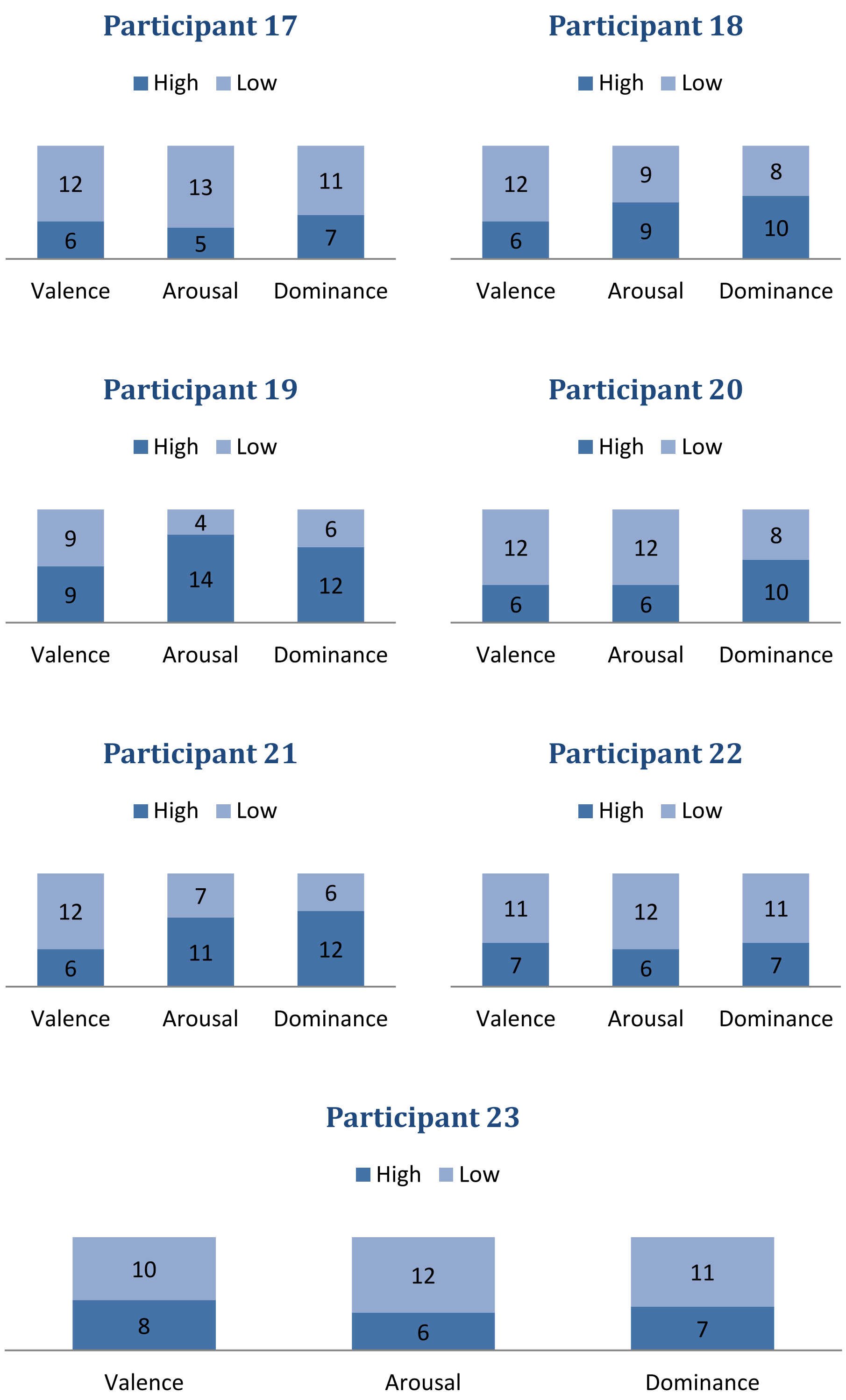


FIGURE 9 CLASS DISTRIBUTION PER PARTICIPANT

Data Analysis

The data of DREAMER is pre-processed and used to create a number of features, which are then utilized as input by the emotions classifiers constructed for this experiment. The classification results are then compared to the results of similar research.

Pre-processing

Before feeding the data to the classifier, it's necessary to pre-process the raw EEG data in order to get rid of noise and artefacts that were detected but did not originate from the brain. Signals caused from cardiac activity, eye movement and muscular activity, as well as power line noise are captured by the EEG device and thus downgrade the quality of the recorded data, making the use of denoising methods a necessity.

In EEG Signals, most ocular artefacts (eye blinking, eye movement, cardiac interferences, etc.) are dominant below 4 Hz, muscle movements produce artefacts above 30 Hz [28] and power line noise usually lies at 50 or 60 Hz, while the frequency bands that contain information relative to the affect recognition task lie in the range of 4 – 30 Hz. This frequency range is commonly divided in theta, alpha, and beta bands. Three separate band-pass Butterworth filters are applied using python module scipy.signal in order to extract only the frequencies inside the ranges of interest. Artefacts that may have been introduced at the beginning or the end of the EEG data in the form of DC offsets are handled by padding the data by a DC constant at the beginning and at the end, before re-sampling. Finally the filtered EEG data are shifted by the filter's group delay. The final step for preparing the EEG data for further analysis is the application of the Common Average Reference (CAR) method, as recommended by Cohen[29], which computes the average value over all the electrodes and subtracts it from each sample of each electrode. In that way we have analyzed the EEG Data into six different data arrays, three arrays containing data for the Alpha, Beta and Theta Bands for the baseline features. The other three arrays contain data for the Alpha, Beta and Theta Bands for the trial features.

Features Extraction

It is well documented that the power spectral densities (PSDs) of EEG signals in different bands are correlated with the affective state of a human [31]. Soleymani et al. [26] showed that the higher frequency components of EEG signals carry more important information regarding positive emotions compared to negative ones (high and low valence respectively), while a correlation between beta power and positive emotional self-induction has also been reported [32]. Koelstra et al. [30] also found strong correlations between valence and EEG signals at all frequency bands. Furthermore, their study also found negative correlations between arousal and the theta, alpha and gamma bands of EEG signals, while prior studies [33] [34] have reported an inverse relationship between the power of the alpha band and the general arousal level.

After the preprocessing step, the captured EEG signals were separated into the theta (4 Hz - 8 Hz), alpha (8 Hz – 13 Hz), and beta (13 Hz – 20 Hz) frequency bands. Welch's overlapped segment averaging estimator is then used to estimate the PSD of each EEG band, using a 256 samples window with an overlap of 128 samples. The logarithms of the PSD from each of the aforementioned bands are extracted from the signal of each of the 14 electrodes in order to be used as features, as also proposed in [22],[26],[30] , leading to a total of 42 features (3 of each of the 14 electrodes). The aforementioned calculation occurs both to the baseline features and to the trial features.

Before proceeding to generating the feature vectors, it's necessary to make the data comparable, since their magnitude and range heavily depend on particular conditions, source signal characteristics, and different subjects. The magnitude of the features extracted from the EEG varies significantly depending on the type of feature and its source. For example, when extracting features from EEG signals, the PSD for higher frequencies has a much smaller magnitude than the PSD for lower frequencies [35]. Physiological signals tend to have high variance between different subjects, as well as between the same features measured at different moments for each individual subject [36]. Feature normalization is employed in order to address these issues. After calculating the PSD values for both the Trial Features and the Baseline features, each of the trial features is divided by the corresponding baseline features. Due to the division of the extracted features by the baseline features, the derived normalized feature contains information about the relation of the initial feature to the background activity, i.e. activity that is present in the data but is not modulated by the actual affective stimuli [37]. As a result, the feature normalization method is employed to remove or strongly attenuate the background activity in order to obtain only stimuli-related changes in the EEG recordings.

Finally, all the features are concatenated into the final feature vector F_{EEG} as follows: Let $F_{i\theta}$, $F_{i\alpha}$, and $F_{i\beta}$, be the normalized logarithm of the PSD for the signal of the i-th electrode, i= 1,2,…,14, for the theta, alpha and beta bands respectively. The final feature vector is defined as $F_{EEG} = [F_{1\theta}\ \ F_{1\alpha}\ \ F_{1\beta} \ldots F_{14\theta}\ \ F_{14\alpha}\ \ F_{14\beta}\ ]$

Classification Experiments

Three classification methods were tested on the data, using two different approaches to validate the results. The classification methods compared included the Gaussian Naïve Bayes Algorithm, Support-Vector Machines and Gradient Tree Boosting using XGBoost. Classification experiments were carried out independently for each different participant. During the first approach, the data was split into training data and into test data, using a ratio of 2:1. The classification model was trained using 67% of the Data before it was tested using the rest 33% of the data. For each participant the metrics of Accuracy, Precision, Recall and F1-score were measured. For the Classification of Arousal and Dominance, it can be considered that the Class Distribution is balanced, thus Accuracy is a valid metric for concluding on the Classifier's performance. For the Classification of Valence, the class distribution is less balanced, thus the F1-score is also considered as a means of comparison. The results of the Classifier are also compared to Katsigiannis et al[23], considering all the Classification Experiments are performed using the same Dataset. The classification method which shows the best results is then validated using the second approach, following a 10-fold cross validation protocol. This allows comparing the Classifier's performance using various different training sets and thus provides a better representation of the Classifier's performance outside the contexts of this experiment.

To further improve the performance of the Classifier, attempting to perform Grid Search over the Classifier's hyper parameters for the data of each participant is attempted. The classifier is then validated again using 10-fold Cross Validation, comparing the results with the previous ones.

TABLE 2. GRID SEARCH HYPERPARAMETERS

Hyperparameter	Potential Values
'min_child_weight'	[1,5]
'gamma'	[0, 0.5, 1 , 1.5, 2]
'subsample'	[0.5, 0.6, 0.8, 1.0]
'colsample_bytree'	[0.6, 0.8, 1.0]
'max_depth'	[5,6,7]

This approach significantly increases the amount of time required to train the classifier, however, once it has been trained there is no impact on the time it is required to classify a feature vector.

After the Classifier has been tested on its performance, in the case where the classifier is being trained independently for each different participant, the experiment additionally attempts to measure the performance of the Classifier for the opposite case. Using a leave-one-out approach, all participants except one, are added as training data to the classifier. Then the classifier is tested on the data of that one participant that was left out of the training data. This procedure is repeated, cycling the test participant, until all participants have been used as both training data and testing data. It is expected, that the Classifier will fail to perform well in this experiment as the patterns which can be noticed within EEG data not only vary between different individuals, but also show differentiations in the same individual when comparing their recordings from a specific date to another.[37]

Conceptual Game Design

To utilize the output of the classifier and make further experiments on it, this document proceeds to propose a serious game design, which allows adapting the emotional state of the user into the mechanics of the game.

