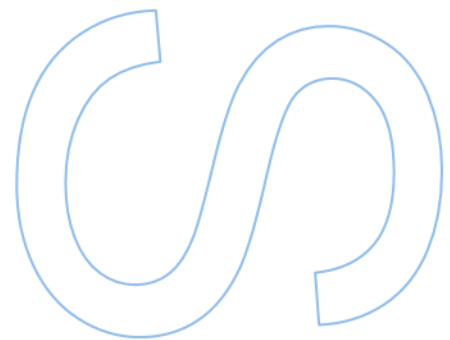
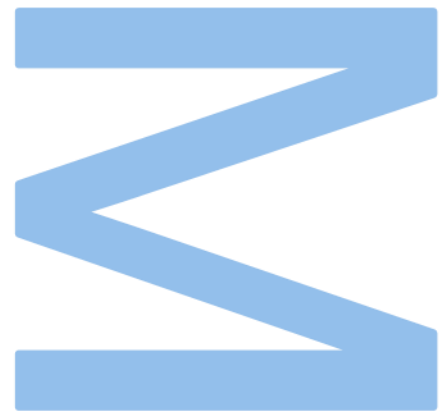




Multimodal Characterization and Classification of Anxiety and Affect



Mariana Vaz

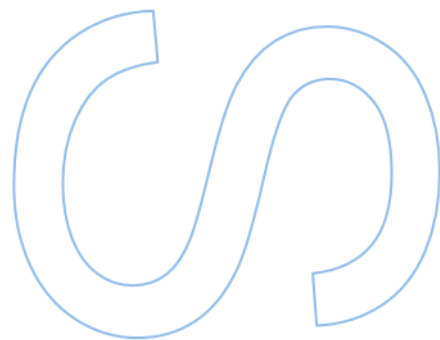
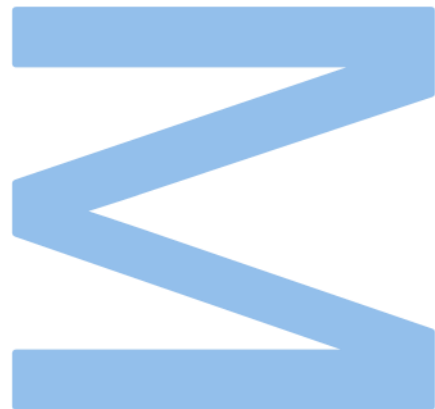
Master's in Bioinformatics and Computational Biology
Department of Biology
2022

Supervisor

Raquel Sebastião, Researcher, University of Aveiro

Co-supervisors

Rita Ribeiro, Assistant Professor, Faculty of Sciences of University
of Porto and INESC TEC
Teresa Summavielle, Researcher, Institute for Research and
Innovation in Health



Sworn Statement

I, Mariana Isabel Ribau Vaz, enrolled in the master's degree in Bioinformatics and Computational Biology at the Faculty of Sciences of the University of Porto hereby declare, in accordance with the provisions of paragraph a) of Article 14 of the Code of Ethical Conduct of the University of Porto, that the content of this dissertation perspective, research work and my interpretations at the time of its submission.

By submitting this dissertation, I also declare that it contains the results of my research work and contributions that have not been previously submitted to this or any other institution.

I further declare that all references to other authors fully comply with the rules of attribution and are referenced in the text by citation and identified in the bibliographic references Section. This dissertation does not include any content whose reproduction is protected by copyright laws.

I am aware that the practice of plagiarism and self-plagiarism constitute a form of academic offence.

Mariana Isabel Ribau Vaz

Porto, 30 de setembro de 2022

Acknowledgements

Gostaria de expressar a minha imensa gratidão a todos os que contribuíram para a realização desta dissertação.

Às minhas orientadoras pela orientação, por todo o apoio, paciência e motivação.

À minha família, pelo amor incondicional e palavras confortantes nas fases difíceis.

A todos os meus amigos, por estarem na minha vida, pelo amor e suporte.

Resumo

Ansiedade e emoções estão associadas a alterações fisiológicas que são cruciais para a manutenção do bem-estar físico e psicológico. Estas respostas podendo estar fortemente relacionadas com distúrbios mentais. Vários estudos demonstram a associação entre estes distúrbios mentais e uma desregulação do Sistema Nervoso Autónomo (SNA) nomeadamente a nível da variabilidade do ritmo cardíaco e do suor. Compreender como as respostas dos sinais fisiológicos e a sua combinação pode contribuir para uma identificação eficaz de estados de ansiedade e afetivos. Esta dissertação, através de uma perspetiva fisiológica, aborda a caracterização e classificação da ansiedade, afetividade negativa e positiva. Para tal, utilizamos sinais fisiológicos: Eletrocardiograma (ECG), Atividade eletrodérmica (EDA) e o Eletromiograma (EMG) Uma combinação de variáveis destes sinais foi usada para descrever a ansiedade e afetividade usando uma amostra de indivíduos saudáveis, sob uma condição neutra. Este estudo é dividido em (1) extração, análise e seleção de variáveis; (2) caracterização de ansiedade e afetividade; (3) classificação da ansiedade e afetividade utilizando técnicas de aprendizagem computacional e de balanceamento de dados.

Os resultados da dissertação permitiram-nos caracterizar ansiedade, afetividade negativa e afetividade positiva através de variáveis discriminatórias extraídas dos sinais fisiológicos mencionados. As técnicas supervisionadas de seleção de variáveis foram críticas para a maioria dos classificadores nas três tarefas de classificação, assim como as técnicas de balanceamento, especialmente Borderline SMOTE 2. Com a combinação da seleção de variáveis e do balanceamento dos dados obtivemos os seguintes melhores desempenhos de classificação: para a ansiedade, sem seleção supervisionada de variáveis, XGB e ADASYN um F1 de 0.870 (EP=0.051) e com RFECV, RF e Borderline SMOTE 2 uma Roc-Auc de 0.980 (EP=0.013); para a afetividade negativa, com SelectKBest, RF e Borderline SMOTE 2 um F1 de 0.880 (EP=0.072) e com RFECV, RF e Random Oversampling uma Roc-Auc de 1.000 (EP=0.000); para a afetividade positiva, um F1 de 0.997 (EP=0.003) com LR com SelectKBest e RFECV e com SVM e RFECV e uma Roc-Auc 1.000 (EP=0.000) com LR e SVM com e sem seleção de variáveis supervisionada.

Palavras-Chave Ansiedade, Afetividade, Classificação, Sensores portáteis, Dados Multimodais, Sinais fisiológicos, Questionários

Abstract

Anxiety and emotional responses are associated with distinct body alterations and are crucial to someone's well-being. They strongly impact mental health, resulting in increased depression and anxiety disorders. Multiple studies show an association between these mental disorders and dysregulation in the Autonomic Nervous System (ANS) such as changes in Heart Rate Variability (HRV) and sweating. Thus, understanding how informative physiological signals and their combination are would contribute to more effective identification and interpretation of anxiety and affect. This dissertation targets a physiological approach to the characterization and classification of anxiety, negative and positive affect states. For that purpose, we analyzed Electrocardiogram (ECG), Electrodermal activity (EDA), and Electromyogram (EMG) signals, to characterize and learn several models capable of predicting anxiety and affect. A combination of features extracted from these signals was used to describe anxiety and affect, using a sample of healthy subjects under a neutral condition. Our study was divided into (1) feature extraction, analysis, and selection; (2) characterization of anxiety, and negative and positive affect; (3) classification of anxiety, negative and positive affect using strategies to cope with imbalanced classification tasks.

The findings of the dissertation allowed us to characterize anxiety, negative affect, and positive affect with class discriminatory features from all the physiological signals. The supervised feature selection techniques showed valuable for most classifiers in all three classification tasks, as well as the data balance techniques, especially Borderline SMOTE 2. The combination of feature selection and data balancing with the different classifiers achieved the following best classification scores: (1) for anxiety, without supervised feature selection, XGB and ADASYN an F1-score of 0.870 (SE=0.051); while performing feature selection with RFECV, RF and Borderline SMOTE 2 a Roc-Auc of 0.980 (SE=0.013); (2) for negative affect, with SelectKBest, RF and Borderline SMOTE 2 an F1 of 0.880 (SE=0.072) and with RFECV, RF and Random Oversampling a Roc-Auc of 1.000 (SE=0.000); (3) for positive affect classification, an F1 of 0.997 (0.003) with LR with SelectKBest and RFECV and with SVM with RFECV and a Roc-Auc of 1.000 (SE=0.000) with LR and SVM, with and without supervised feature selection.

Keywords Anxiety, Affect, Classification, Wearable sensors, Multimodal dataset, Machine learning, Physiological signals, Self-reports

Table of Contents

List of Tables	vi
List of Figures	viii
List of Abbreviations	xii
1. Introduction.....	1
1.1. Contextualization	1
1.2. Motivation and Challenges.....	2
1.3. Objectives and Contributions	3
1.4. Structure of the Dissertation	3
2. Literature Review.....	5
2.1. Psychophysiology of Mental Disorders	5
2.1.1. Anxiety.....	5
2.1.2. Affect	7
2.2. Physiological Signals	8
2.3. Machine Learning in Physiological Domains	10
2.3.1. Related Work in Multimodal Classification Tasks.....	11
2.3.2. Imbalanced Classification	15
2.3.3. Data Pre-processing Techniques.....	15
2.3.4. Suitable Performance Metrics	19
2.4. Dissertation Summary	20
3. Data and Methods	21
3.1. Dataset.....	21
3.2. Methodology	22
3.2.1. Raw Signals Filtering.....	23
3.2.2. Class Assignment.....	24
3.2.3. Feature Extraction	26
3.2.4. Feature Analysis and Selection	29
3.2.5. Pipeline Description.....	33
3.2.6. Supervised Feature Selection – Filter and Wrapper Methods	33
3.2.7. Feature Normalization	35
3.2.8. Data Balance Techniques.....	35

3.2.9.	Classification of Anxiety and Affect levels	36
3.2.10.	Anxiety and Affect Classes Characterization	33
4.	Results and Discussion.....	38
4.1.	Univariate Analysis	38
4.1.1.	Anxiety Characterization.....	39
4.1.2.	Negative Affect Characterization.....	40
4.1.3.	Positive Affect Characterization	42
4.2.	Unsupervised Feature Selection.....	45
4.3.	Supervised Feature Selection and Classification Results.....	46
4.3.1.	Anxiety Feature Selection	46
4.3.2.	Anxiety Classification.....	48
4.3.3.	Negative Affect Feature Selection.....	55
4.3.4.	Negative Affect Classification	56
4.3.5.	Positive Affect Feature Selection	61
4.3.6.	Positive Affect Classification	63
4.3.7.	Summary of the Classification Results.....	65
5.	Conclusion.....	66
5.1.	Contributions	66
5.2.	Limitations and Future work.....	67
	References	69
6.	Attachments.....	76
6.1.	Features Description.....	76
6.2.	Discriminant negative and positive affect features	84
6.3.	Questionnaires' description.....	88
6.4.	Models Parameters and Parameters Optimization	89
6.5.	Data Balance Techniques Parameters.....	90
6.6.	Combination of Features and Data Balance Technique	90
6.7.	Anxiety Classification Results	95
6.8.	Negative Affect Classification Results.....	98
6.9.	Positive Affect Classification Results	101

List of Tables

Table 3-1	Extracted features from the filtered physiological signals.	29
Table 4-1	Significant features for anxiety classification.	38
Table 4-2	Significant features for negative affect classification.	38
Table 4-3	Significant features for positive affect classification.	38
Table 4-4	Features selected after the unsupervised feature selection and before the supervised feature selection.	46
Table 6-1	STAI questionnaire items.	88
Table 6-2	PANAS questionnaire items.	88
Table 6-3	Models' hyperparameters and optimization values.	90
Table 6-4	Data balance techniques functions and the selected parameters.	90
Table 6-5	Best combinations for anxiety classification using SelectKBest.	91
Table 6-6	Best combinations for anxiety classification using RFECV.	93
Table 6-7	Best combinations for negative affect classification using SelectKBest.	93
Table 6-8	Best combinations for negative affect classification using RFECV.	95
Table 6-9	LR anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).	95
Table 6-10	LDA anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).	96
Table 6-11	DT anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).	96
Table 6-12	SVM anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).	96
Table 6-13	ADB anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).	97
Table 6-14	RF anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).	97
Table 6-15	XGB anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).	98
Table 6-16	LR negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R).	98
Table 6-17	LDA negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R).	99

Table 6-18 DT negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R). 99

Table 6-19 SVM negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R)..... 100

Table 6-20 ADB negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R). 100

Table 6-21 RF negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R). 100

Table 6-22 XGB negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R). 101

Table 6-23 Positive affect classification scores without and with SelectKBest (S) and RFECV (R). 102