Primarily, it is compulsory to include features which function as indicators to the recorded emotional state. The indications can be visual, auditory or both. To indicate the recorded value of Valence a visual filter for adjusting the warmth of the visible colors can be applied, creating a colder bluish visual when inputting a low Valence score or a warmer reddish hue when a high Valence score is recorded. The saturation values of the screen can be matched to the recorded Arousal, with low Arousal leading to a lower saturation. Accordingly, Dominance can be assigned to the brightness of the game's visuals, with the scenes dimming darker when a high Dominance score is recorded. Furthermore, an overlay of visual effects can be added in order to enhance the visual impact of the previously suggested indication mechanisms. The blue tint produced when the recorded valence is low, can be supplemented with the visual effect of the corners of the screen freezing, the drop in saturation produced to indicate a low arousal can be supplemented with an overlay of RGB noise added to the edges of the player's display. The auditory features that can be included to improve these indications contain the inclusion of sound effects that match any displayed visual effects as well as tweaking the pitch of the background music, reducing it for lower Dominance values. Having the same background track recorded with different instruments and mixing their volumes based on the values of the VAD Vector is also a valid option.

Additionally, features which control the state of the game and features that promote a specific goal for the game are designed. Within the context of this document, the design of the game supports features which enable it to be used as an emotions regulation device. These features require to be tailored directly into the mechanics and the gameplay, so they will be discussed in parallel with the explanation of the game's core elements.

Core Gameplay

The player navigates their avatar through a maze, using buttons on his keyboard in order to change the direction of their movement. Their goal is to collect enough points before the time limit is reached. The points they collect within the maze affect their survival parameters, which includes stats that represent the physical, mental and financial situation of their Avatar.

Every game starts with a small maze, which is very easy and simple to clear in a small amount of time. At the end of every level, the player is allowed to make a choice between picking one of the tiles that appear on their screen or passing that opportunity. After picking a tile, they get to place it on their maze, extending its size. The expansion of the maze becomes necessary as the game progresses and the player finds themselves struggling to keep their Survival Parameters over their critical limits.

Every tile is defined by four properties. The first property includes the spatial layout of the maze paths upon the tile. The second property is affiliated with the type of the tile; an attribute that also defines which survival parameters will be affected when collecting points from this tile. The third property determines a threshold of points that needs to be collected in each level upon the tile to prevent activating a negative effect on the tile. The final property describes the negative effect which occurs when the previous threshold is not achieved.

The points that a player collects on the maze are directly tied to one or more of his Survival Parameters. When these parameters are low, they apply a number of negative effects on all the tiles of the board. In addition to that, when the player reaches the end a level, for each of those parameters below a critical point, they lose 1 Life. When their lives drop to 0, it is game over. Their Survival Parameters include the following: Money, Social, Health and Entertainment.

It is possible for the player to collect special collectibles within the maze, which increase their Skills. Some of the tiles that the player wants to place on his maze, might require having over a threshold of points collected to a specific skill before you can place them. The available skills are the following: Education, Humor, Housekeeping and Fitness.

For some tiles, when the player fails to collect enough points within a level, it is possible for the tile to spawn a Chaser during the next Level. Chasers are hostile creatures that traverse the maze, trying to catch the player. If a Chaser comes in touch with the player's avatar, then they receive a significant reduction to their Survival Parameters and the chaser is destroyed. Chasers are also generated by tiles that contain Chaser spawners. These spawners generate chasers between intervals up to their limit and when a chaser is destroyed, dropping below that limit, they spawn new ones until the limit is reached again.

By regulating their emotions and staying relaxed, the player can slow down the hostile chasers and slowly destroy them. Using such a method, the player is rewarded for managing to achieve serenity and for being able to maintain a low arousal, high valence profile.

Tile Generation

The maze is made up of multiple tiles, placed adjacent to one another. Between levels, the player is allowed to choose one of three automatically generated tiles. During the generation step of a tile, all its properties are calculated.

The spatial layout of a tile is initiated as a twenty by twenty integer array, which is filled with zeros to represent parts of the tile that the player avatar can't move through. This array defines a coordinate system for use within the tile. The first step of the algorithm generates an eighteen by eighteen rectangle, centered to the center of the tile. All tiles have this pathway, which is referred to as the **Outer Loop**. For the second step of the algorithm, the algorithm generates three more rectangles of varying sizes. For each of those rectangles the following values are calculated:

- **Width:** The size of the rectangle from its top-left corner to its top-right corner.
- **Height:** The size of the rectangle from its top-left corner to its bottom-left corner.
- **X Position:** The horizontal position of the top-left corner of the rectangle, relative to the coordinates of the Tile. The X Position can be negative, in which case it is considered to be outside of the tile's borders. The X Position's lowest value is equal to the *Width* of the Rectangle – *2.*
- **Y Position:** The vertical position of the top-left corner of the rectangle, relative to the coordinates of the tile. The Y Position can be negative, in which case it is considered to be outside of the tile's borders. The Y Position's lowest value is equal to the *Height* of the Rectangle – *2.*

Each of these Rectangles is generated separately given a different size parameter. The first rectangle to be generated is of size 2. The second is of size 1 and the third and final one of size 0. This size parameter is used to determine the ranges from which the Width and Height values will be selected from.

The first step during the generation of a rectangle is to calculate three variables. For each rectangle, these variables are regenerated:

- **bigSize:** This variable is randomly assigned an integer between 10 to 17.
- **midSize:** This variable is randomly assigned an integer between 6 and *bigSize*-1.
- **smallSize:** This variable is randomly assigned an integer between 3 and *midsize*-1.

Then depending on the Size Parameter, the Width and the Height of the Rectangle are calculated as following:

- **Size = 2:** The Width is randomly assigned either the bigSize or the midSize variable's value. The same assignment procedure is followed for the Height.
- **Size = 1:** The Width is randomly assigned any of the three variables' values. With the midSize option having double the chance of being selected than the other two. The same assignment procedure is followed for the Height.
- **Size = 0:** The Width is randomly assigned either the midSize or the smallSize variable's value. The same assignment procedure is followed for the Height.

Finally, the X and Y Positions of the Rectangle are generated as following:

- **X Position:** Is randomly assigned a value between (-Width +2, 17).
- **Y Position:** Is randomly assigned a value between (-Height +2, 17).

After the rectangles have been generated, they are used to draw paths upon the tile. Each rectangle is positioned based on the X Position and its Y Position. Any parts of the rectangle within the **Outer Loop** are used to re-assign the values of the Pathways Array. If a rectangle doesn't intersect with the Outer Loop or any of the existing pathways on the Array, then a line is generated between it and the closest element of the Array with a non-zero value. This concludes the Spatial Layout Generation of the Tile. Upon being placed on the maze, the tile generates a random path between its Outer Loop to the Outer Loop of at least one Adjacent Tile.

Then generation of the other properties of each Tile is simpler. The type of the tile is randomly picked from an enumerated list.

- **Job tile:** The points collected from this tile are converted to Money.
- **Hobby tile:** The points collected from this tile are converted to Entertainment.
- **Person tile:** The points collected from this tile are converted to Social.
- **Healthy Habit tile:** The points collected from this tile are converted to Health.
- **Survival tile:** The points collected from this tile are converted to your Lowest Survival Parameter.

The tile's daily goal, meaning the threshold of points that should be collected over the tile while playing a level, is calculated based on a percentage of the points that can be generated upon the tile within the span of the level. The percentage is randomly selected, ranging from 20% to 80%. The exact number of points that should be collected from the tile is displayed to the player instead of the percentage itself. All tiles generate points at a standard rate. By regulating their emotions, the player can increase that rate.

The final property of the tile, which is activated upon failing to complete the tile's daily goal, is randomly picked from an enumerated list. The possible Negligence Penalties are listed below:

- The tile will contain a Chaser spawner starting from the next level. The Chaser spawner has an upper limit of 1 Chaser. If this negligence penalty is applied on this tile again, the spawner's upper limit is increased by 1.
- The tile's corruption value is increased by 1. For every 1 point in this value, the chance of a Negative point being generated instead of a normal point increases. The first point increases that chance to 20%. The second point increases it further to 30%. The third point increases it to 35%. The fourth to 37% and with every point after that it increases by another 1%.
- The tile is destroyed and removed from the maze.
- All your Survival Parameters are reduced by 25%.

The Negligence Penalty of each tile changes after every time that it is applied.

Maze Generation

The first level of the game begins with one Survival Tile. After that level is completed, the player is given the option to choose between one of three tiles. They can also choose to skip this phase, not expanding their maze. If they pick one of the tiles however, the player should then proceed and place the tile on the maze, adjacent to one of the already existing tiles. Before they begin playing the next level, they can change the positions of all the tiles as long as their position remains legal. The position of the starting tile cannot change.

When a new level starts, the game generates paths connecting adjacent tiles with one another. The Starting tile is considered the root of a tree, while the other tiles are converted into nodes of that tree. A path is necessarily generated between parents and their children nodes. After this step is completed, the tree is iterated in reverse and it is randomly determined whether to add more connections between adjacent tiles.

After the Maze's layout is generated, the game spawns the initial points within the maze. The points generated within each tile are equivalent to 20% of the number of points that the maze will have generated by the time that the time limit is reached.

The Levels

After the Maze has been generated successfully, the player is loaded within the level and the game begins. The player rushes to collect as many points as possible, making sure to complete the goal of each tile along the way. When the time limit is reached an exit spawns to a random corner of the maze. The player should get to the exit as fast as possible, as the tiles start generating more negative points and also this delay would carry over to the next level, either by reducing the time limit or by reducing the speed of the player's avatar for the duration of that delay during the beginning of the next level.

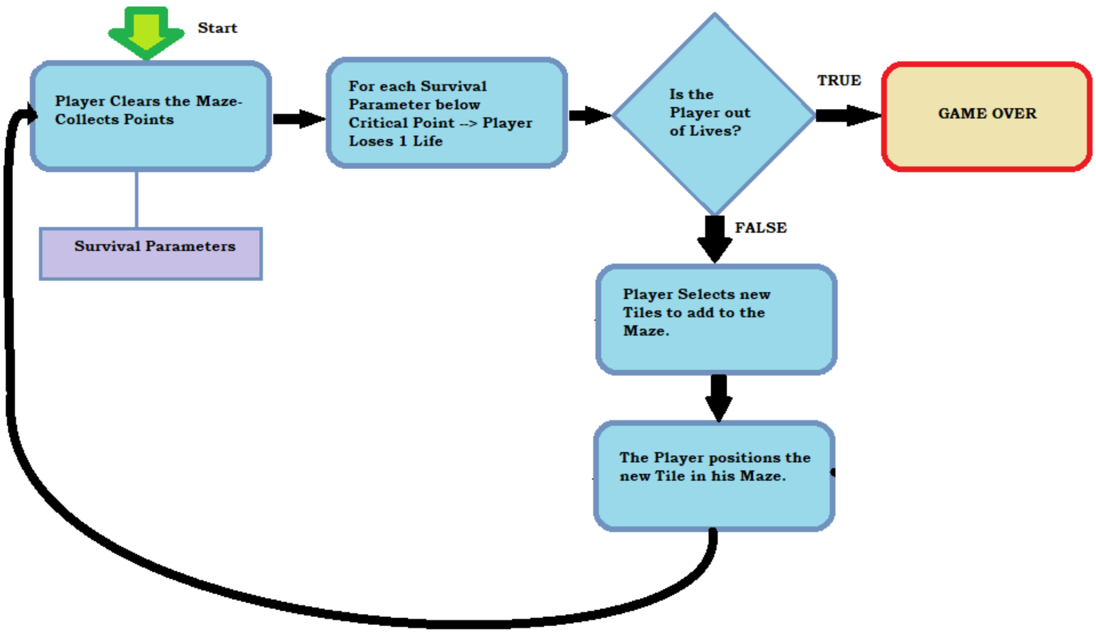


Figure 10 - Core Gameplay Loop

ALLEGORIES

The game is designed to allegorize the daily struggle of a human living within an urban society. Each level represents a day of the player avatar's life, with each tile representing the different aspects in that individual's life. Gathering points from within a certain tile, represents giving effort to that particular aspect of life, with the rewards being relative to the type of the tile. As such, Job tiles provide the player's avatar with Money; his Hobby tiles provide the player's avatar with Entertainment; while Person tiles provide him with ways to socialize.

The daily goal of each tile represents an expectation for the Player's Avatar. Failing to meet these expectations burdens the player's avatar. The presence of negative points within the tile signifies how this part of the player avatar's life has become more stressful and less rewarding. The presence of chasers, represents the fears of messing up.

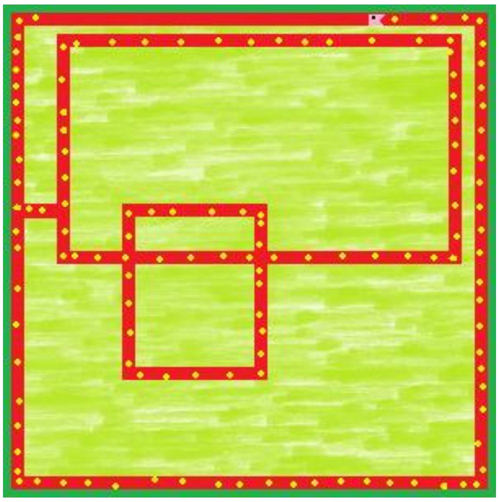


FIGURE 11 - EXAMPLE OF A TILE

FIGURE 12 - EXAMPLE OF A MAZE

Results

Three different classification methods are compared: GNB, SVM and XGBoost. For this comparison, each of them is validated using the simple split approach, by randomly splitting the data in training data and testing data. 67% of the data is used as training data with the remaining 33% of the data of each participant being used to test the classifier. The results of this comparison are displayed in Table 3.

Table 3. Accuracy and F1 Scores for the Examined Classification Methods

	Accuracy			F1-Score		
	Valence	Arousal	Dominance	Valence	Arousal	Dominance
Katsigiannis et al. [23] (SVM)	0.6249	0.6217	0.6184	0.5184	0.5767	0.6166
GNB	0.6023	0.6204	0.6458	0.5434	0.5788	0.6200
SVM	0.6838	0.6603	0.6639	0.5898	0.5732	0.6145
XGBoost	**0.8514**	**0.8541**	**0.8269**	**0.8027**	**0.8138**	**0.7946**

The Gradient Tree Boosting using XGBoost Method is then validated using a 10-fold cross-validation approach, in order to ensure its performance using various sets of train data. Following this approach, an accuracy of **88.12%** was measured for the classification task of the Valence attribute. An accuracy of **88.02%** and an accuracy of **87.71%** were noted for the classification tasks of Arousal and Dominance accordingly. This level of accuracy shows that the Classifier could be reproduced by training random sets of data, without a significant alteration to the performance of the classifier.

Table 4. Valence 10-Fold Validation Metrics

Participant	Accuracy%	Precision%	Recall%	F1-Score%	ROC AUC%	Class Distribution%
1	79.29	78.44	80.71	78.20	87.00	50.00
2	96.52	94.67	95.00	93.56	99.50	27.78
3	86.05	83.11	87.14	83.67	97.14	44.44
4	89.00	83.33	59.17	68.33	91.88	22.22
5	81.81	82.31	61.69	67.36	91.19	33.33
6	91.52	92.88	92.18	92.04	97.88	44.44
7	91.62	88.17	84.33	84.15	97.18	27.78
8	91.67	91.33	84.07	87.30	97.01	33.33
9	90.24	89.49	89.55	88.44	96.41	44.44
10	90.29	90.72	94.71	92.07	94.16	66.66
11	90.86	85.89	87.50	85.97	97.03	33.33
12	84.67	87.12	85.34	84.87	91.31	50.00
13	86.76	87.35	91.44	87.42	95.01	55.55
14	81.19	81.83	77.91	77.55	90.64	44.44
15	88.10	95.00	62.67	73.92	93.29	27.78

16	92.33	87.50	87.40	86.26	97.22	33.33
17	86.10	83.00	65.76	72.00	91.63	33.33
18	87.48	85.98	79.64	82.16	95.24	33.33
19	98.62	100.00	97.08	98.42	100.00	50.00
20	76.86	67.69	68.05	64.86	86.82	33.33
21	90.90	93.06	82.00	84.86	96.63	33.33
22	84.00	82.40	79.19	78.99	91.28	38.88
23	90.90	93.00	85.56	88.55	97.74	44.44
Average	**88.12**	**87.14**	**81.65**	**82.65**	**94.48**	**39.37**

Table 5. Arousal 10-Fold Validation Metrics

Participant	Accuracy%	Precision%	Recall%	F1-Score%	ROC AUC%	Class Distribution%
1	82.67	82.50	57.00	62.60	92.32	27.78
2	85.38	88.89	81.30	84.14	93.36	50.00
3	91.71	90.50	89.49	89.29	96.38	38.88
4	88.90	87.14	79.12	80.09	93.86	27.78
5	80.48	82.13	72.95	75.92	89.25	44.44
6	79.81	82.00	91.64	85.74	84.39	66.66
7	90.29	95.00	84.92	87.99	99.08	38.88
8	93.71	92.50	83.50	85.37	99.33	27.78
9	86.00	86.14	96.13	90.48	93.47	72.22
10	93.71	95.32	89.29	91.05	97.64	38.88
11	83.95	85.71	75.21	78.30	95.95	38.88
12	86.00	84.65	81.28	82.22	94.61	38.88
13	83.90	80.48	84.64	82.23	91.32	44.44
14	90.38	81.00	75.14	77.22	95.32	33.33
15	91.67	91.87	91.77	91.55	96.46	50.00
16	88.81	89.33	85.83	86.56	94.87	44.44
17	86.81	81.55	71.55	73.62	93.65	27.78
18	81.38	79.06	85.48	80.82	88.73	50.00
19	92.33	92.28	98.23	94.94	97.97	77.78
20	90.86	89.64	85.92	85.88	96.51	33.33
21	88.95	86.60	96.41	90.73	97.09	61.11
22	93.76	97.50	81.56	87.65	96.52	33.33
23	93.19	91.21	90.74	90.50	98.54	33.33
Average	**88.02**	**87.52**	**83.87**	**84.12**	**94.63**	**43.71**

Table 6. Dominance 10-Fold Validation Metrics

Participant	Accuracy%	Precision%	Recall%	F1-Score%	ROC AUC%	Class Distribution%
1	79.86	83.37	81.57	80.87	89.18	55.55
2	92.33	93.43	91.92	92.22	97.21	50.00
3	90.19	95.00	73.45	80.90	95.64	27.78

4	85.38	72.50	67.75	65.07	88.17	27.78
5	83.33	84.48	73.49	76.94	93.62	38.88
6	87.38	86.38	90.98	88.72	92.91	55.55
7	82.62	81.92	80.06	80.21	89.87	44.44
8	89.67	81.83	73.50	71.57	96.66	22.22
9	81.81	81.81	95.54	87.01	91.32	66.67
10	93.00	94.40	90.95	92.26	97.19	44.44
11	90.76	93.75	90.84	91.98	97.43	61.11
12	84.19	84.24	75.27	79.08	90.42	38.89
13	84.57	85.20	87.88	85.84	89.49	55.55
14	84.67	81.25	66.33	70.92	89.24	33.33
15	95.10	94.59	99.09	96.59	98.87	77.78
16	87.38	86.03	85.49	84.68	95.03	44.44
17	89.52	90.00	84.21	85.70	96.03	38.88
18	86.86	86.54	89.63	87.26	96.74	55.55
19	92.43	91.09	99.00	94.62	96.38	66.67
20	91.10	94.46	89.92	91.63	96.84	55.55
21	88.24	88.39	94.28	91.04	95.95	66.67
22	88.86	84.87	89.11	86.55	96.51	38.88
23	88.19	89.64	81.08	84.02	97.52	38.88
Average	**87.71**	**87.29**	**84.84**	**84.59**	**94.27**	**48.06**

The accuracy metrics were calculated as the mean of accuracies of each separate participant. Examining the classification results for each independent participant can provide a better insight to the overall performance of the Classifier.

The comparison between the results of the participants with the highest accuracies, to the results of the participants with the lowest accuracies for each variable, shows that the Classifier provides its best results in both balanced and unbalanced class distributions. This proves that the classifier’s performance is not biased to the class distribution of the data. The worst results provide a clear insight to the Classifier’s worst performance.

To further review these observations, the results that were validated using a cross validation approach are used to generate box-plot diagrams, allowing perceiving the distribution of the results. The box-plots were generated for the metrics of: Accuracy, Precision, Recall, F1-Score and ROC Area under the Curve.