List of Figures

Fig. 2-1 - Segment of the ECG signal of one of the study's participants.	9
Fig. 2-2 - Segment of the EDA signal of one of the study's participants.	9
Fig. 2-3 - Segment of the EMG signal from one of the study's participants.	10
Fig. 2-4 - PCA analysis of the SMOTE variants applied to our data without supervised feature selection.	18
Fig. 3-1 - Schematic representation of the WESAD dataset protocol.	21
Fig. 3-2 - Graphical representation of the study's workflow.	23
Fig. 3-3 - On the right, the bar plot shows the number of individuals per anxiety class. On the left, the bar plot shows the number of participants for anxiety measured in the STAI questionnaire.	25
Fig. 3-4 - On the right, the bar plot shows the number of individuals per negative affect class. On the left, the bar plot shows the number of participants with negative affect scores measured in the PANAS questionnaire.	26
Fig. 3-5 - On the right, the bar plot shows the number of individuals per positive affect class. On the left, the bar plot shows the number of participants with positive affect scores measured in the PANAS questionnaire.	26
Fig. 3-6 - Plot of the ECG signal peaks: P, Q, S, T (on the left) and R (on the right). ..	28
Fig. 3-7 - On the left, the EDA filtered signal is presented in blue. The red line is - the SCR onset; the blue line is - the SCR peak; the green line is - the SCR recovery point. On the right, the figure shows the EDA filtered signal and the Tonic and Phasic components.	28
Fig. 3-8 - Plot of the EMG filtered signal (in yellow) and the activation peaks marked with red dots.	28
Fig. 3-9 - Correlation of the ECG features – Max ECG, Min ECG, Med ECG, Mean ECG, SD ECG, Var ECG, Range ECG.	31
Fig. 3-10 - Correlation of the ECG peak features – SDPP, SDQQ, SDRR, SDSS, SDTT, MeanPP, MeanQQ, MeanRR, MeanSS, MeanTT, MedPP, MedQQ, MedRR, MedSS, MedTT.	31
Fig. 3-11 - Correlation of the HRV time features - RMSSD, MeanNN, SDNN, CVNN, CVSD, MedianNN, MadNN, MCVNN, pNN50, pNN20, TINN.	31
Fig. 3-12 - Correlation of the HRV frequency and non-linear features - LF, HF, VHF, LFHF, LFn, HFn, LnHF, ApEn, SampEn.	31

Fig. 3-13 - Correlation of the AUCC and the HR features – Min HR, Max HR, Mean HR, Med HR, SD HR, Var HR, and Range HR. 31

Fig. 3-14 - Correlation of the EMG features – Max EMG, Min EMG, Mean EMG, Med EMG, SD EMG, Var EMG, Range EMG, Perc 10 EMG, and Perc 90 EMG. 31

Fig. 3-15 - Correlation of the EMG Envelope features – Max Env, Min Env, Mean Env, Med Env, SD Env, Var Env, Range Env, Perc 90 Env, and Perc 10 Env. 32

Fig. 3-16 - Correlation of the EMG Envelope features - Nr Pulse Onsets/s, Nr Activ Peaks/s, Med/Mean Activ Peaks, and Pulse Onsets Amp. 32

Fig. 3-17 - Correlation of the EDA features – Max EDA, Min EDA, Mean EDA, Med EDA, Var EDA, SD EDA, and Range EDA..... 32

Fig. 3-18 - Correlation of the SCL features – Max SCL, Min SCL, Mean SCL, Med SCL, SD SCL, Var SCL, Range SCL, and Corr SCL. 32

Fig. 3-19 - Correlation of the SCR features – Mean/Med peaks Amp, Mean/Med Recovery and Onsets Amp, and Nr SCR Peaks/s. 32

Fig. 3-20 - Correlation of the SCR features – Max SCR, Min SCR, Mean SCR, Med SCR, SD SCR, Var SCR, Range SCR, Corr SCR, Mean/Med Rise, and Recovery Time. 32

Fig. 3-21 - Pipeline scheme with supervised feature selection techniques. 33

Fig. 3-22 - Pipeline scheme with data balance techniques. 33

Fig. 4-1 - Boxplot of the EMG signal feature - Med EMG. 39

Fig. 4-2 - Boxplots of the EMG signal features - Max EMG (on the left) and Nr Activ Peaks/s (on the right)..... 41

Fig. 4-3 - Boxplots of the EDA signal features - SD SCR (on the left) and Nr SCR peaks/s (on the right). 42

Fig. 4-4 - Boxplots of the HRV features – MedianNN (on the left) and SDNN (on the right). 42

Fig. 4-5 - Boxplot of the heart rate features – Med HR (on the left) and Min HR (on the right). 44

Fig. 4-6 - Boxplot of the EMG signal feature - Med Env. 44

Fig. 4-7 - Boxplots of the EDA signal features - Med SCL (on the left) and SD SCL (on the right). 45

Fig. 4-8 - Features selected at least three times with SelectKBest (on the left) and with RFECV (on the right) for anxiety classification..... 47

Fig. 4-9 - Anxiety classification results without supervised feature selection (top: F1-score, bottom: Roc-Auc). 49

Fig. 4-10 - PCA analysis of the anxiety classes..... 51

Fig. 4-11 - PCA analysis of the data balance techniques applied to the anxiety classes.
 51

Fig. 4-12 - Anxiety classification results using SelectKBest (top: F1-score, bottom: Roc-
 Auc)..... 52

Fig. 4-13 - Anxiety classification results using RFECV (top: F1-score, bottom: Roc-Auc).
 53

Fig. 4-14 - Features selected at least three times with SelectKBest (on the left) and with
 RFECV (on the right) for negative affect classification. 55

Fig. 4-15 - Negative affect classification results without supervised feature selection (top:
 F1-score, bottom: Roc-Auc)..... 57

Fig. 4-16 - PCA analysis of the negative affect classes..... 58

Fig. 4-17 - PCA analysis of the data balance techniques applied to the negative affect
 classes. 58

Fig. 4-18 - Negative affect classification results using SelectKBest (top: F1-score,
 bottom: Roc-Auc)..... 59

Fig. 4-19 - Negative affect classification results using RFECV (top: F1-score, bottom:
 Roc-Auc). 60

Fig. 4-20 - Features selected at least three times with SelectKBest (on the left) and with
 RFECV (on the right) for positive affect classification..... 62

Fig. 4-21 - Positive affect classification results without supervised feature selection, with
 SelectKBest and RFECV (top: F1-score, bottom: Roc-Auc)..... 63

Fig. 6-1 - ECG signal extracted features: HRV time, frequency, and non-linear features.
 76

Fig. 6-2 - ECG signal extracted features: HR and ECG signal statistical features. 77

Fig. 6-3 - ECG signal extracted features: non-normalized peaks' features. 78

Fig. 6-4 - EDA signal extracted features: EDA signal statistical features. 78

Fig. 6-5 - EDA signal extracted features: SCR statistical features. 79

Fig. 6-6 - EDA signal extracted features: mean and median time of the rise and recovery
 of the SCR peaks. 80

Fig. 6-7 - EDA signal extracted features: SCL statistical features..... 80

Fig. 6-8 - EDA signal extracted features: SCR peaks' features. 81

Fig. 6-9 - EMG signal extracted features: statistical features..... 82

Fig. 6-10 - EMG signal extracted features: EMG envelope statistical features. 83

Fig. 6-11 - Boxplots of the EDA signal significant features for negative affect. 84

Fig. 6-12 - Boxplots of the Standard deviation of the time interval between ECG peaks, significant for positive affect..... 84

Fig. 6-13 - Boxplots of positive affect significant features that measure the variation of the time interval between R peaks..... 85

Fig. 6-14 - Boxplots of the features mean interval between ECG peaks, significant for positive affect. 85

Fig. 6-15 - Boxplots of the median intervals between ECG peaks, significant for positive affect. 86

Fig. 6-16 - Boxplots of positive affect significant features that measure the interval between R peaks..... 86

Fig. 6-17 - Boxplots of the significant features Mean HR and AUCC, for positive affect. 87

Fig. 6-18 - Boxplots of the EDA signal significant features for positive affect. 87

List of Abbreviations

ACC	MOTION ACCELEROMETER
ADA	ADASYN
ADB	ADAPTATIVE BOOSTING
ANN	ARTIFICIAL NEURAL NETWORKS
ANOVA	ANOVA - ANALYSIS OF VARIANCE
ANS	AUTONOMIC NERVOUS SYSTEM
APEN	APPROXIMATE ENTROPY
BORDER	BORDERLINE SMOTE 2
BP	BLOOD PRESSURE
BV	RESPIRATION VOLUME
BVP	BLOOD VOLUME PULSE
CBFS	CORRELATION-BASED FEATURE EXTRACTION
CV	COEFFICIENT OF VARIATION
CVNN	SDNN DIVIDED BY THE MEANNN
CVSD	RMSSD DIVIDED BY THE MEANNN
DT	DECISION TREE
ECG	ELECTROCARDIOGRAM
EDA	ELECTRODERMAL ACTIVITY
EMG	ELECTROMYOGRAM
FCUP	FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO
FMRI	FUNCTIONAL MAGNETIC RESONANCE IMAGING
FPR	FALSE POSITIVE RATE
F1	F1-SCORE
GBM	GRADIENT BOOSTING MACHINES
GSR	GALVANIC SKIN RESPONSE
HF	HIGH-FREQUENCY HRV
HMM	HIDDEN MARKOV MODELS

HPA	HYPOTHALAMIC-PITUITARY-ADRENAL
HR	HEART RATE
INESC TEC	INSTITUTE FOR SYSTEMS AND COMPUTER ENGINEERING, TECHNOLOGY AND SCIENCE
IR	IMBALANCE RATIO
KNN	K-NEAREST NEIGHBORS
W/O	WITHOUT
LDA	LINEAR DISCRIMINANT ANALYSIS
LF	LOW-FREQUENCY
LF/HF	RATIO BETWEEN LOW-FREQUENCY HRV AND HIGH- FREQUENCY HRV RATIO
LR	LOGISTIC REGRESSION
MADNN	MEDIAN ABSOLUTE DEVIATION OF THE N-N INTERVALS
MAE	MEAN ABSOLUTE ERROR
MSCVNN	MADNN DIVIDED BY MEDIANNMARKOV MODELS
MM	MARKOV MODELS
MW	MANN-WHITNEY U
N	NORMALIZED R PEAKS
NA	NEGATIVE AFFECT
NB	NAÏVE BAYES
OCDE	ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT
OVER	RANDOM OVERSAMPLING
PA	POSITIVE AFFECT
PANAS	POSITIVE AND NEGATIVE AFFECT SCHEDULE
PCA	PRINCIPAL COMPONENT ANALYSIS
PD	PUPIL DIAMETER
PNN20	PROPORTION OF N-N INTERVALS GREATER THAN 20MS

PNN50	PROPORTION OF N-N INTERVALS GREATER THAN 50MS
PSD	PSD - POWER SPECTRAL DENSITY
RE	RELATIVE ERROR
RESP	RESPIRATORY ACTIVITY
RF	RANDOM FOREST
RFE	RECURSIVE FEATURE ELIMINATION
RFECV	RECURSIVE FEATURE ELIMINATION WITH CROSS-VALIDATION
RMSSD	SQUARE ROOT OF THE MEAN OF THE SUM OF SUCCESSIVE DIFFERENCES BETWEEN ADJACENT R-R INTERVALS
ROC CURVE	RECEIVER OPERATING CHARACTERISTIC CURVE
ROC-AUC	AREA UNDER THE ROC CURVE
RR	RESPIRATION RATE
SAMPEN	SAMPLE ENTROPY
SBS	SEQUENTIAL BACKWARD SEARCH
SCL	SKIN CONDUCTANCE LEVEL
SCR	SKIN CONDUCTANCE RESPONSE
SDNN	STANDARD DEVIATION OF THE N-N INTERVALS
SDSD	STANDARD DEVIATION OF THE SUCCESSIVE DIFFERENCES BETWEEN N-N INTERVALS
SE	STANDARD ERROR
SFS	SEQUENTIAL FORWARD SEARCH
SMO	SMOTE
SMOTE	SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE
SNR	SIGNAL-TO-NOISE RATIO
ST	SKIN TEMPERATURE
STAI	STATE-TRAIT ANXIETY INDEX
SVM	SUPPORT VECTOR MACHINES
SW	SHAPIRO-WILK

TEMP	TEMPERATURE
TI	THERMAL IMAGING
TINN	TRIANGULAR INTERPOLATION OF N-N INTERVAL HISTOGRAM
TPR	TRUE POSITIVE RATE
VLF	VERY LOW-FREQUENCY HRV
XGB	EXTREME GRADIENT BOOSTING

1. Introduction

This chapter will contextualize the master dissertation and explain the motivation, and challenges of the research. Also, it will expose the main objectives of this work and present the general structure of the document.

1.1. Contextualization

Stress and anxiety have become buzzwords of present times, affecting many people worldwide, irrespective of gender, age group, or work profile. Possible reasons for this trend may be the challenging demands in contemporary work culture, changing lifestyles, and technological interactions [1]–[3].

Indeed, public health statistics reveal the alarming impact that anxiety-associated complications have on all aspects of society. From the 2021 health statistics report of the Organization for Economic Co-operation and Development (OCDE), in 2019 Portugal was the OCDE country with the highest anxiolytics consumption and the second in antidepressants consumption in 2020 [4].