VALENCE

ACCURACY

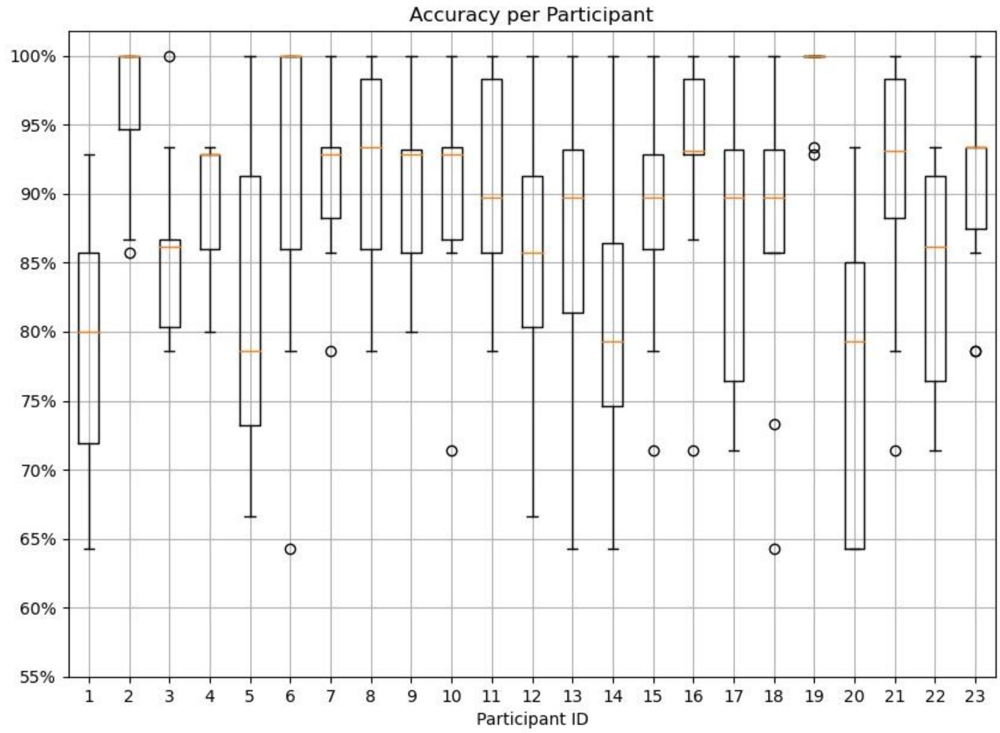


FIGURE 13 - VALENCE ACCURACY BOX PLOT

PRECISION

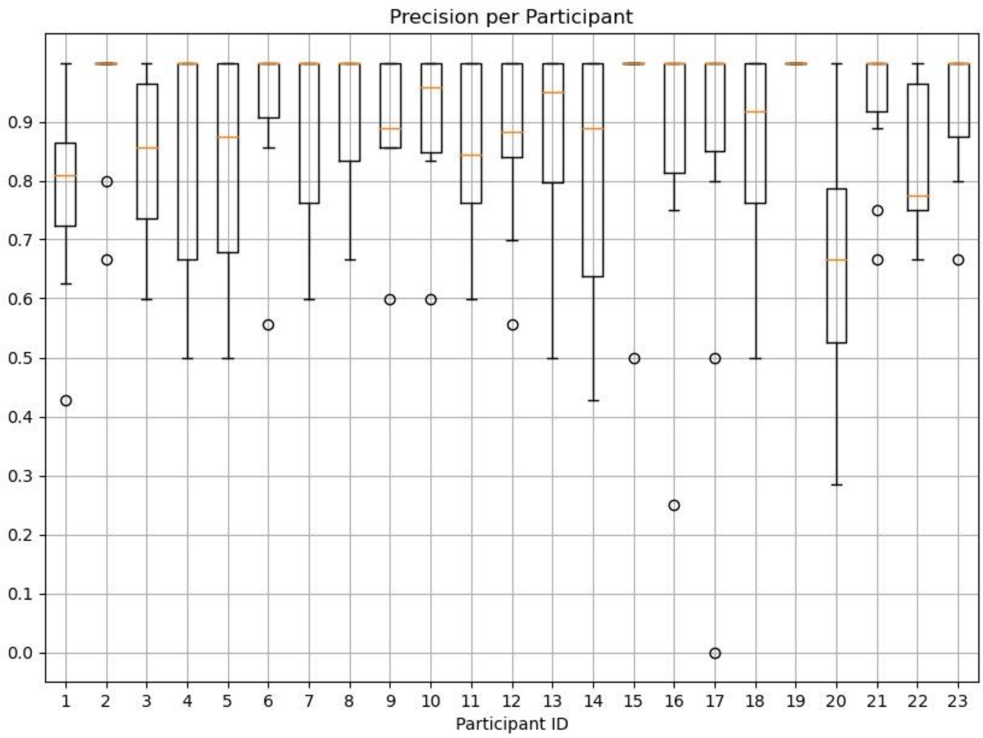


FIGURE 14 - VALENCE PRECISION BOX PLOT

RECALL

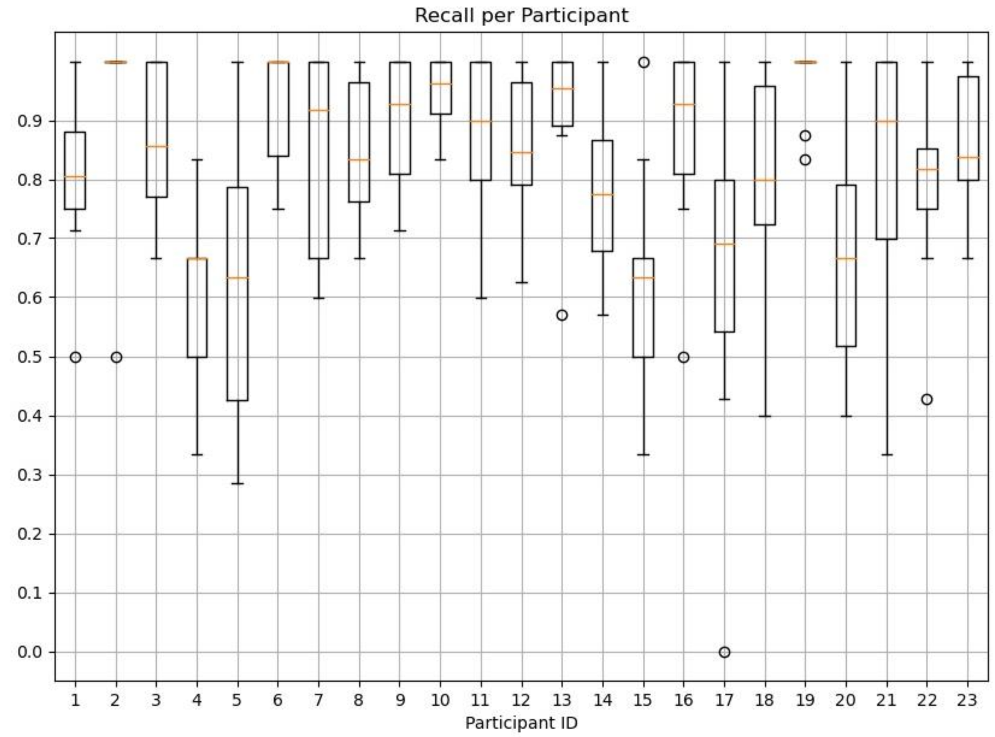


FIGURE 15 - VALENCE RECALL BOX PLOT

F1-SCORE

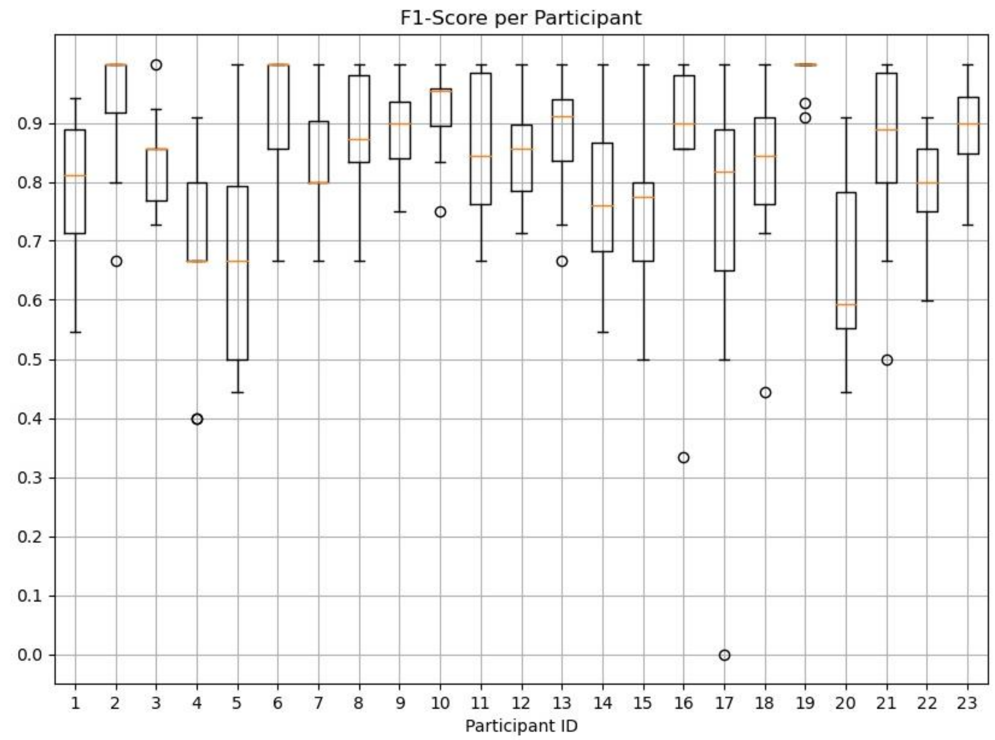


FIGURE 16 - VALENCE F1 SCORE BOX PLOT

ROC- Area Under the Curve

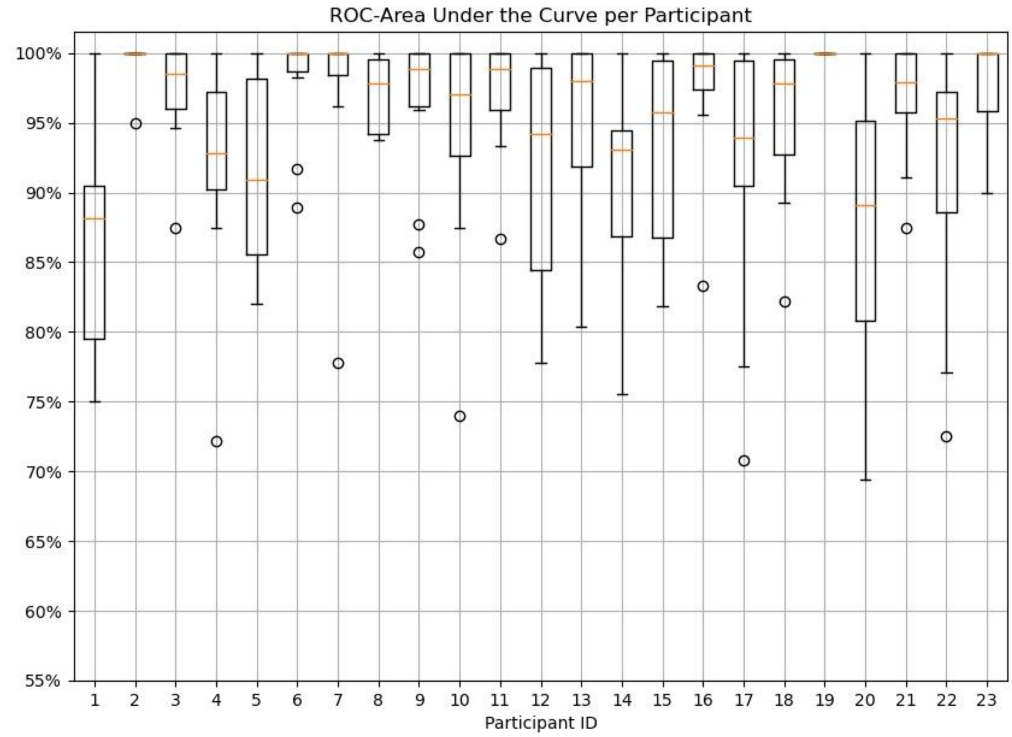


Figure 17 - Valence ROC AUC Box Plot

Arousal

Accuracy

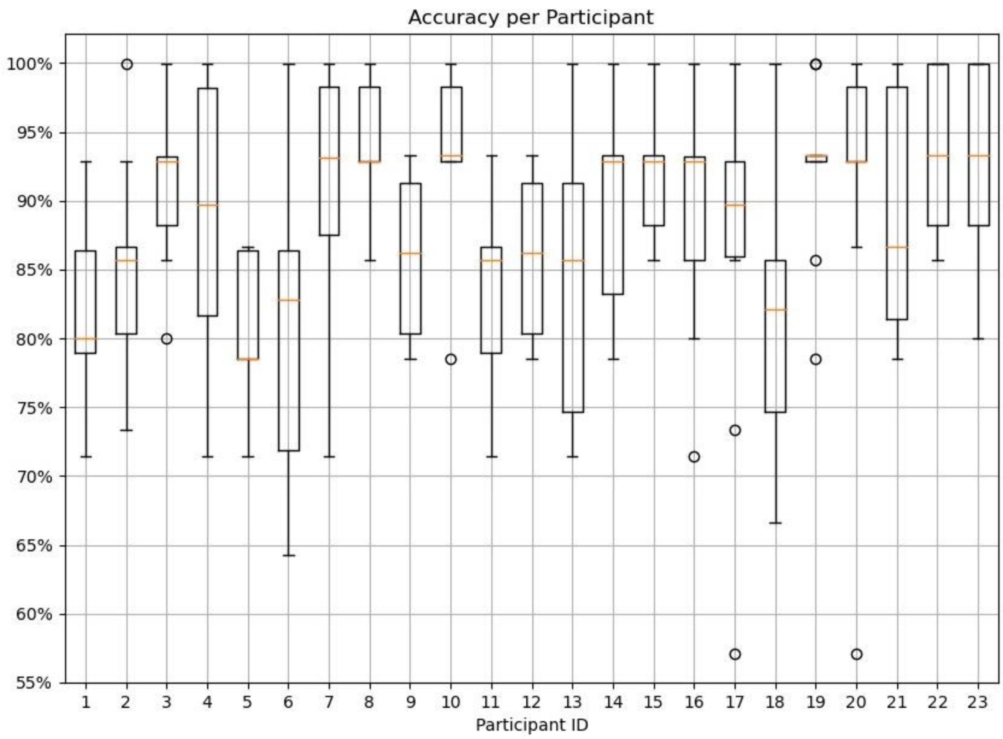


Figure 18 - Arousal Accuracy Box Plot

Precision

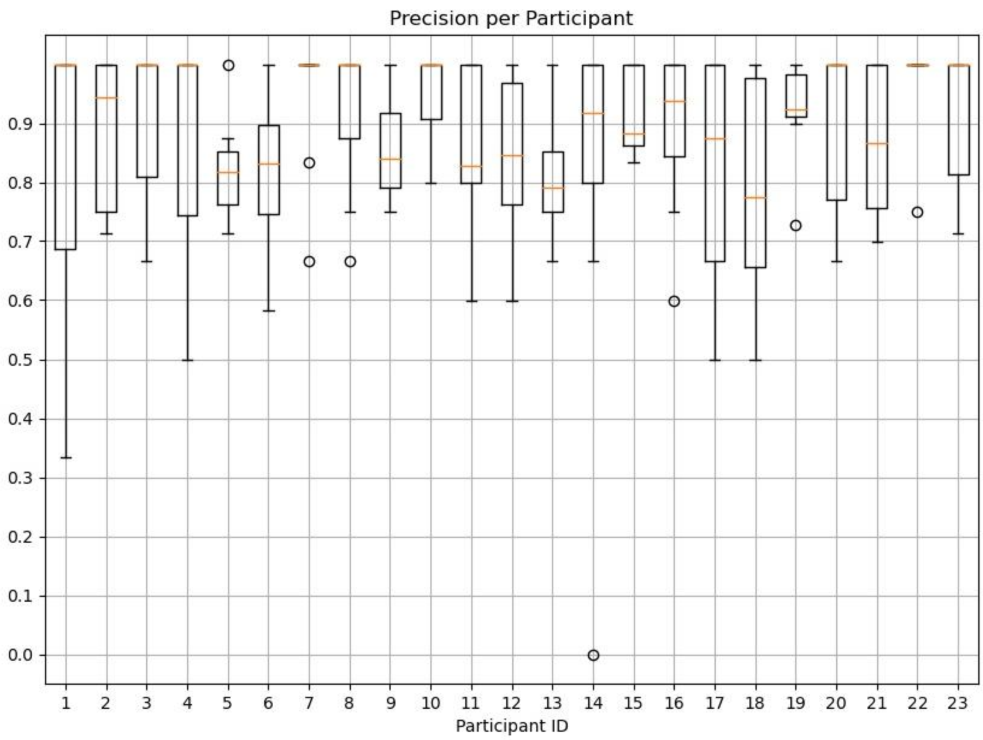


Figure 19 - Arousal Precision Box Plot

RECALL

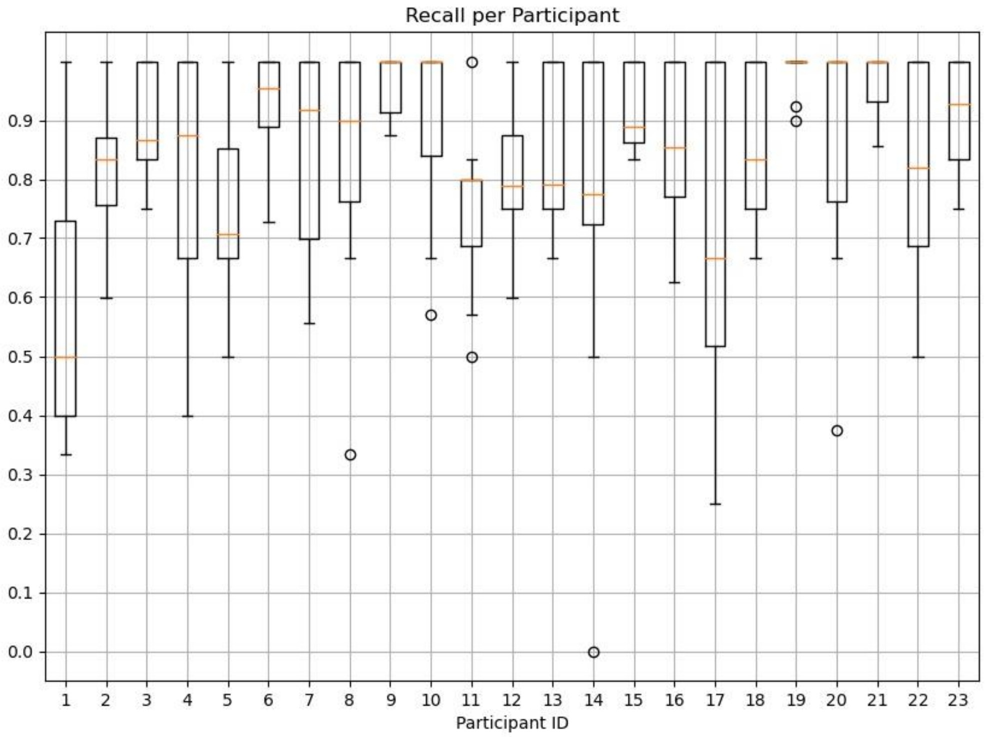


FIGURE 14 - AROUSAL RECALL BOX PLOT

F1-SCORE

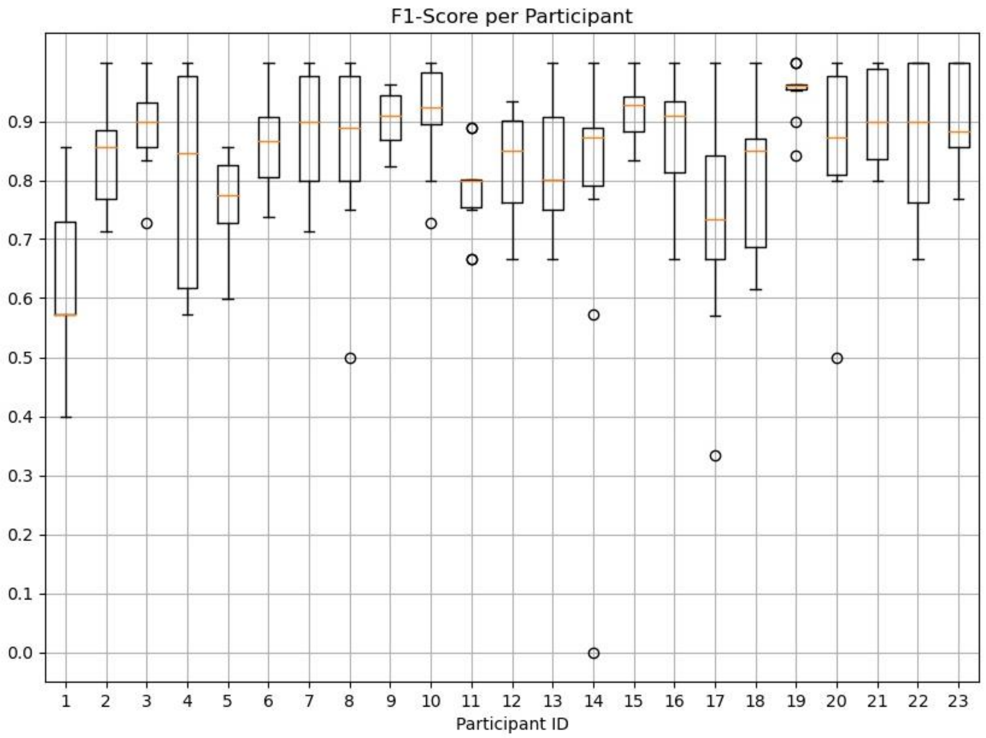


FIGURE 21 - AROUSAL F1 SCORE BOX PLOT

ROC Area Under the Curve

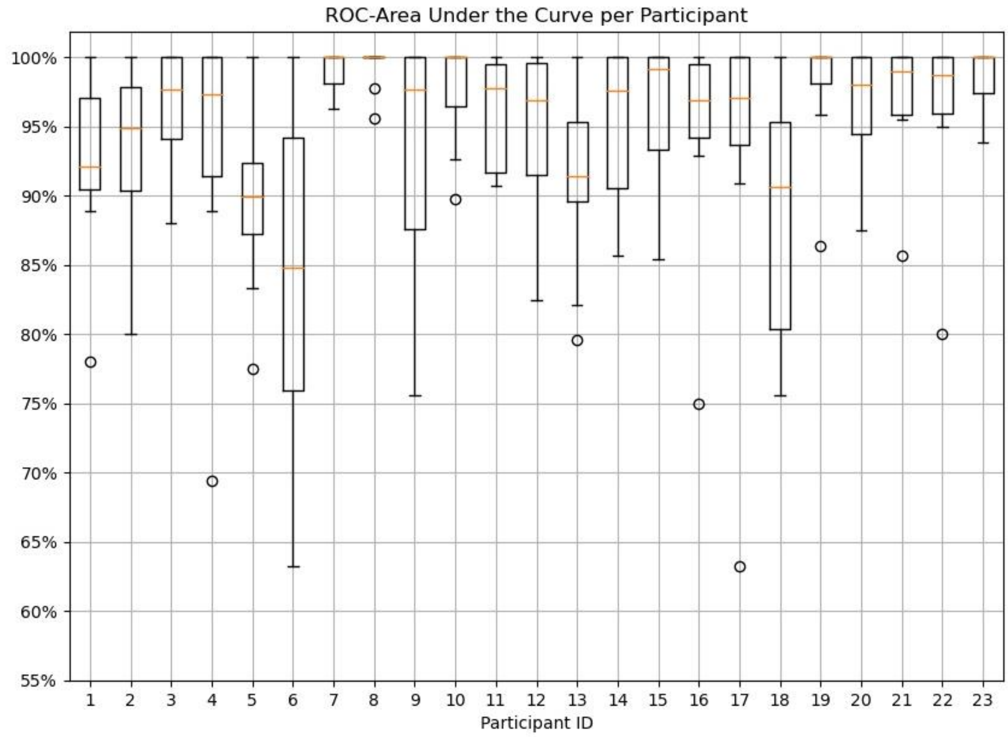


Figure 22 - Arousal ROC AUC Box Plot

Dominance

Accuracy

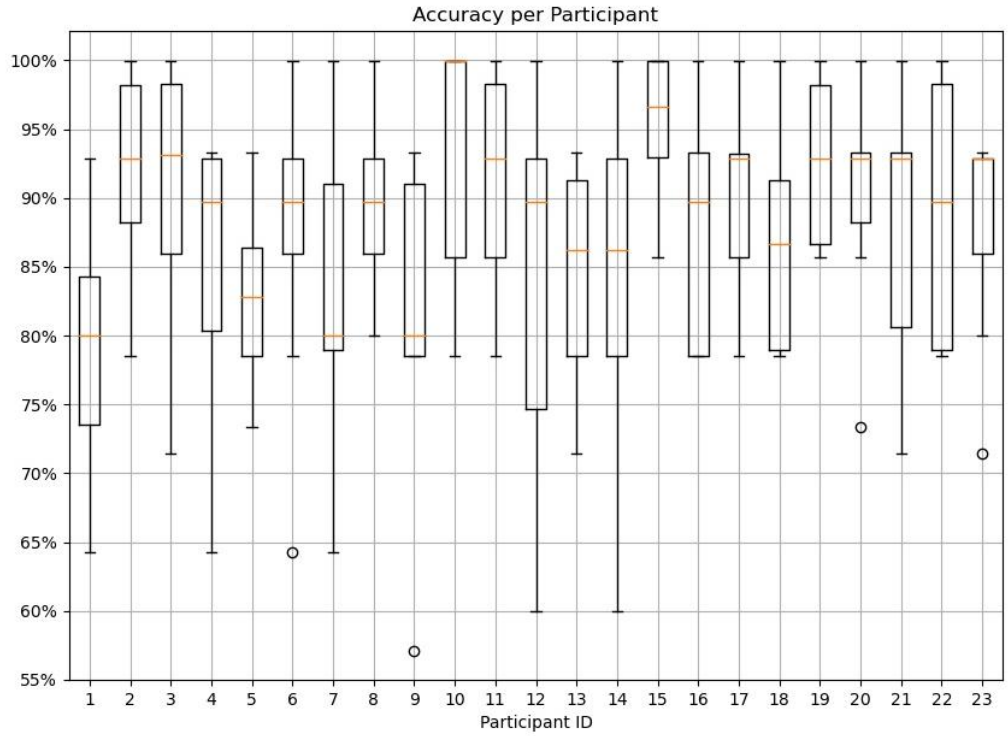


Figure 23 - Dominance Accuracy Box Plot

Precision

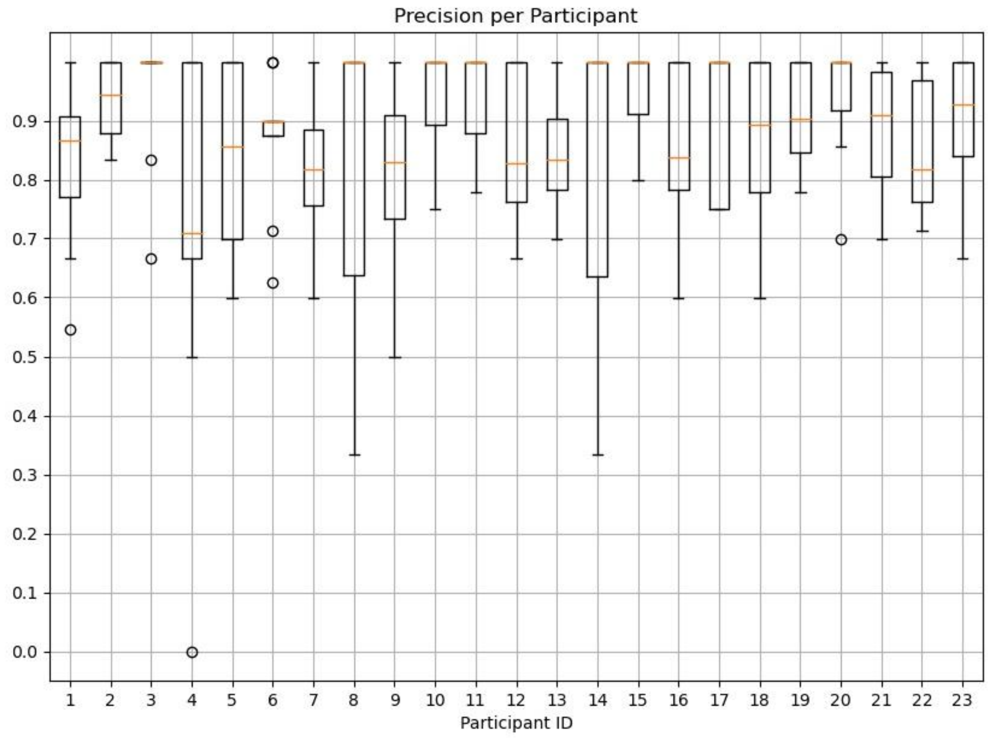


Figure 24- Dominance Precision Box Plot

Recall

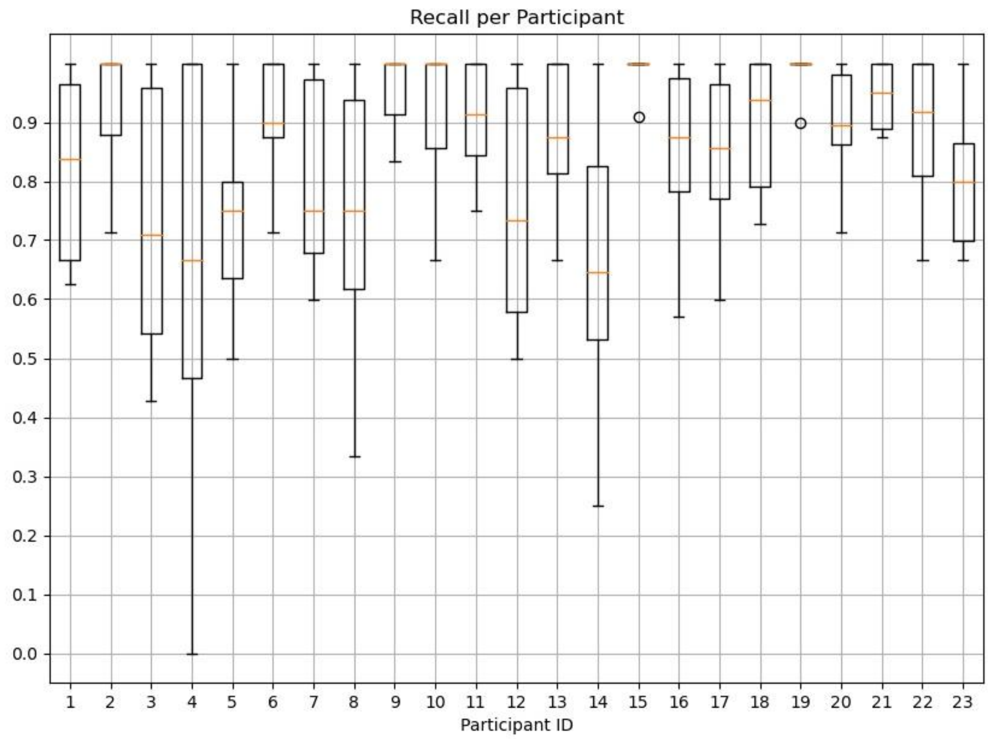


Figure 25 - Dominance Recall Box Plot

F1 Score

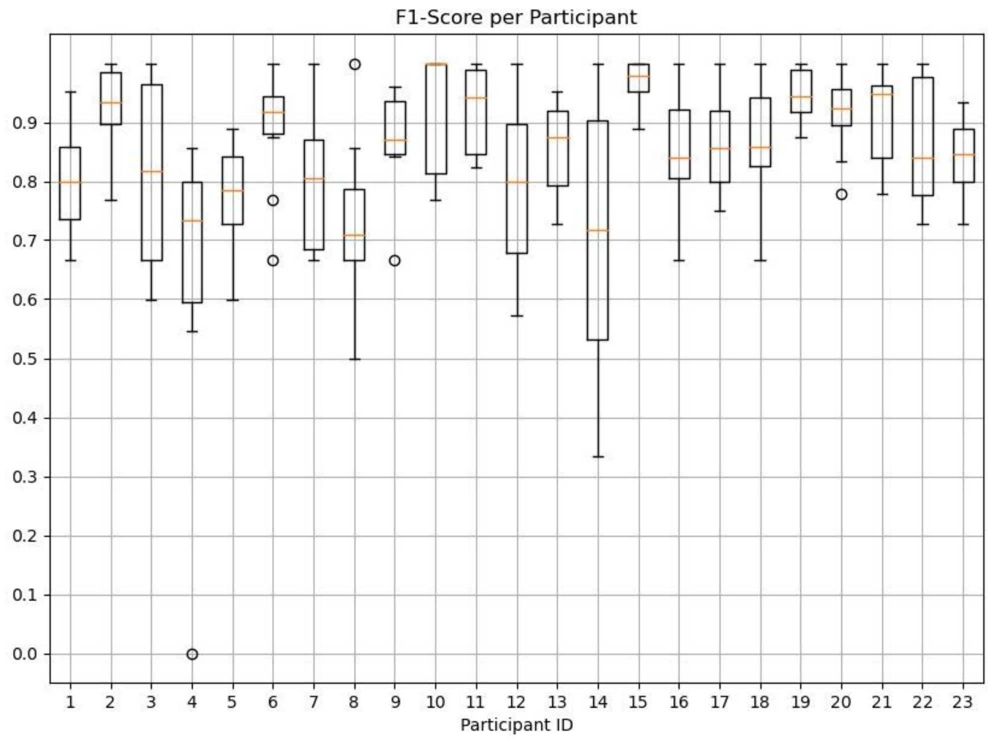


Figure 26- Dominance F1 Score Box Plot

ROC Area Under the Curve

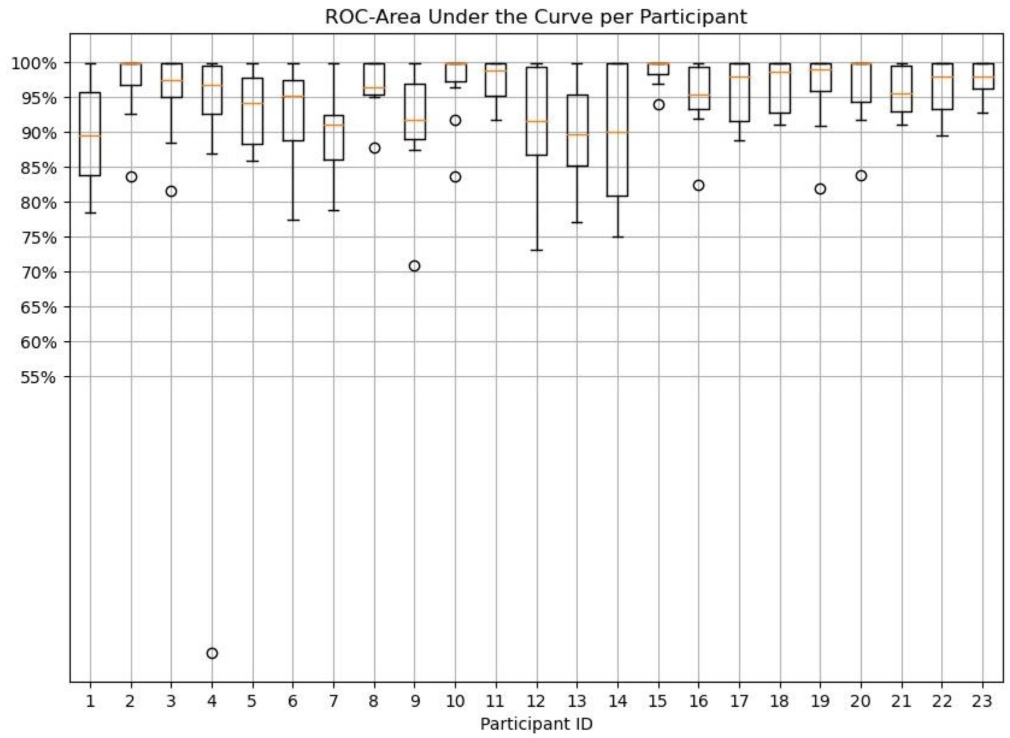


FIGURE 15 - DOMINANCE ROC AUC BOX PLOT

Observing that the Classifier is performing well, an additional step is implemented into the Classifiers logic, significantly increasing the training time in order to calculate a set of hyperparameters for the Classifier. The results of this Grid Search operation are displayed below:

TABLE 7. GRID SEARCH RESULTS COMPARISON (VALENCE)

Participant	Previous Accuracy	Accuracy w/ GridSearch	Colsample_bytree	Gamma	Max_depth	Min_child_weight	subsample
1	79.29	72.38	0.6	1	5	1	1.0
2	96.52	95.19	1.0	0.5	5	1	0.5
3	86.05	88.19	0.6	1	5	1	0.5
4	89.00	91.05	0.8	0	5	1	0.6
5	81.81	81.95	0.8	0	5	1	0.8
6	91.52	90.24	1.0	0.5	5	1	0.5
7	91.62	93.78	1.0	0	5	1	0.5
8	91.67	92.33	0.6	0	5	1	0.8
9	90.24	90.86	0.6	0.5	5	1	0.8
10	90.29	90.29	1.0	0	5	1	0.6
11	90.86	91.57	1.0	0	5	1	0.8
12	84.67	84.76	0.8	0.5	5	1	1.0
13	86.76	88.14	0.6	1.5	5	1	0.8
14	81.19	74.24	1.0	2	5	5	0.8
15	88.10	89.62	0.6	0.5	5	1	1.0
16	92.33	92.33	1.0	0	5	1	1.0
17	86.10	88.14	1.0	0	5	1	1.0
18	87.48	84.00	1.0	1	5	1	1.0

19	98.62	96.57	0.6	0	5	1	0.5
20	76.86	80.52	0.6	0	5	1	0.5
21	90.90	93.00	1.0	0.5	5	1	1.0
22	84.00	87.52	0.8	0	5	1	0.6
23	90.90	88.14	0.6	0	5	1	0.6
Mean	**88.12**	**88.03**					

TABLE 8. GRID SEARCH RESULTS COMPARISON (AROUSAL)

Participant	Previous Accuracy	Accuracy w/ GridSearch	Colsample_bytree	Gamma	Max_depth	Min_child_weight	subsample
1	82.67	82.67	0.6	0.5	5	1	0.5
2	85.38	83.29	1.0	0.5	5	1	0.8
3	91.71	87.43	0.6	0	5	1	0.8
4	88.90	86.14	0.6	0	5	1	0.8
5	80.48	80.57	0.6	2	5	1	0.5
6	79.81	79.10	0.6	0.5	6	1	1.0
7	90.29	88.71	0.6	0	5	1	0.6
8	93.71	92.38	0.6	1	5	1	1.0
9	86.00	86.71	1.0	0.5	5	1	0.5
10	93.71	93.71	1.0	0	5	1	1.0
11	83.95	88.10	1.0	1	5	1	0.8
12	86.00	84.67	0.8	0.5	5	1	1.0
13	83.90	79.10	1.0	1	5	1	0.5
14	90.38	92.39	0.6	0	5	1	0.8
15	91.67	91.62	1.0	2	5	1	1.0
16	88.81	88.14	1.0	0	5	1	1.0
17	86.81	90.29	0.6	0	5	1	0.5
18	81.38	84.67	0.8	0.5	5	1	0.5
19	92.33	95.10	0.6	0	5	1	0.8
20	90.86	88.71	0.8	1.5	5	1	0.8
21	88.95	90.33	1.0	0	5	1	0.8
22	93.76	92.38	0.8	0.5	5	1	1.0
23	93.19	92.48	1.0	0	5	1	0.8
Mean	**88.02**	**87.76**					

TABLE 9. GRID SEARCH RESULTS COMPARISON (DOMINANCE)

Participant	Previous Accuracy	Accuracy w/ GridSearch	Colsample_bytree	Gamma	Max_depth	Min_child_weight	subsample
1	79.86	78.48	0.8	0.5	5	1	0.8
2	92.33	88.86	0.6	0	5	1	0.8
3	90.19	90.90	0.8	0	5	1	1.0
4	85.38	83.19	0.6	1.5	5	1	0.5
5	83.33	84.05	0.6	1	5	1	0.5
6	87.38	84.76	0.6	2	5	1	0.8
7	82.62	87.38	0.8	0	5	1	0.5
8	89.67	88.86	0.6	1.5	5	1	0.8
9	81.81	83.19	0.6	0.5	5	1	0.5
10	93.00	93.00	0.8	0.5	5	1	0.6
11	90.76	90.14	0.6	0	5	1	0.6
12	84.19	82.00	1.0	0	5	1	0.6
13	84.57	79.10	1.0	1.5	5	1	0.8

14	84.67	86.86	0.6	0	5	1	0.6
15	95.10	94.33	0.6	0	5	1	0.8
16	87.38	91.62	1.0	0	5	1	0.8
17	89.52	86.71	0.8	1	5	1	0.5
18	86.86	85.33	0.6	0.5	5	1	0.5
19	92.43	90.38	0.6	0.5	5	1	0.8
20	91.10	89.76	0.8	2	5	1	1.0
21	88.24	90.19	0.6	0.5	5	1	0.6
22	88.86	90.29	0.6	0	5	1	0.8
23	88.19	90.90	0.6	0	5	1	0.8
Mean	**87.71**	**87.40**					