Anxiety is defined as being a specific reaction to stress and persists even after a concern has passed, while affect is a term that encompasses both emotions and moods. Anxiety and affect can manifest psychologically and physiologically through specific responses of the nervous system. A strong link has been reported between prolonged and severe anxiety and affect with mental disorders, such as depression and anxiety disorders, as well as an increased risk for other health complications like cardiovascular diseases [5]–[9].

Furthermore, multiple studies show an association between anxiety disorders and depression, and an imbalance in the Autonomic nervous system (ANS), characterized by increased sympathetic activity and reduced parasympathetic or vagal activity [6], [10], [11].

The ANS [5] maintains the homeostasis of unconscious bodily functions such as heart and respiration rates, blood pressure, digestion, urination, or sweating, among others. The two main branches of ANS are the Sympathetic nervous system (SNS) and the Parasympathetic nervous system (PNS). The SNS [12] responds to emergencies, activating our organs and functions to respond to stress (fight-or-flight responses), while the PNS [12] relaxes these functions back to rest (rest-and-digest activity).

1.2. Motivation and Challenges

The ability to understand and perceive mental disorders has even more relevance considering the new challenges associated with the COVID-19 pandemic, which include increased anxiety and depression rates at a global level.

Anxiety and affect recognition aspire to detect states of anxiety and affect in a person. Its detection is a highly interdisciplinary research field with links to signal processing, machine learning, and neuroscience [13].

Anxiety disorders' prevalence across the population vastly exceeds the capacity of mental health services to provide face-to-face therapy for all those affected, demanding novel approaches for delivering therapy [14]. Traditional monitoring methods rely on retrospective reports which are subject to recall bias and limit the ability to accurately understand behaviour in real-time settings [15]. Therefore, having access to valuable physiological information would greatly enrich the process of intervening in these situations [14].

This master dissertation relies on physiological signals, namely Electrocardiogram (ECG), Electrodermal Activity (EDA), Electromyogram (EMG), and self-reports to characterize and classify the data according to anxiety and affect states.

The main challenges are, precisely, concerned with their physiological characterization through these different signals, arising either from the combination of several features extracted from them or by the potential of each signal to contribute more to a better description. These tasks require that feature importance and selection be carefully attained.

Another challenge comprises the fact that frequently, in a non-clinical population, more severe cases of anxiety and affect are rarer [1]. Nevertheless, training machine learning models to predict those states accurately is of crucial importance as they represent higher health and well-being risks.

The WESAD dataset [1], used in this work, is imbalanced as more severe cases of anxiety and affect are not as common and consequently, the various levels are not equally represented. This poses difficulties to standard classification algorithms. Data pre-processing techniques [16] are widely applied strategies to cope with imbalanced domain learning. In this work, we resort to techniques that involve oversampling and synthetic data generation as a means to improve the classification of anxiety and affect.

1.3. Objectives and Contributions

This dissertation proposes to attain the following objectives that can help in the current increased needs for the management of mental health:

1. physiological characterization of anxiety and positive and negative affect states, providing further knowledge about the physiological interactions that rule these emotional and physiological responses.

2. automatic detection by classifying anxiety, and positive and negative affect levels using machine learning algorithms.

3. test how to improve the classification of anxiety and affect states in such an imbalanced domain and if different feature selection techniques and data-level pre-processing techniques can be good solutions to the problem of imbalanced class distribution in mental health datasets.

We aim to achieve these goals by the combination of physiological signals recorded by wearable sensors, using the public dataset WESAD [1] described in Chapter 3, Section 3.1.

Our approach differs from conventional studies as it will address anxiety and affect states without any controlled elicitation for those. This ensures the characterization and classification of anxiety and emotions in a neutral condition, providing a useful way for people to better manage their health and well-being [17]. Moreover, by tackling this from a multi-signal perspective, we ensure higher classification accuracy, as research has pointed out that recognition systems that base their decisions on multiple physiological data tend to be more accurate than their single-signal counterparts [13].

The main contributions of this work are the findings of the study, which expect to achieve a deeper understanding of mental disorders and physiological relations. By resorting to machine learning techniques and physiological signals collected from wearable devices, this work also supports the development of real-time systems for the early detection of anxiety and affect and related undesirable health outcomes.

1.4. Structure of the Dissertation

Chapter 1 introduces the dissertation's motivation, goals, and contributions. Chapter 2 presents key concepts and related work in the literature. Chapter 3 describes the data and methods used in this study. Chapter 4 shows and discusses the obtained

results and Chapter 5 addresses the main conclusions of the dissertation, the study's limitations, and future work.

2. Literature Review

In this chapter, some relevant concepts will be presented along with related work developed so far. In recent years, several studies have been conducted to detect anxiety and different emotions based on physiologic parameters [1], [18], [19].

These studies comprise a combination of three different areas of knowledge: neuroscience, to better understand the biological basis of the behavioural and psychological responses to anxiety and affect; biomedical signal processing, which allows the analysis of the collected physiological signals to retrieve relevant information from the mechanisms underlying anxiety and affect; and machine learning, which uses computational techniques to detect patterns in the physiological data. The combinations of these areas will be addressed in this dissertation to characterize and classify different states of anxiety, and positive and negative affect.

2.1. Psychophysiology of Mental Disorders

The brain is the organ that determines what is threatening, and therefore “stressful”, and that orchestrates behavioural and physiological responses to a stimulus. In response to a stressful trigger, the brain and body respond to cope and adapt. Physiologically, the Sympathetic nervous system (SNS), Parasympathetic nervous system (PNS), Hypothalamic-pituitary-adrenal (HPA), the immune system and metabolic hormones, and the molecular processes within all organs operate non-linearly to promote adaptation via allostasis, meaning achieving stability via activation of these systems [20].

However, these same mechanisms that help us respond to challenging situations through allostasis can simultaneously promote pathophysiology when overused. Allostatic load and allostatic overload refer to the cumulative result of an allostatic state and can be defined as “the price the body pays for containing the effects of arousing stimuli and the expectation of negative consequences” [7], [20].

2.1.1. Anxiety

In the current study, we address anxiety states. Anxiety is a person’s specific response to stress with an internal origin. Anxiety normally refers to the anticipation of future concerns and is typically characterized by a “persistent feeling of apprehension or dread” [21]. Anxiety is usually noticeable in two dimensions: psychological and physical. Psychological symptoms of anxiety are recognized as irritability, tension, disturbance, and restlessness, while physical reactions involve the adrenal glands, the autonomous

nervous system, and the production of hormones [22]. Recurrent and/or intense exposure to anxiety may lead to anxiety disorders. Anxiety disorders differ from normal feelings of nervousness or anxiousness by involving excessive fear and abnormal reactions to common daily stressors and include various phobias, panic, social anxiety, generalized anxiety disorders, etc [23].

It is understood that anxiety's underlying physiological mechanisms are controlled by the ANS. The ANS plays a central role in regulating unconscious bodily functions such as heart and respiration rates, blood pressure, and digestion [18], and is composed of two main branches: the SNS and the PNS [5]. The SNS controls activities that are mobilized during emergencies, which characterize the fight-or-flight response. The PNS controls the basic functions of rest, repair, and restoration of energy stores (rest-and-digest activity) [12].

Stressful events or emergencies cause dynamic changes in ANS, increasing the activity rate in the SNS and decreasing the PNS activity. Anxiety-induced changes in the ANS are reflected in a series of physiological parameters: a rise in Heart rate (HR), Blood pressure (BP), or Respiration rate (RR); an increased Galvanic skin response (GSR), lowered Blood volume pulse (BVP) and Heart rate variability (HRV), skin temperature (ST) rises, and muscle tension increases [18], [19].

Heart rate variability (HRV) is a popular non-invasive measure to detect cardiovascular conditions and ANS activities [18]. HRV is extracted from the ECG signal and differently from the Heart rate (HR) - which measures the number of heart beats per minute - HRV refers to the fluctuation in the time intervals between adjacent heart beats [24]. Time-domain and frequency-domain measures have been employed to evaluate HRV. An increased/decreased HRV indicates increased activity of the PNS/SNS, respectively. Anxiety disorders are shown to be associated with reduced high-frequency HRV (HF) [5], [18].

Changes in electrodermal activity (EDA) are another effective measure to assess SNS activity [13]. The EDA signal is composed of two main components: the Skin conductance level (SCL) and Skin conductance response (SCR). The SCR is a measurement of changes in sweat gland activity that produces an increase in the electrical conductivity of the skin. The SCR usually reflects the conscious expectancy of a possible negative outcome thereby serving as an objective indicator of anticipation and arousal [5], [18].

Anxiety can also manifest psychologically. This dimension of anxiety is called psychological or perceived anxiety and occurs due to the mental appraisal of stressful situations of an individual. These can be assessed through self-reports, which are regarded as a good measure of an individual's anxiety level, such as the State-Trait Anxiety Index (STAI) [19], [25].

The original STAI has 40 items, with 20 items allocated to each of the state and trait subscales, nonetheless shorter versions have been developed. The one used in this dissertation is the 6-item STAI which has been applied extensively in several medical conditions [26]. Responses for the S-anxiety scale assess the intensity of current feelings "at this moment" while the responses for the T-anxiety scale assess the frequency of feelings "in general", scoring each response from 1 to 4 ("almost never" to "almost always").

2.1.2. Affect

In scientific literature, happiness is composed of two interrelated components: affect and satisfaction with life. Affect plays a central role in human experience and well-being, and the term encompasses feelings, emotions, moods, and affective traits [27].

Moreover, the study of affect has increased in recent years because it impacts a broad range of aspects of human personality [27]. Similarly, to anxiety, affect can be captured by physiological signals. At an affect physiological level, both the autonomic and central nervous systems play a central role and are responsible for specific internal reactions such as changes in heart activity, profuse sweating, atypical facial expressions, and muscle compression [28], [19]. An example of how affect states and physiological changes are linked is when someone tells a joke, we laugh or smile (physiological response of amusement), or when we are afraid, we sweat or get a dry mouth [13].

As per the facts presented, lowered, or increased HRV can distinguish between emotions. Along similar lines, it was demonstrated that GSR can differentiate between the emotions of "fear" versus "anger", "fear" versus "sadness", and "happiness" versus "sadness" [19]. It was also shown that muscle activity increases with negative valence emotions [29].

In general, affect can be divided into two dimensions considering their valence: positive affect (PA) and negative affect (NA) [30]. PA reflects the extent to which a person feels enthusiastic, active, and alert. High PA is a state of high energy, full concentration, and pleasurable engagement, whereas low PA is characterized by sadness and lethargy.

In contrast, NA is a general dimension of subjective distress and unpleasurable engagement, where high NA considers a variety of aversive mood states like anger, contempt, disgust, guilt, fear, and nervousness, and low NA is a state of calmness [30].

These two dimensions of affect can be measured by the Positive and Negative Affect Schedule (PANAS) [30], a self-reported scale developed by Watson et al., widely used in the field of psychology to access how negative or positive someone's affective state or personality affective trait is. The PANAS consists of two 10-item scales for PA and NA, respectively [30]. It is important to emphasize that PA and NA scales are considered independent of each other [31]. Depending on the instructions set, a person's score on the PANAS scale is conceived as a dispositional trait affect ("In general, how often do you feel...") or as a momentary/state affect ("Recently, how often have you felt...") [30]. Although there are time fluctuations in the levels of affect a person experiences, the stability of self-reports is impressive [30] – [33].

The study of the structure of affect has been particularly important in increasing psychopathological and clinical knowledge about mental disorders such as anxiety disorders and depression. NA is associated with several physical and mental health outcomes that have received attention for years. Only more recently, the role of PA has started to be examined, with some studies suggesting it to be an important factor. NA and PA have been associated with anxiety and depression. High NA is present in both mental conditions, but PA only presents abnormally low scores in depression [32].

2.2. Physiological Signals

Biomedical signals can reflect spontaneous ongoing activity or activity that results from external stimulation. Physiological signals of interest for stress, anxiety and affect detection include hormone levels, ECG, EEG, EDA, BP, skin temperature (ST), EMG, Respiration volume (BV), Pupil diameter (PD), Thermal imaging (TI), and functional Magnetic resonance imaging (fMRI) [18], [19]. From the physiological signals, several parameters associated with the physiological processes of anxiety or affect can be obtained, which reinforces the need to pre-process these signals to remove noise and artefacts during collection.

- ECG: The ECG signal, shown in Fig. 2-1, is a graphical recording of the electrical activity produced by an impulse of ions flowing through cardiac muscles, which dissipates into the region around the heart. The main electrical signals are produced by cardiac cells depolarizing and repolarizing. Depolarizing results in atrial heart muscle

constriction which results in a P wave. The impulse then travels through the ventricles of the heart causing septal depolarization, early ventricular depolarization followed by late ventricular depolarization. This series forms a QRS wave. After the end of depolarization, ventricular cells repolarize, resulting in a T wave. ECG signals are periodic and have persistent features such as the R–R interval [18].

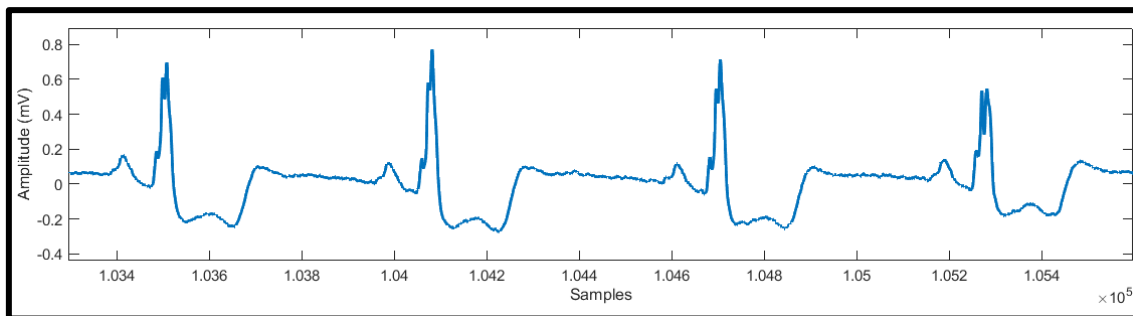


Fig. 2-1 - Segment of the ECG signal of one of the study's participants.

- EDA: The EDA signal, shown in Fig. 2-2, is a measurement of the flow of electricity through the skin which can be taken by measuring electrical potentials between electrodes placed on the surface of the skin [18]. When the individual is under stress, skin conductance is increased due to increased moisture on the surface of the skin, which increases the flow of electricity. The EDA signal is composed of two main components: the SCL component and the SCR component. The SCL [1] represents a slowly varying baseline conductivity, while the SCR is a short-term response to a stimulus. To measure a slowly moving baseline, researchers use the Tonic or SCL component. For arousal event detection, the Phasic or SCR [34] component is employed. EDA signal is reported as one of the best discriminative signals along with the ECG signal [34].

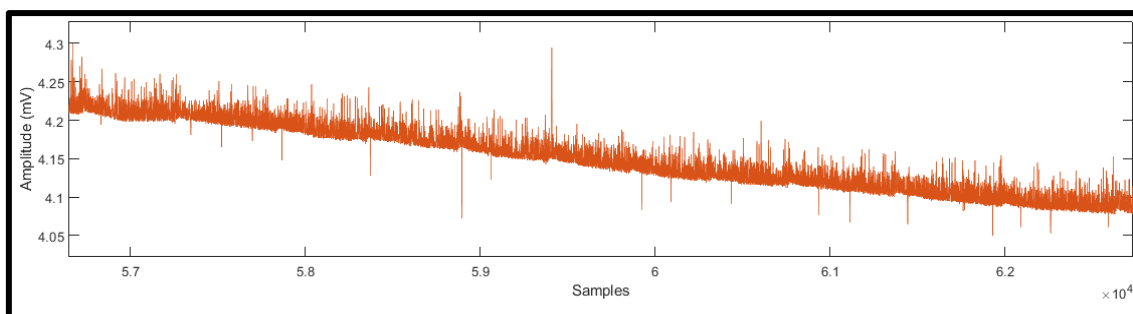


Fig. 2-2 - Segment of the EDA signal of one of the study's participants.

- EMG: The EMG signal, shown in Fig. 2-3, reflects the electrical activity of active skeletal muscles – the muscle action potentials. The EMG signal conveys information about the function of the central and peripheral nervous systems on the muscles. As

such, the EMG signal provides a useful characterization, since many pathological processes, whether arising in the nervous system or the muscle, are manifested by alterations in the signal properties. For anxiety detection, EMG electrodes have been placed on the trapezius muscle (on both sides of the body), which is in the shoulder. If the EMG signal is recorded on the skin over the muscle instead of inside the muscle, it is called Superficial EMG [35].

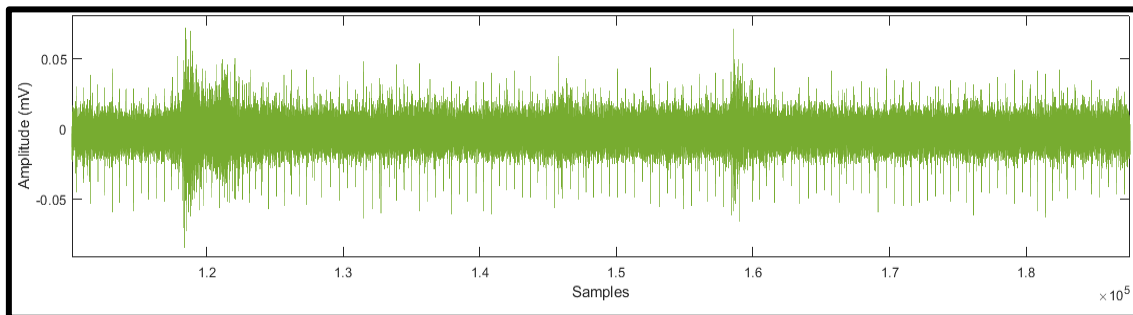


Fig. 2-3 - Segment of the EMG signal from one of the study's participants.

2.3. Machine Learning in Physiological Domains

Machine Learning has been popularly applied in anxiety and affect detection. The specificity of the physiologic response has been used to train machine learning models to classify anxiety and affect levels [1].

Physiological features exhibit several unique characteristics that contribute to the reliability, accuracy, and robustness of anxiety detection and classification systems [13]. Machine learning aims to build models from data so that, using the right features, one can make inferences over new examples. This is possible because almost all non-random data contains patterns that allow a machine to generalize [15]. Machine learning proposes to solve a variety of tasks such as classification, regression, clustering, and association, among many more [36].

A task is an abstract representation of a problem we want to solve regarding a set of domain objects. These objects constitute examples or observations in a data set, each of them described by the same set of features or predictors. Most tasks can be represented as a mapping from these data points to one or more outputs represented by target variables. In these cases, the goal of machine learning algorithms is to build a model that approximates, as close as possible, this mapping function using a so-called training data set. The resulting model should, in principle, effectively capture patterns, such as relationships between examples and features, that would allow it to generalize

the concept underlying the data that enable accurate predictions to be made over future unseen examples [36].

There are different types of machine learning algorithms. Regarding the learning process, some algorithms are designed to learn tasks in a supervised fashion requiring labelled training and test data such as mentioned above. Other algorithms are devoted to solving unsupervised learning tasks, where that data is unlabelled, meaning that there is no output variable [36]. For the first type of task, the result is a predictive model that predicts the target variable, while for the second type of task the result is a descriptive model that identifies interesting structures in data [36]. Nonetheless, regardless of the task or algorithm, exploring the data is a crucial step toward data cleaning and feature extraction, and engineering operations, which are the basis of successful machine learning techniques [36].

Classification is a supervised task where the goal is to learn a model to predict a target variable, which domain is discrete. In our study, we have a binary classification task, as the target variable has only two possible values, designed as the positive and negative class (e.g., medical diagnosis, where the positive class is having a particular disease) [36].

To assess the performance of models in binary classification tasks, a two-by-two contingency table build on true and predicted classes, is used to derive a wide range of performance indicators. Correctly classified positives and negatives are referred to as true positives and true negatives, respectively. Incorrectly classified positives are called false negatives; misclassified negatives are called false positives [36]. Accuracy as the proportion of correct predictions, regardless of the class, is the most used metric. It is often necessary to distinguish performance in the classes [36]. The true positive rate (sensitivity or recall) is the proportion of positives correctly classified [36]. The true negative rate (specificity) is the proportion of negatives correctly classified [36]. These rates represent per-class accuracies.

2.3.1. Related Work in Multimodal Classification Tasks

Several machine learning algorithms have been used to classify anxiety and stress relying on the unique characteristics of its physiological mechanisms. Often self-report results are used as part of the labels which makes the participant in charge of the labelling process. One of the advantages of self-reports is that they can be easily applied although this method carries individual bias [15]. Studies have labelled anxiety in multiple

ways: binary (stressed or relaxed), three classes (no anxiety, low anxiety, and high anxiety), or four classes (no, low, mild, or high anxiety) [22].

Some of the classification algorithms that have been used for anxiety and affect detection are Decision Tree (DT) Adaptive Boosting (ADB), Support Vector Machines (SVM), Naïve Bayes (NB), Markov Models (MM), Logistic Regression (LR), K-Nearest Neighbours (KNN), Random Forest (RF), Artificial Neural Networks (ANN), Linear Discriminant Analysis (LDA), Hidden Markov Models (HMM), etc. [19], [22]. Using neural networks, trained on audio and/or visual data, has achieved high-performance results but these models are quite demanding in terms of computational resources [1]. Some machine learning studies have been conducted in classifying distinct levels of anxiety and affect. Combining different physiological signals has been shown to increase the accuracy of anxiety and affect detection systems [18].

In classifying anxiety, the EDA signal was used alone to detect and measure stress in call centre employees at work and during each call. In a binary classification task - stressful/non-stressful calls - SVMs achieved an accuracy across participants of 78.03% when trained and tested on different days from the same person, and 73.41% when trained and tested on different people using the proposed adaptations to SVMs [34].

Accuracies higher than 95%, were obtained in binary and multi-class classification of anxiety levels, in either subject-dependent or independent studies, using Linear Discriminant Analysis (LDA) with a combination of physiological signals such as GRS, EMG, ECG, and ST as well as respiration data, with as many as three different physiological signals at a time [22].

For classifying anxiety, a combination of BVP, GSR, and skin temperature (ST) was used with an SVM classifier to classify four different anxiety classes ("low", "mild", "moderate", and "high"), and achieved an accuracy of 80,1% [22]. In this research study [22], the feature selection was addressed by using Random Forest (RF), a supervised feature selection method [36] – meaning it considers the relationship of the features with the target variable when selecting the optimal features. Some examples of other supervised feature selection algorithms that can also be found in the literature [37] in studies related to stress and anxiety detection are Sequential Forward Search (SFS), Sequential Backward Search (SBS), and Correlation-based feature extraction (CBFS).

In a study [38], a system for feature extraction and machine learning classification was developed to detect and classify negative affect. This study used an imbalanced dataset. Physiological data - heart activity, skin conductance, and accelerometer signals - was collected from 21 participants of an algorithmic programming contest for nine days to discriminate contest stress, relatively higher cognitive load (lecture), and relaxed time activities. For dimension reduction, Principal Component Analysis (PCA) was used to attain a lower dimensional space. This study used an imbalanced dataset. To deal with the imbalance problem, a combined strategy of undersampling the majority class and oversampling the minority class was used. When combined with ECG and EDA signals, a multilayer perceptron algorithm achieved the best classification accuracy of 92.15%.

The classification algorithms used in this study were chosen based on literature, which has reported good results, on their efficacy, suitability for binary classification problems, and the characteristics of the data used. Furthermore, the goal was to evaluate different machine learning algorithms that are easily reproducible, manually optimized, and do not take much computational power. This motivated the following choices.

Logistic Regression (LR), a linear algorithm that predicts the output of categorical dependent variables, is normally used in the classification of binary problems. LR is defined as a discriminative probabilistic model. It models the posterior probability distribution – $P(Y|X)$, where Y is the target variable and X is the set of known features. That is, given X it returns a probability distribution over Y , modelling a decision boundary that linearly separates classes by focusing on where classes overlap [36].

Linear Discriminant Analysis (LDA) is also a linear classification algorithm that models a decision boundary, which is generated by fitting class conditional densities to the data and using Bayes' rule. Simplifying, it develops a probabilistic model per class based on the distribution of observations for each input variable. An example is then classified by calculating the conditional probability of belonging to each class and selecting the class with the highest probability. LDA assumes that the input variables are numeric, normally distributed, and have equal variance. Nonetheless, if all these conditions are not completely met, the model can be robust to those violations [36].

The Decision Tree (DT) algorithm induces a tree model that can be used for regression or classification and consists of a structured sequence of rules composed of conditions over the variables. The answers determine what is next condition to be assessed if any should be. The result is a set of rules that are represented in a tree-like

structure, where the output is given by the values at the terminal nodes, i.e., the leaves of the tree. The main elements of any decision tree algorithm are the splitting criterion at a node based on the value of one variable; the stopping criterion for deciding when a node is terminal and thus should not be split any further, and a prediction for the target variable in each terminal node [36].

The Support Vector Machines (SVM) algorithm's objective is to find a hyperplane in K-dimensional space, being K the number of features, that distinctly classifies the data points. The central idea underlying this method is to perform a mapping from the input space into a (usually) very high dimensional feature space where the data become easier to be separated. According to the type of mapping that is performed for data separation, several kernels can be used [36].

Three different tree ensemble algorithms were also chosen for this study: Random Forest (RF) Adaptive Boosting (ADB) and Extreme Gradient Boosting (XGB). These are machine learning algorithms based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and improve the performance of the model.

Random Forest (RF) builds several decision trees on various subsets of the given dataset. Instead of relying on one decision tree, the RF takes the prediction from each tree, and based on the majority votes of predictions it predicts the final output. A greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting [36].

The Adaptive Boosting (ADB) algorithm is an ensemble learning algorithm that uses the boosting technique. Boosting is a method used to reduce bias and variance for supervised learning. It works on the principle that learners are grown sequentially. Except for the first, each subsequent learner is grown from previously grown learners on the assumption that better learners will be progressively added to the ensemble. In that sense, several weak learners are converted into a strong learner. ADB typically uses DTs as weak learners. It is called Adaptive Boosting because the weights are re-assigned to each instance, with higher weights to incorrectly classified instances [36].