Finally, it was attempted to measure the performance of the Classifier as non-participant specific tool. The Classifier was observed 23 times, each of them with 22 Participants used as Training Set and 1 Participant used as a Testing Set. The Participant used for the Testing Set was cycled among the 23 Participants, choosing a different Participant for each of the 23 iterations of this experiment.

TABLE 10. NON-PARTICIPANT SPECIFIC PERFORMANCE (VALENCE)

Participant Left out for the Testing Set	Classifier's Accuracy%
1	54.17
2	64.58
3	54.86
4	70.83
5	63.19
6	47.92
7	64.58
8	47.22
9	50.00
10	36.11
11	52.78
12	51.39
13	45.83
14	59.03
15	68.06
16	55.56
17	59.72
18	59.03
19	58.33
20	57.64
21	63.89
22	59.03
23	50.00
MEAN	**56.25**

Table 11. Non-Participant Specific Performance (Arousal)

Participant Left out for the Testing Set	Classifier's Accuracy%
1	62.50
2	59.03
3	51.39
4	56.25
5	50.69
6	43.06
7	49.31
8	47.22
9	32.64
10	50.69
11	47.92
12	43.75
13	52.78
14	54.86
15	41.67
16	60.42
17	61.81
18	43.75
19	32.64
20	58.33
21	49.31
22	54.17
23	64.58
MEAN	**50.81**

Table 12. Non-Participant Specific Performance (Dominance)

Participant Left out for the Testing Set	Classifier's Accuracy%
1	40.97
2	42.36
3	54.86
4	57.64
5	50.69
6	56.94
7	39.58
8	60.42
9	47.22
10	56.94
11	47.92
12	45.83
13	50.00
14	52.78
15	59.03
16	59.03
17	58.33
18	42.36
19	46.53
20	63.89
21	46.53
22	47.22
23	48.61
MEAN	**51.11**

Discussion

As shown in Tables 3, 4, 5 and 6, the utilization of XGBoost significantly improves the Accuracy Performance of the Valence Classifier by 25.64% , of the Arousal Classifier by 25.85% and of the Dominance Classifier by 25.87%, when compared to the baseline research as it was conducted by Katsigiannis et Al[23] using the same Database (DREAMER). The Accuracy and ROC-AUC scores in Tables 4,5 and 6, confirm the high quality of the Classifier, something which signifies that the output of the Classifier can be used for the suggested application in the development of a Serious Game. With an average accuracy below 95%, it might prove efficient to temporarily store the most recent results of the Classifier in a buffer. Then upon small fixed intervals, clear the buffer and import its contents to a simple Voting Function which determines the final classes, before they are employed into the Serious Game.

Even though GridSearching through a set of values for the Hyperparameters of the Classifier, leads to improved results for some participants, it seems that the set of values which are searched is not extensive enough to provide the best results for the majority of participants nor does it lead to increase to the mean performance of the Classifier. A different range of values for each Hyperparameter might lead to better results, however increasing the number of possible values for those Hyperparameters would lead to increasing this operation significantly. For the participants who show an increase to their performance in these measurements, the average increase to that performance is equal to **1.83%** with the biggest deviation being equal to 4.76%. Therefore, it can be presumed that by optimizing this GridSearch operation the increase of the Classifier's performance would be within the range of 1.5% to 5%.

As shown in Tables 10, 11 and 12, when the Classifier is trained for more than one participant it operates poorly. This outcome was expected as the EEG Recordings of each individual contain multiple person-specific quirks and patterns, which further out perplexes the Classifier's training. This has been an issue with multiple EEG-based technologies, which these applications to require personalized configuration for each user. Zanesco et Al's research[37] provides hopeful insight to the solution of this problem, with their findings on EEG Microstates. Instead of training the Classifier for each different target user, it might be possible to train a Classifier for each different Microstate and then introduce a new classifier to determine the Microstate for each recorded EEG Feature Vector. Analyzing this assumption could be an interesting step to the future of this research.

The classifier was only tested on data from the DREAMER database. In future studies, it would be required to test the classifier with more data, as well as with signals acquired through a brain computer interface in real time. For the latter, it would be required to devise a routine that would allow collecting the training data required for the configuration of the Classifier to a new user. To implement the classifier in a serious game, this operation is compulsory.