Finally, Extreme Gradient Boosting (XGB) is an implementation of Gradient boosting machines (GBM), that provides parallel tree boosting. The main objective of Gradient Boosting is to minimize the loss function by adding weak learners, using a

gradient descent optimization algorithm. XGB is an extension of gradient-boosted decision trees, designed to be highly efficient and flexible [39].

2.3.2. Imbalanced Classification

In most collected datasets, severe states of affect, stress, and anxiety are less frequent, especially in non-clinical populations [18], [19]. This poses challenges to classification since standard classification models aim to maximize their performance across all the observations and, thus, neglect their performance on the rarer ones, which might be the most important ones from the domain perspective [16].

Imbalanced domain data poses several challenges to standard machine learning algorithms. The problem is related to the fact that the distribution of the target variable is not balanced and the most important cases from the domain perspective are poorly represented in the available data [16].

Standard learning algorithms are built on preference criteria that aim to minimize the overall error over the domain of the target variable. To that be attained, the algorithm will build a model that is more accurate in the most common cases, neglecting the performance of the rare but of utmost importance cases [40].

Several strategies have been developed to address the between-class problem, mainly in the classification setting and, in particular, for binary classification [16]. The many solutions that appear in the literature can be grouped into four types of strategies: special-purpose learning methods, which imply modifications to the learning algorithms; data pre-processing, which comprises changes in the data before the learning process; prediction post-processing, where transformations are applied to the predictions of the learned models; and hybrid methods that combine distinct types of strategies [16].

In this study, we will use data pre-processing approaches that change the data distribution before learning algorithms. This means that instead of applying a learning algorithm directly to the provided training sample, we will first pre-process this data following our goals. Any standard learning algorithm can then be applied to the pre-processed dataset [16].

We will also use performance metrics that are suitable for this imbalanced binary classification task. We will detail these pre-processing techniques in Section 2.3.3 and the suitable evaluation metrics in Section 2.3.4.

2.3.3. Data Pre-processing Techniques

Data pre-processing techniques can be grouped into two main types: distribution change, which changes the data distribution to address the issue of the poor representativeness of the more relevant cases; and the weighting of the data space, which modifies the training set distribution using information concerning misclassification costs, such that the learned model avoids costly errors [16], [41]. For binary classification problems, changing the class distribution of the training data may improve classifiers' performance in an imbalanced context because there is a connection with non-uniform misclassification costs [16].

The approaches for changing the data distribution can be aggregated into sampling, synthesizing new data, or combinations of the previous methods. A Sampling approach includes strategies that remove and/or add examples to the original dataset. These are based on a diverse set of techniques, such as Random under/oversampling, distance methods, data cleaning approaches, and clustering algorithms. Approaches that synthesize new data involve the generation of new artificially generated examples that are added to the original dataset [16].

In this dissertation, and because we don't have a big number of examples, the methods tested will oversample the minority class and these techniques will include: Random oversampling (Over) and synthetic data generation techniques [16]. Some important drawbacks of sampling methods are that some sampling approaches only work on between-class imbalance and not within-class imbalance [40]. Between-class imbalance is the imbalance occurring only due to the minority class, and within-class imbalance refers to rare cases present within the minority or the majority class [40].

Furthermore, changing the data distribution may not be as easy as expected, since deciding the optimal distribution for the user preference biases is not straightforward. A frequently used approach consists of trying to perfectly balance the data distribution, but, for some classifiers, this distribution does not always provide optimal results [16].

Sampling involves creating a new transformed version of the training dataset, in which the selected examples have a different class distribution. The simplest strategy is to choose examples for the transformed dataset randomly. There are two main approaches to random resampling for imbalanced classification which are: Random oversampling and Random under-sampling. For the reasons above mentioned, only Random oversampling will be considered for this study. Random oversampling involves

randomly selecting examples from the minority class, with replacements, and adding them to the training dataset [36]. However, this method may increase the likelihood of overfitting, especially for higher over-sampling rates [16].

Another important approach for dealing with the imbalance problem is the generation of new synthetic data. Several methods exist for building new synthetic examples that carry several advantages, namely: reducing the risk of overfitting and improving the ability of generalization, which may be compromised by the random over-sampling methods [16]. Inside these different methods, there are random methods like the Synthetic Minority Oversampling Technique (SMOTE), which randomly chooses examples of the minority class on which it creates synthetic examples, and informed methods, where the location of the examples in the input space is considered for that purpose. Informed methods, such as ADASYN (ADA) and Borderline SMOTE derive from SMOTE (SMO) and direct their efforts to critical areas of the input space, for instance, sparse areas, safe areas, or areas close to the decision boundary, and consequently may avoid the generation of noise and tackle imbalances within classes [41].

SMOTE (SMO) was proposed to improve random oversampling. A random example from the minority class is first chosen, and then k of the nearest neighbours for that example is found [41]. A randomly selected neighbour is chosen, and a new synthetic example is created by interpolating the values of features of the two examples [42]. The SMOTE has been applied with several different classifiers and integrated with boosting and bagging. Nevertheless, SMOTE generates synthetic examples with the positive class label, disregarding the negative class examples. This strategy may be especially problematic in the case of highly skewed class distributions where the minority class examples are very sparse, thus resulting in a greater chance of class mixture. Another major concern is that SMOTE may further amplify the noise present in the data since the method does not distinguish overlapping class regions from so-called safe areas [41].

ADASYN (ADA) is a version of SMOTE that works similarly but with one important difference. After creating the synthetic samples as described in “normal” SMOTE, the ADASYN version uses a density distribution as a criterion to automatically decide the number of synthetic samples that need to be generated for each minority data example. If the ratio between the number of minority examples versus the number of majority examples is less or equal to a defined balance threshold, ADASYN synthesizes new

minority instances. In the n neighbours of a minority example (which form a group named a neighbourhood), the lower this ratio, the more instances of the minority class are added to each neighbourhood, until the balance threshold defined is reached. The resulting dataset will not only provide a balanced representation of the data distribution but will also force the learning algorithm to focus on those difficult-to-learn examples [43].

Borderline SMOTE is also a version of SMOTE, but unlike SMOTE, only makes synthetic data along the decision boundary between the two classes. If half of the defined m neighbours belong to the majority class, that minority example is considered in danger and part of a decision boundary and Borderline SMOTE will act on it. There are two kinds of Borderline SMOTE: Borderline SMOTE 1 and Borderline SMOTE 2 (Border). Borderline SMOTE 1 over-samples the minority classes in this critical area using only its minority class neighbours (being more like SMOTE in this sense) [44].

Borderline SMOTE 2 (Border), used in this study, not only generates synthetic examples using each minority example and its positive nearest neighbours but also does that from each minority example and its nearest negative neighbour. The difference between the positive example and its nearest negative neighbour is multiplied by a random number between 0 and 0.5, thus the newly generated examples are closer to the minority examples. This aims to strengthen the decision border area [44]. Fig. 2-4 shows the different SMOTE variants and their application to our dataset.

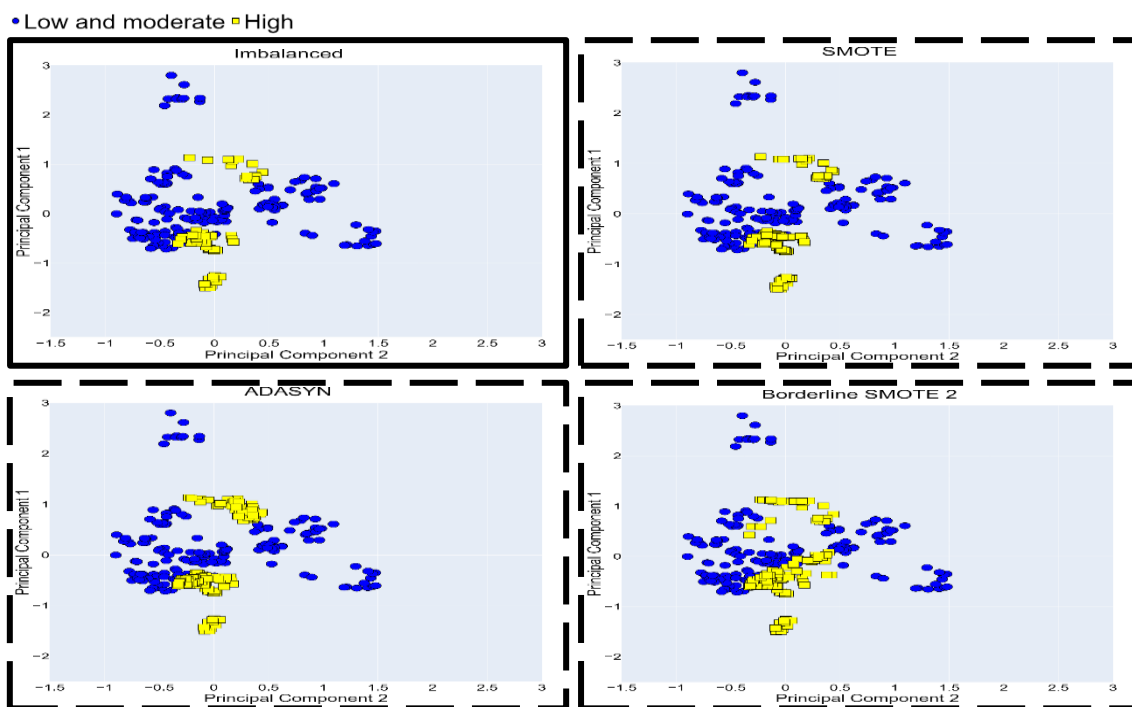


Fig. 2-4 - PCA analysis of the SMOTE variants applied to our data without supervised feature selection.

2.3.4. Suitable Performance Metrics

The lack of representativeness of the important class (the positive class in binary classification tasks) causes standard evaluation metrics (e.g. accuracy, error rate) to be misleading, as they are biased toward average performance and will not correctly assess the performance of the models on these rare events [16]. In imbalanced domains, several solutions have been proposed to overcome the inadequacy of more traditional evaluation metrics, as the choice of appropriate performance metrics is of key importance [16]. Adequate metrics should not only provide means to compare the models according to the user preferences but, ultimately, also drive the learning of these models [16].

Accuracy evaluates the fraction of predictions the model got right. For binary classification, accuracy is calculated in terms of positives and negatives. The positive class is the class of the user's preference [36]. In this case, if the positive class is the minority class, there is a bias toward the majority class and examples that are considered less relevant to be correctly classified. Thus, in imbalanced domains, accuracy is not the best measure of model performance since the least represented will have a low impact [16].

The metrics used in imbalanced domains must consider the domain preferences and the data distribution. Among the several performance measures: precision and recall are two metrics that assess the performance of the classifier concerning the positive class. Precision is defined as the fraction of relevant instances among all retrieved instances, representing the fraction of true positives among those classified as positives. Recall is the fraction of retrieved instances among all relevant instances. It is also called the true positive rate (TPR) i.e., the proportion of positives that are correctly classified as such. A perfect classifier has both precision and recall equal to 1 [16], [36].

It is often possible to calibrate a model's results by improving one at the expense of the other [16], [36]. As these two measures exhibit a trade-off, it may be impractical to simultaneously monitor them. To cope with those situations combined measures, such as the $F\beta$, or the Area under the ROC curve (Roc-Auc) were proposed.

The $F\beta$ can be interpreted as a harmonic mean of precision and recall, where $F\beta$ reaches its best value at 1 and worst at 0. β is a coefficient to adjust the relative importance of recall concerning precision (if $\beta = 1$ precision and recall have equal weight, if $\beta > 1$ the weight of recall will increase, while if $\beta < 1$ more importance to precision will be given). The $F\beta$ is more informative in imbalanced settings than accuracy. This metric

value is high when both recall, and precision are high. The most common setting is to use $\beta = 1$ to obtain the F1-score (F1), also known as the harmonic mean between precision and recall, and calculated as follows: $F1 = (2 \times (Precision \times Recall)) / (Precision + Recall)$ [16].

The Area under the ROC Curve (Roc-Auc) is a popular performance measure also used in the context of imbalanced domains since it is not biased towards the minority class [16]. Roc-Auc measures the area underneath the receiver operating characteristic curve (ROC curve). The ROC curve is a graph showing the performance of a classification model at all classification thresholds, with two components: the true positive rate (TPR), and the false positive rate (FPR). The FPR is defined by the ratio between false positives and the sum of false positives and true negatives. Roc-Auc ranges in value from 0 to 1. A model whose all predictions are wrong has a Roc-Auc of 0 whereas a model whose predictions are all correct has a Roc-Auc of 1 [40]. The Roc-Auc measures the probability of the classifier assigning a higher rank to a randomly chosen positive example than a randomly chosen negative example [15].

It is important to point out that relying on the discussion of results on a single metric to evaluate performance should be avoided since using several different metrics at the same time gives a better overview of the real performance and robustness of a classifier or prediction approach.

2.4. Dissertation Summary

Summing up, to approach the problem presented above – characterize and predict different levels of anxiety and affect – we used a combination of different physiological signals instead of only one at a time, collected in a neutral condition. This differentiates our study from previous ones since we are mixing distinct signals collected in an environment where we are not stimulating anxiety or affect. For the classification tasks, we addressed them as three separate imbalanced binary classification tasks – anxiety classification, negative affect classification, and positive affect classification - using different classifiers. Approaching feature selection, we decided to apply a combination of unsupervised methods as the correlation threshold, and supervised methods – a filter method or a wrapper method - to determine the best features to deal with the detailed problem. Additionally, we deal with the data imbalance using data-level techniques, both informed and random techniques, which change the data distribution by oversampling the minority class to achieve perfectly balanced classes.

3. Data and Methods

This chapter will describe the data used for this dissertation as well as the methodology employed during the workflow chain.

3.1. Dataset

To achieve the proposed goals, we used a publicly available multimodal dataset for stress and affect detection – the WESAD dataset – containing self-reports, motion, and physiological data [1]. The data were collected under a study protocol in which the subjects were exposed to different affective stimuli (neutral, stress, and amusement) and two meditation periods (to de-excite the participants).

Fig. 3-1 describes the protocol. There are two versions of the protocol – the version presented in Fig. 3-1 describes the one in which the stress condition was conducted right after the neutral condition and before the amusement condition. In the second version of the protocol, in which half of the participants participated, the amusement condition comes right after the neutral condition and before the stress condition. More details on the protocol can be found in the article [1].

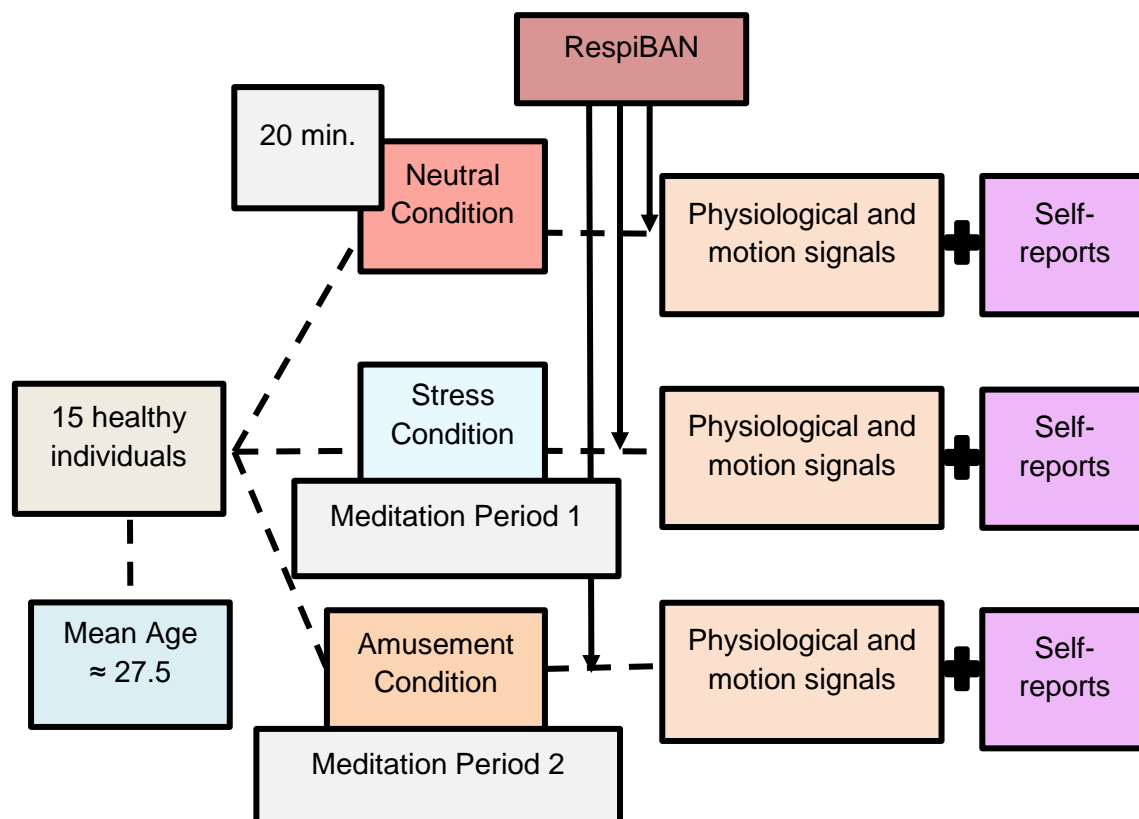


Fig. 3-1 - Schematic representation of the WESAD dataset protocol.

The dataset contains high-resolution data, sampled at 700 Hz from a chest-worn device (RespiBAN) and a lower-resolution wrist-worn device (Empatica E43). These devices are multimodal sensors that allow obtaining the following physiological signals: ECG, EDA, EMG, respiratory activity (RESP), temperature (TEMP), and motion accelerometer (ACC). The two sensors were synchronized, and the data was labelled with one of the three conditions: neutral, stress, or amusement. The dataset also includes context notes about the participants and self-report results.

In this study, the EMG data used was recorded on the upper trapezius muscle on both sides of the spine, and the skin over the muscle. The EDA signal used was recorded on the rectus abdominis. The three physiological signals - ECG, EDA, and EMG - were filtered to remove the noise and isolate frequencies of interest. From these signals were extracted and selected features.

From the seventeen initial subjects that participated in the study, two subjects were excluded due to sensor malfunction, resulting in a total of 15 healthy subjects (12 males). The mean age of the subjects is around 27.5 years, ranging from 25 to 29 years old.

Self-reports were collected from each participant after each defined condition (neutral, amusement, stress, meditation 1, meditation 2). Of those answered after the neutral condition, two of them were used: the 6-STAI questionnaire (questionnaire's items are presented in Table 6-1), to assess the current anxiety level of the participants, and the PANAS questionnaire (questionnaire's items are shown in Table 6-1), to determine the negative and positive momentary affect valences of the participants.

As this study aims to address the characterization and classification of address anxiety and affect states without any controlled elicitation for those, nor any pre-designed affective stimuli, to ensure a neutral affective state, only the data collected during the 20-minute neutral condition was used. The selection of the physiological signals for anxiety and affect is supported by the literature [18], [19], those that are described as reliable indicators of anxiety and affect.

3.2. Methodology

This research follows a data processing chain, consisting of the following steps: pre-processing of the physiological signals, segmentation, feature extraction, feature selection, classification, and classification performance evaluation. Fig. 3-2 illustrates these different steps, further detailed in this Section.

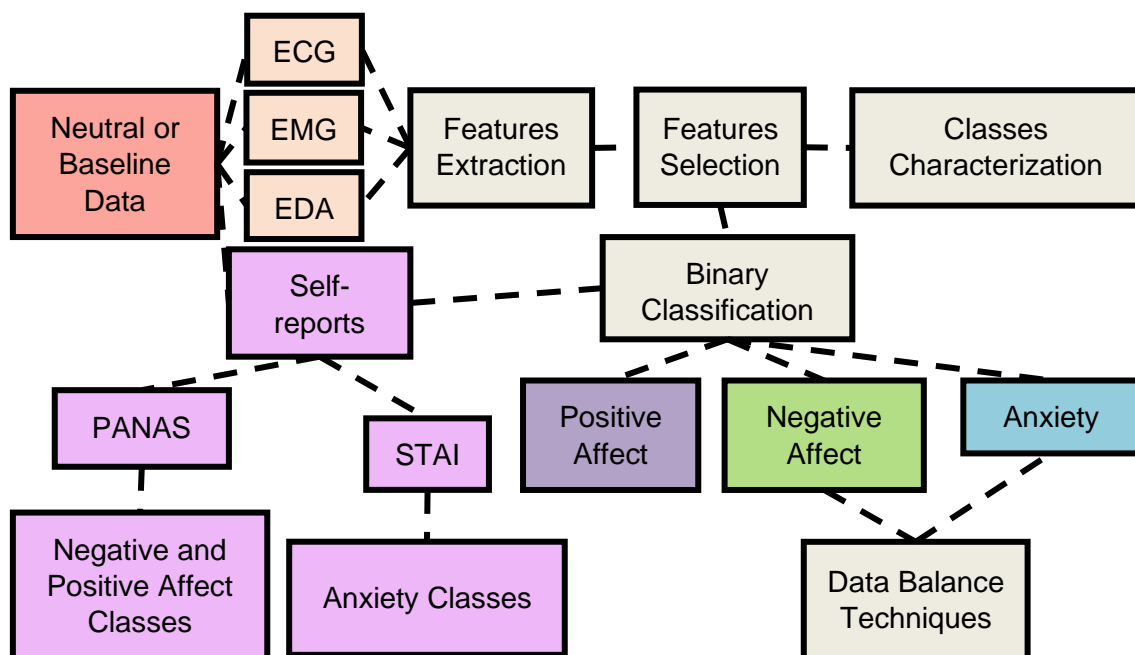


Fig. 3-2 - Graphical representation of the study's workflow.

3.2.1. Raw Signals Filtering

After extracting the data from the neutral condition, all signals were shortened from 20 minutes to a total of 19 minutes to ensure the same length. A common step in pre-processing is to apply filters, to improve overall signal quality by removing different noises and artefacts.

Besides the powerline interference (around 50Hz), different noises affect the ECG, EMG, and EDA signals:

ECG: Mainly two types of noises are present in the ECG signal. High-frequency noises include electromyogram noise and additive white Gaussian noise, and low-frequency noise, including baseline wandering, due to movement and respiration (normally around 0.5 Hz) [35].

EDA: It shows expected skin conductance morphology at a low amplitude, therefore is mostly affected by the noise caused by electrical interference, which can manifest in the data as a small amount of fuzz on top of the signal [45].

EMG: Common artefacts of low frequency occur when the EMG signal cables move due to the subject's motion or incorrectly attached electrodes [35]. Also, other muscles in the body can generate an EMG signal – if they are close to the testing site, crosstalk can be recorded. EMG recordings close to the subject's heart may detect the

subject's pulse as a regular beat underlying the EMG signal that is being collected for investigation.

The filtering of the physiological signals was performed after power spectral density (PSD) analysis, allowing us to determine which frequencies contain the most information. The appropriate filters and their parameters were selected by analysing the mean absolute error (MAE), the relative error (RE), the coefficient of variation (CV), and the signal-to-noise ratio (SNR).

1. ECG Signal: The baseline wander, although not significant, was present in all participants. Thus, a Butterworth bandpass filter was applied between 0.5-40Hz.
2. EDA Signal: The EDA signal was filtered using a 5Hz lowpass filter.
3. EMG Signal: The powerline interference was removed in the 50Hz range, and a bandpass filter was applied between 20-200Hz.

3.2.2. Class Assignment

Based on the self-report results on the participant's anxiety, and positive and negative affect levels, we distinguish three binary classification tasks:

- 1) The first is a classification of anxiety levels in which we distinguished two classes: "Low and moderate" anxiety vs "High" anxiety.
- 2) The second classification task is a problem to classify negative affect levels defined as "Low and moderate" negative affect vs "High" negative affect.
- 3) The third one is the classification of the positive affect in which a problem was defined as "Low" positive affect vs "Moderate and high" positive affect.

Through the self-report results, we used thresholds to assign the classes of anxiety and affect. These classes were then used as labels in the classification task and the features obtained from the physiological signals served as predictors to classify different states of anxiety and affect.

When analysing all self-report results, in Fig. 3-3, Fig. 3-4 and Fig. 3-5, the distribution of classes present in each dataset is not balanced, since the anxiety and affect classes are not equally represented. Regarding positive affect, less elevated levels of PA are more frequent in the dataset and represent the most important class to be correctly classified. However, the more elevated levels of anxiety and negative affect are less represented in the dataset, even though they are more important to detect and

classify correctly. Therefore, as described in the literature, balance techniques were implemented to improve the classification performance of anxiety and negative affect.

3.2.2.1. Anxiety Class division

Considerable evidence attests to the construct and concurrent validity of the scale. Short forms were highly correlated with the 20-item STAI score, and all internal consistency reliabilities were greater than 0.90. Commonly, the 20 items of the STAI S or T questionnaire divide individuals into 3 levels of anxiety: No/Low anxiety (1 – 11.1), moderate anxiety (11.4 - 13.2), and high anxiety (13.5 - 24) [46]. The cut points described in the literature reflect the contexts in which the studies are conducted. Normative values are available for adults, college students, and psychiatric samples [26]. Based on the cut points described in the literature as common, we defined the following classes: “Low and moderate” anxiety (6 - 13) and “High” anxiety (14 - 24). The median value of our scores is 12 and the mean is 12.567 with a standard deviation (SD) of ± 1.726 . The scores ranged from 10 to 16.

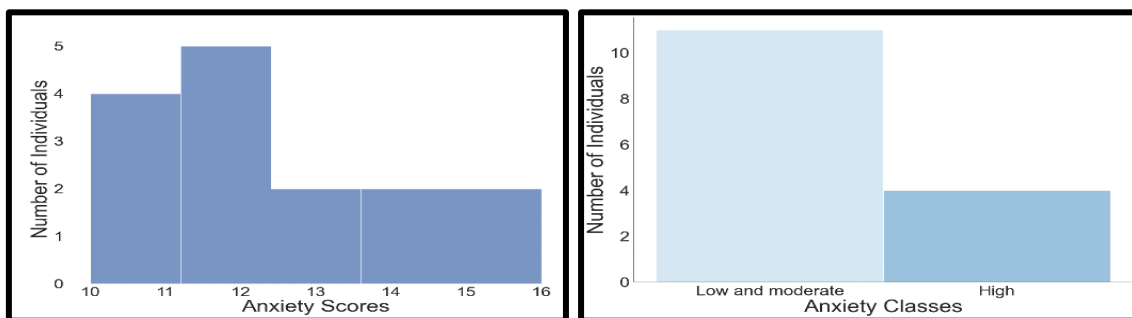


Fig. 3-3 - On the right, the bar plot shows the number of individuals per anxiety class. On the left, the bar plot shows the number of participants for anxiety measured in the STAI questionnaire.