It is possible to modify the Classifier’s performance by constructing a different features vector during the feature extraction step, or positioning the electrodes to different positions. For the current Feature Vectors, the frontal lobe asymmetry has been overlooked. Experimenting with different feature vectors that make use of such details might lead to the better performance of the Classifier.

A valid step forward to the future of this research would be the development and implementation of the proposed game design concept, something which would allow testing the performance of the classifier in a virtual environment. Further expansions to the game design, should include a method for training the classifier in the form of a calibration session before the beginning of the game. Short films can be employed for the completion of this task, assimilating the method used for the construction of the DREAMER database. In addition to that, the game can be improved by adding minor story elements to further stimulate the player and his emotions. The story elements could make use of the existing allegories to create this supplementary immersive experience. The story elements could be implemented by generating characters bound on each tile and adding dialogue and choices.

The Classifier stores a trained model, generated based on the users EEG recordings and the Feature Vectors generated from them. Anyone with access to that model and the Classifier, could potentially use them in order to classify the emotions of the user, considering he has access to his EEG Data. This application doesn’t store the user’s EEG Data, so the substantiality of privacy and security concern regarding the user’s data and the exploitation of the user by third parties depends on the accessibility of the user’s EEG Data to those third parties. Furthermore, the Classifier stores the model locally, in the user’s system meaning that as long as the user doesn’t share the trained model to third parties, or that his system isn’t accessed by a malicious user, the data is safe. Encrypting the model would provide an additional layer of security to the user’s data, however in that case it would be required to add the appropriate decryption algorithms in the classifier.

A moral concern is created when it comes to determining whether such technology should be used in multi-user environments, such as Multiplayer games. Even if the emotions of the user are not exposed to other users and they are employed simply to facilitate operations of the application private to the user, it is still possible for a large part of the player’s activity to be recorded in the logs of the third party responsible for managing the aforementioned environment. Due to the existence of General Data Protection Regulation (GDPR), such activity is prevented for genetic, biometric and health data, as well as personal data revealing racial and ethnic origin, political opinions, religious or ideological convictions or trade union membership. Considering emotions to be part of health data, they are also protected by GDPR. Sadly, laws and regulations won't prevent malicious third-parties to engage in such behaviors.

Conclusion

In this work a Classifier for Emotions recognition using XGBoost is proposed, which matches the emotional state of a target user to a triplet of classes. The Classifier operates with an accuracy of 88.12%, 88.02% and 87.71% for the binary classes of Valence, Arousal and Dominance, respectively. This performance is superior to the results of previously reported research and brings the Classifier to an acceptable level for its utilization as a tool in future research or in the development of Serious games for health. To showcase the possible usage of Emotions Recognition within Serious games, a simple game was designed and its mechanics were analyzed.

BIBLIOGRAPHY

[1] J. Tao, T. Tan, and R. W. Picard, Affective computing and intelligent interaction : first international conference, ACII 2005, Beijing, China, October 22-24, 2005 : proceedings. 2005.

[2] A. Mehrabian and J. A. Russell, “Evidence for a Three-Factor Theory of Emotions,” *J. Res. Pers.*, vol. 11, pp. 273–294, 1977.

[3] J. J. Shih, D. J. Krusienski, and J. R. Wolpaw, “Brain-computer interfaces in medicine,” *Mayo Clin. Proc.*, vol. 87, no. 3, pp. 268–279, 2012.

[4] H. H. Jasper, “The 10/20 international electrode system,” *Electroencephalography and Clinical Neurophysiology*, vol. 10, no. 2. pp. 370–375, 1958.

[5] A. Subasi and E. Erçelebi, “Classification of EEG signals using neural network and logistic regression,” *Comput. Methods Programs Biomed.*, vol. 78, no. 2, pp. 87–99, 2005.

[6] X. Wu et al., Top 10 algorithms in data mining, vol. 14, no. 1. 2008.

[7] Hastie, T.; Tibshirani, R.; Friedman, J. H. (2009). "10. Boosting and Additive Trees". The Elements of Statistical Learning (2nd ed.). New York: Springer. pp. 337–384. ISBN 978-0-387-84857-0. Archived from the original on 2009-11-10.

[8] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. 13-17-August-2016, pp. 785–794, 2016.

[9] Introduction to boosted trees. Introduction to Boosted Trees - xgboost 1.6.0-dev documentation. (n.d.). Retrieved October 19, 2021, from https://xgboost.readthedocs.io/en/latest/tutorials/model.html.

[10] S. Spiegel and R. Hoinkes, “Immersive serious games for large scale multiplayer dialogue and cocreation,” *Serious Games Mech. Eff.*, no. 2005, pp. 469–485, 2009.

[11] K. Mitsis, K. Zarkogianni, K. Dalakleidi, G. Mourkousis and K. S. Nikita, "Evaluation of a Serious Game Promoting Nutrition and Food Literacy: Experiment Design and Preliminary Results," 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), 2019, pp. 497-502, doi: 10.1109/BIBE.2019.00096.

[12] K. Mitsis, E. Kalafatis, K. Zarkogianni, G. Mourkousis and K. S. Nikita, "Procedural content generation based on a genetic algorithm in a serious game for obstructive sleep apnea," 2020 IEEE Conference on Games (CoG), 2020, pp. 694-697, doi: 10.1109/CoG47356.2020.9231785.

[13] K. Mitsis, K. Zarkogianni, N. Bountouni, M. Athanasiou and K. S. Nikita, "An Ontology-Based Serious Game Design for the Development of Nutrition and Food Literacy Skills," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 1405-1408, doi: 10.1109/EMBC.2019.8856604.

[14] V. Wattanasoontorn, I. Boada, R. García, and M. Sbert, *Serious games for health*, vol. 4, no. 4. 2013.

[15] Zarkogianni, K, Athanasiou, M, Mitsis, K, Chatzidaki, E, Polychronaki, N, Perakis, K, Vergeti, D, Antonopoulou, D, Papachristou, E, Chioti, V, others. "A Comprehensive Approach To Empower Self-Management of Health in Childhood Obesity Based on Gamification Mechanisms and Biofeedback". Diabetes Technology & Therapeutics, pp. A57--A58, 2021.

[16] E. Fragkiadakis, K. V. Dalakleidi and K. S. Nikita, "Design and Development of a Sitting Posture Recognition System," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 3364-3367, doi: 10.1109/EMBC.2019.8856635.

[17] J. Bird, "EEG Brainwave Dataset: Feeling emotions," Kaggle, 19-Dec-2018. [Online]. Available: https://www.kaggle.com/birdy654/eeg-brainwave-dataset-feeling-emotions. [Accessed: 26-Oct-2021].

[18] J. Bird, "EEG Brainwave Dataset: Mental state," Kaggle, 09-Mar-2021. [Online]. Available: https://www.kaggle.com/birdy654/eeg-brainwave-dataset-mental-state. [Accessed: 26-Oct-2021].

[19] C. Y. Park, N. Cha, S. Kang, A. Kim, A. H. Khandoker, L. Hadjileontiadis, A. Oh, Y. Jeong, and U. Lee, "K-EmoCon, a multimodal sensor dataset for continuous emotion recognition in naturalistic conversations," Nature News, 08-Sep-2020. [Online]. Available: https://www.nature.com/articles/s41597-020-00630-y. [Accessed: 26-Oct-2021].

[20] Ruo-Nan Duan, Jia-Yi Zhu and Bao-Liang Lu, Differential Entropy Feature for EEG-based Emotion Classification, Proc. of the 6th International IEEE EMBS Conference on Neural Engineering (NER). 2013: 81-84. [link] [BibTex]

[21] Wei-Long Zheng, and Bao-Liang Lu, Investigating Critical Frequency Bands and Channels for EEG-based Emotion Recognition with Deep Neural Networks, accepted by IEEE Transactions on Autonomous Mental Development (IEEE TAMD) 7(3): 162-175, 2015. [link] [BibTex]

[22] Wei-Long Zheng, Wei Liu, Yifei Lu, Bao-Liang Lu, and Andrzej Cichocki, EmotionMeter: A Multimodal Framework for Recognizing Human Emotions. IEEE Transactions on Cybernetics, Volume: 49, Issue: 3, March 2019, Pages: 1110-1122, DOI: 10.1109/TCYB.2018.2797176. [link] [BibTex]

[23] S. Katsigiannis and N. Ramzan, “DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices,” IEEE J. Biomed. Heal. Informatics, vol. 22, no. 1, pp. 98–107, 2018.

[24] "DEAP: A Database for Emotion Analysis using Physiological Signals (PDF)", S. Koelstra, C. Muehl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, EEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18-31, 2012

[25] C. A. Gabert-Quillen, E. E. Bartolini, B. T. Abravanel, and C. A. Sanislow, “Ratings for emotion film clips,” Behav. Res. Methods, vol. 47, no. 3, pp. 773–787, 2015.

[26] M. Soleymani, S. Asghari-Esfeden, Y. Fu, and M. Pantic, “Analysis of EEG Signals and Facial Expressions for Continuous Emotion Detection,” IEEE Trans. Affect. Comput., vol. 7, no. 1, pp. 17–28, 2016.

[27] A. Schaefer, F. Nils, P. Philippot, and X. Sanchez, “Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers,” Cogn. Emot., vol. 24, no. 7, pp. 1153–1172, 2010.

[28] M. M. Bradley and P. J. Lang, “Measuring emotion: The self-assessment manikin and the semantic differential,” *J. Behav. Ther. Exp. Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[29] M. X. Cohen, ”Analyzing Neural Time Series Data: Theory and Practice. The MIT Press, London, 2014.

[30] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, “DEAP: A database for emotion analysis ;using physiological signals,” IEEE Trans. Affect. Comput., vol. 3, no. 1, pp. 18–31, Jan. 2012.

[31] R. J. Davidson, “Affective neuroscience and psychophysiology: toward a synthesis,” Psychophysiology, vol. 40, no. 5, pp. 655–665, Sep 2003.

[32] H. W. Cole and W. J. Ray, “EEG correlates of emotional tasks related to attentional demands,” Int J Psychophysiol, vol. 3, no. 1, pp. 33–41, Jul 1985.

[33] R. J. Barry, A. R. Clarke, S. J. Johnstone, C. A. Magee, and J. A. Rushby, “EEG differences between eyes-closed and eyes-open resting conditions,” Clin Neurophysiol, vol. 118, no. 12, pp. 2765–2773, Dec 2007.

[34] R. J. Barry, A. R. Clarke, S. J. Johnstone, and C. R. Brown, “EEG differences in children between eyes-closed and eyes-open resting conditions,” Clin Neurophysiol, vol. 120, no. 10, pp. 1806–1811, Oct 2009.

[35] M. X. Cohen, ”Analyzing Neural Time Series Data: Theory and Practice. The MIT Press, London, 2014.

[36] D. Novak, M. Mihelj, and M. Munih, “A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing,” Interact. Comput., vol. 24, no. 3, pp. 154–172, May 2012.

[37] A. P. Zanesco, B. G. King, A. C. Skwara, and C. D. Saron, “Within and between-person correlates of the temporal dynamics of resting EEG microstates,” *Neuroimage*, vol. 211, p. 116631, 2020.