3.2.2.2. Negative and Positive affect class division

The mean score for the development sample of the PANAS (moment) was 29.7 (SD = 7.9) for positive affect and 14.8 (SD = 6.2) for negative affect [30].

The mean score of the negative affect dimension of our PANAS is 12,667 (± 2.086) which shows to be lower than the normative mean scores and lower than 1 SD interval from the normative mean established for the PANAS momentary scale [30]. The median value is 12. The scores ranged from 10 to 16. This indicates that our sample is less negative than the normative sample. The normative development values were used to divide the negative affect scale into two classes: “Low and moderate” negative affect (10 - 14.8); “High” negative affect (14.9 - 50).

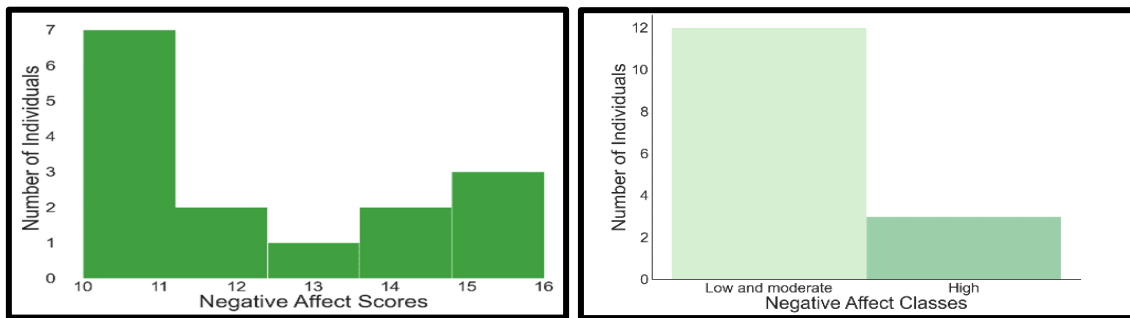


Fig. 3-4 - On the right, the bar plot shows the number of individuals per negative affect class. On the left, the bar plot shows the number of participants with negative affect scores measured in the PANAS questionnaire.

The mean score for positive affect is 25.467 (\pm 6.209), which is also lower than the normative mean scores [30] but within 1 SD from the mean. The median value is 26. The scores ranged from 16 to 37. We can see that there are more values under the normative mean than equal or over the normative mean, which indicates that our sample is more positive than the normative sample. The normative development values were used to divide the positive affect scale into two classes: “Low” positive affect (10 - 29.6); “Moderate and high” positive affect (29.7 - 50).

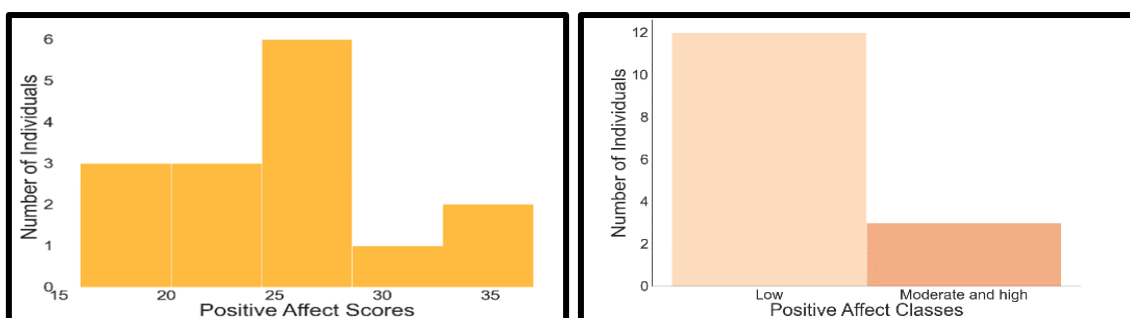


Fig. 3-5 - On the right, the bar plot shows the number of individuals per positive affect class. On the left, the bar plot shows the number of participants with positive affect scores measured in the PANAS questionnaire.

3.2.3. Feature Extraction

An important part of creating a model from physiological signals is feature extraction. Thus, after filtering the raw signals, features in time and/or frequency domains, as well as non-linear features, were extracted from each physiological signal. To extract the features from ECG, EDA, and EMG, the “Neurokit2” python package [47] was used.

After this step, to augment the number of observations of each feature for the classification task, the features were extracted after dividing the signal into equal time segments. It is important to select a suitable signal window size, specifically because some features of HR and HRV have a minimal time window required to be collected.

Each physiological signal was divided into short-term segments with a 5-minute time length and a 4-minute overlapping. 15-time segments were obtained for each subject, and for each segment, the features described in Table 3-1 were computed.

From the ECG-filtered signal, peak detection algorithms were used to find the heart beats, and peak features (peak interval standard deviation, peak mean, and median interval) were computed. Statistical features from the ECG-filtered signal were also calculated. Through the detected peaks, the HR and corresponding statistical features were computed. From the location of the heart beats, the HRV was derived as its corresponding statistical features.

HRV time-domain indices quantify the HRV observed during monitoring periods that range from around 1 min to more than 24h. Generally, resting values obtained from short-term monitoring periods correlate poorly with 24h indices. Thus, it is important to select HR and HRV features that can be extracted with validity in a 5-minute time window (short-time measures) or less (ultra-short-term measures) [48]. From the ultra-short-term features, the ones extracted were the ones that also showed the validity of estimation and correlation with the 5 minutes measures [48].

For short-term data, the time domain measures mean and median N-N intervals (MeanNN and MedianNN), N-N meaning the interval between normalized R peaks (N) intervals, the Standard deviation of the N-N intervals (SDNN), square root of the mean of the sum of successive differences between adjacent R-R intervals (RMSSD), and proportion of N-N intervals greater than 50ms or 20ms (pNN50 or pNN20) out of the total of RR intervals, and the frequency domain measures total power, very low-frequency HRV (VLF), low-frequency HRV (LF), high-frequency HRV (HF), and the ratio between low-frequency HRV and high-frequency HRV ratio (LF/HF) can be used. Their normalized and log-transformed versions are also considered appropriate [48].

Some ultra-short-term (< 5 min) studies mentioned Approximate entropy (ApEn), Sample entropy (SampEn), triangular interpolation of N-N interval histogram (TINN), the standard deviation of the successive differences between N-N intervals (SDSD), median absolute deviation of the N-N intervals (MadNN), SDNN divided by the MeanNN (CVNN), RMSSD divided by the MeanNN (CVSD), and MadNN divided by MedianNN (MNVN) were suitable for short-term measures of HRV [48]. The detection of ECG peaks, using the function “events_plot” for the R peaks and “ecg_delineate” for the P, Q, S, and T peaks, from the “Neurokit2” python package [47], is shown in Fig. 3-6.

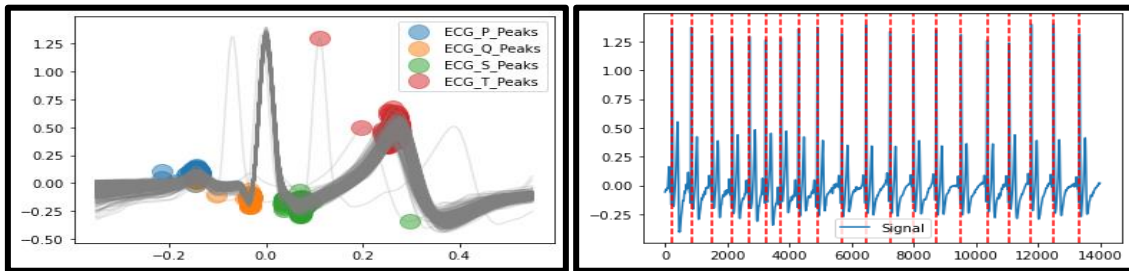


Fig. 3-6 - Plot of the ECG signal peaks: P, Q, S, T (on the left) and R (on the right).

From the EDA filtered signal, statistical features were computed. Furthermore, the two components of the EDA signal - SCL and SCR were separated, and additional statistics from both components and peak features from the SCR were extracted. Fig. 3-7 represents visually some of the features extracted from the EDA signal through the plots created using the function “plot” from the “matplotlib” python package [49] and the function “events_plot” from the “Neurokit2” python package [47].

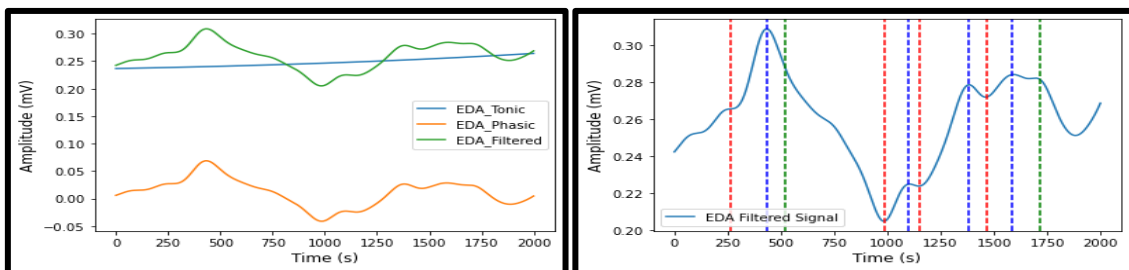


Fig. 3-7 - On the left, the EDA filtered signal is presented in blue. The red line is - the SCR onset; the blue line is - the SCR peak; the green line is - the SCR recovery point. On the right, the figure shows the EDA filtered signal and the Tonic and Phasic components.

Concerning the EMG-filtered signal, statistical and peak features were computed to assess muscle activity. The EMG linear envelope was also extracted, and its statistical and peak features were computed [34]. The EMG linear envelope consists of the smoothed EMG signal obtained through the low pass filtering of the full wave rectified signal. In Fig. 3-8, the activation peaks extracted from the EMG envelope are represented in the bottom plot, generated using the function “emg_plot” from the “Neurokit2” python package [47].

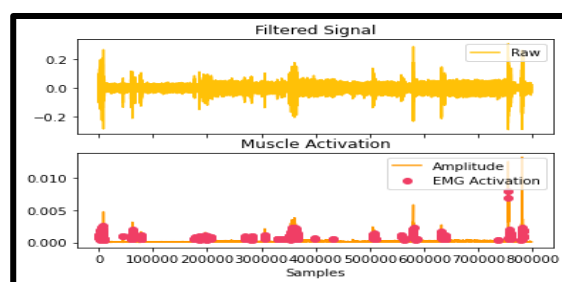


Fig. 3-8 - Plot of the EMG filtered signal (in yellow) and the activation peaks marked with red dots.

ECG	Minimum and Maximum ECG (Max/Min ECG), Mean and Median ECG (Mean/Med ECG), Standard deviation and Variance ECG (SD/Var ECG), Dynamic range ECG (Range ECG), Mean Area under the Cardiac Cycles (AUCC)
	Minimum and Maximum HR (Min/Max HR), Mean and Median HR (Mean/Med HR), Standard deviation and Variance HR (SD/Var HR), Dynamic range HR (Range HR)
	RMSSD, MeanNN, CVNN, MadNN, CVSD, MCVNN, SDNN, SDSD, MedianNN, pNN50/20, TINN, ULF, VLF, LF, HF, VHF, LFHF, LFn, HFn, LnHF, ApEn, SampEn
	Q, R, S, T, P peaks interval - Standard deviation (SDQQ, SDRR, SDSS, SDTT, SDPP), Mean (MeanQQ, MeanRR, MeanSS, MeanTT, MeanPP) and Median (MedQQ, MedRR, MedSS, MedTT, MedPP)
EDA	Minimum and Maximum EDA (Min/Max EDA), Mean and Median EDA (Mean/Med EDA), Standard deviation and Variance EDA (SD/Var EDA), Mean and Median Rise/Recovery time (Mean/Med Rise/Recovery Time), Dynamic range EDA (Range EDA)
	Minimum and Maximum SCR (Min/Max SCR), Mean and Median SCR (Mean/Med SCR), Standard deviation and Variance SCR (SD/Var SCR), Dynamic range SCR (Range SCR), Autocorrelation SCR (Corr SCR)
	Number of SCR peaks per second (Nr SCR Peaks/s), Mean and Median SCR Peaks/Recovery/Onsets Amplitude (Mean/Med SCR Peaks/Recovery/Onsets Amp)
	Minimum and Maximum SCL (Min/Max SCL), Mean and Median SCL (Mean/Med SCL), Standard deviation and Variance SCL (SD/Var SCL), Dynamic range SCL (Range SCL), Autocorrelation SCL (Corr SCL)
EMG	Minimum and Maximum EMG (Min/Max EMG), Mean and Median EMG (Mean/Med EMG), Standard deviation and Variance EMG (SD/Var EMG), Dynamic range EMG (Range EMG), Percentile 10 and 90 EMG (Perc 10/90 EMG)
	Minimum and Maximum Envelope (Min/Max Env), Mean and Median Envelope (Mean/Med Env), Standard deviation and Variance Envelope (SD/Var Env), Dynamic range Envelope (Range Env), Percentile 10 and 90 Envelope (Perc 10/90 Env)
	Number of Activation Peaks/Pulse Onsets per second (Nr Activ Peaks/Pulse Onsets/s), Mean and Median Activation Peaks/Pulse Onsets Amplitude (Mean/Med Activ Peaks/Pulse Onsets Amp)

Table 3-1 Extracted features from the filtered physiological signals.

3.2.4. Feature Analysis and Selection

First, an evaluation of the 109 features (each with 15 observations, one observation per subject) described in Table 3-1 was performed, and some statistics were computed. To understand if there are features with similar distribution, thus the mean value, the standard deviation and variance, the 25, 50, and 75 percentiles, and the

number of missing values were analyzed. Features VLF and ULF did not have any observations and were eliminated. The boxplots of the features are presented in Fig. 6-1 to Fig. 6-10.

Feature selection is a crucial step before processing any high-dimensional data for clustering or classification and can be classified as supervised or unsupervised depending on whether it considers the target variable or not. The main objective is to reduce the data dimensionality by removing redundant or non-discriminatory features. This helps in the efficient execution of the classification step. Notwithstanding, not all features are relevant, leading to the need of determining the ones that best represent the analyzed conditions – the two anxiety, positive, and negative affect classes.

After the data augmentation (15 observations per subject), the feature selection consisted of three sequential steps: a) missing values and variance analysis using the function “map” from the python package “Statistics” [50]; b) unsupervised feature selection - correlation threshold evaluation using function “corr” from the python package “SciPy” [51]; c) supervised feature selection - a filter or wrapper method using functions “SelectKBest” and “RFECV” from the python package “Sklearn” [52].

No feature showed zero variance or missing values. After this step, all features were normalized using the Min-Max normalization method described in Section 3.2.8.

Concerning unsupervised methods for feature selection, the correlation between features was analyzed. This step aims to remove redundant variables that can cause bias in classification. Redundancy or collinearity may indicate very correlated processes or just different ways of representing the same thing.

The normality of the data was tested using a Shapiro-Wilk (SW) test. With a 95% interval confidence, the features LnHF and Corr SCR showed to follow a normal distribution. Thus, since not all variables followed a normal distribution, we used the Spearman correlation coefficient and checked with the Pearson coefficient for the normal ones. The correlation coefficients that showed to be significant (with a 0.01 and 0.05 significance level) were analyzed.

The correlation coefficient threshold method – with a threshold of 0.90 – was used in which between two highly correlated features, the one with the higher variance is selected. Fig. 3-9 to Fig. 3-20 present the correlation between features.

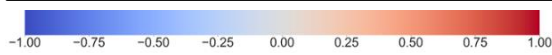
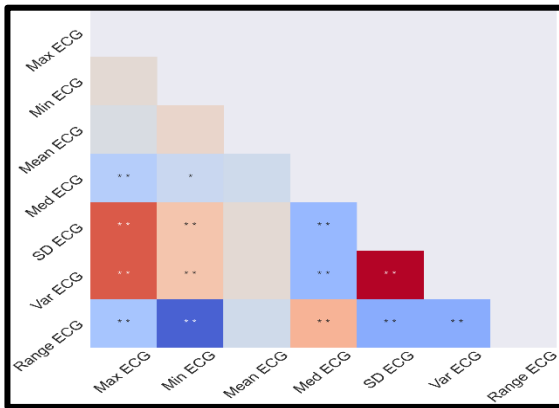


Fig. 3-9 - Correlation of the ECG features – Max ECG, Min ECG, Med ECG, Mean ECG, SD ECG, Var ECG, Range ECG.

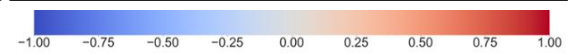
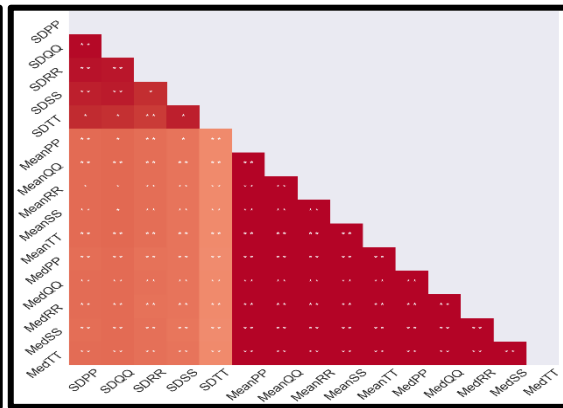


Fig. 3-10 - Correlation of the ECG peak features – SDPP, SDQQ, SDRR, SDSS, SDTT, MeanPP, MeanQQ, MeanRR, MeanSS, MeanTT, MedPP, MedQQ, MedRR, MedSS, MedTT.

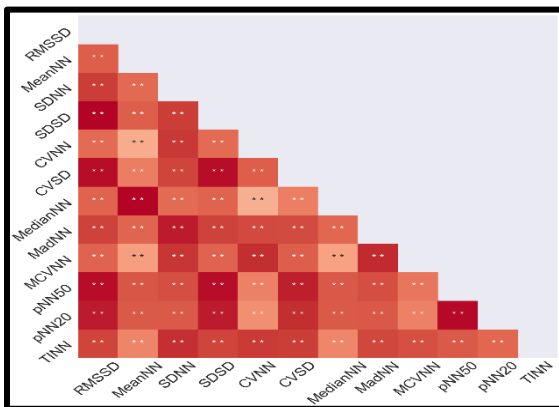


Fig. 3-11 - Correlation of the HRV time features - RMSSD, MeanNN, SDNN, CVNN, CVSD, MedianNN, MadNN, MCVNN, pNN50, pNN20, TINN.

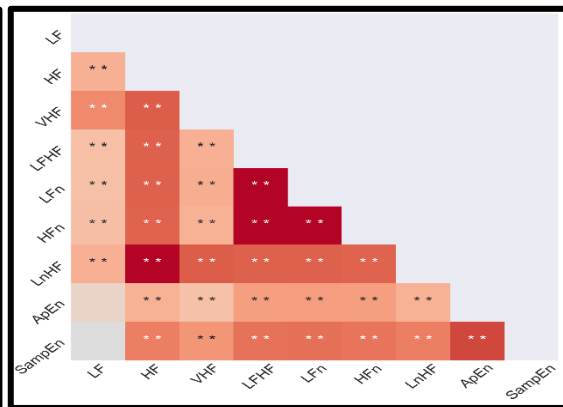


Fig. 3-12 - Correlation of the HRV frequency and non-linear features - LF, HF, VHF, LFHF, LFn, HFn, LnHF, ApEn, SampEn.

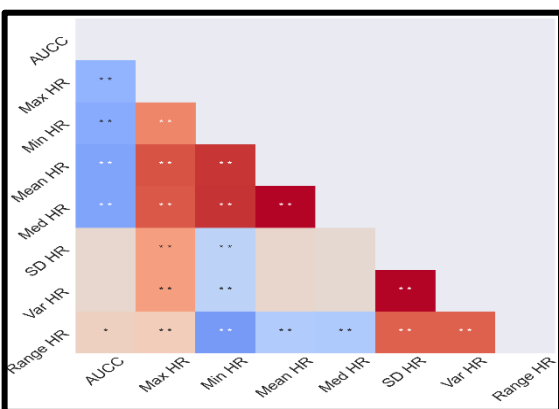


Fig. 3-13 - Correlation of the AUCC and the HR features – Min HR, Max HR, Mean HR, Med HR, SD HR, Var HR, and Range HR.

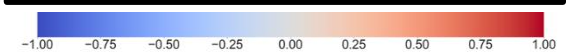
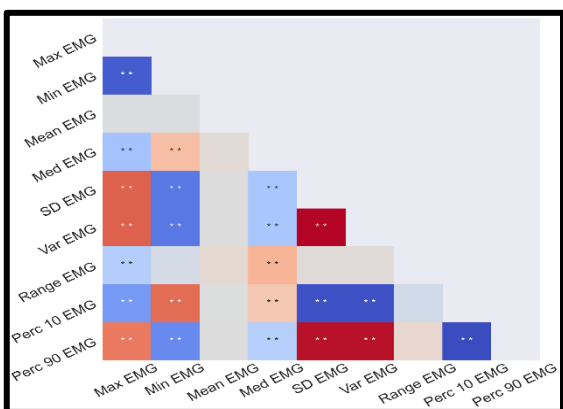


Fig. 3-14 - Correlation of the EMG features – Max EMG, Min EMG, Mean EMG, Med EMG, SD EMG, Var EMG, Range EMG, Perc 10 EMG, and Perc 90 EMG.

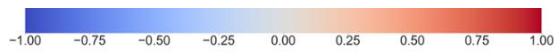
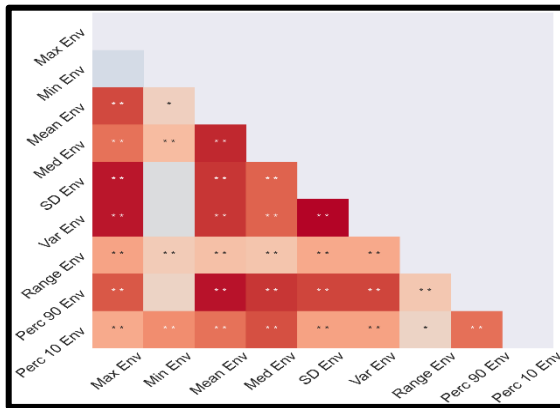


Fig. 3-15 - Correlation of the EMG Envelope features – Max Env, Min Env, Mean Env, Med Env, SD Env, Var Env, Range Env, Perc 90 Env, and Perc 10 Env.

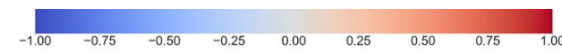
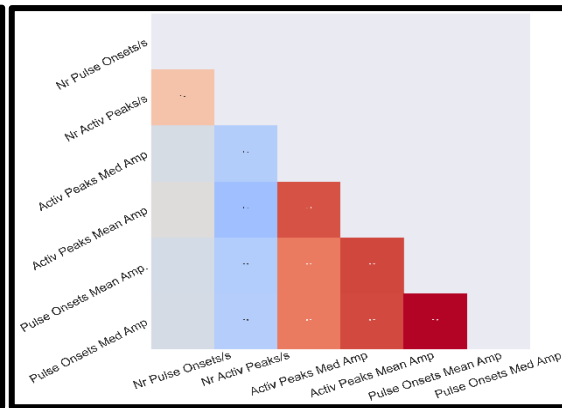


Fig. 3-16 - Correlation of the EMG Envelope features - Nr Pulse Onsets/s, Nr Activ Peaks/s, Med/Mean Activ Peaks, and Pulse Onsets Amp.

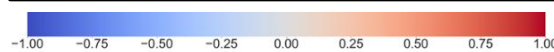
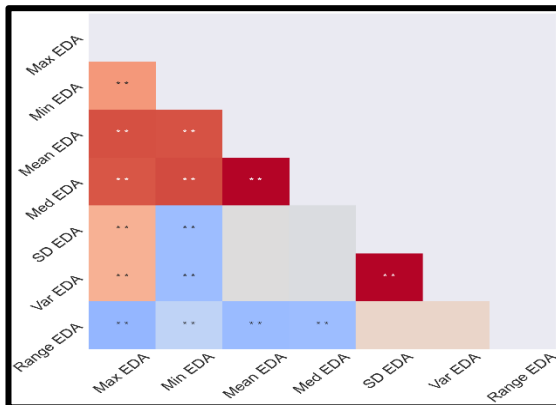


Fig. 3-17 - Correlation of the EDA features – Max EDA, Min EDA, Mean EDA, Med EDA, Var EDA, SD EDA, and Range EDA.

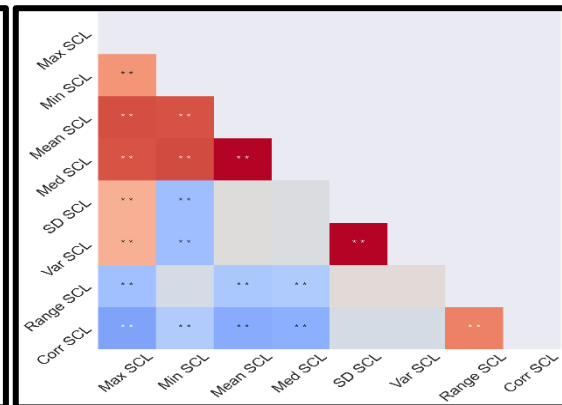


Fig. 3-18 - Correlation of the SCL features – Max SCL, Min SCL, Mean SCL, Med SCL, SD SCL, Var SCL, Range SCL, and Corr SCL.

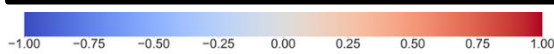
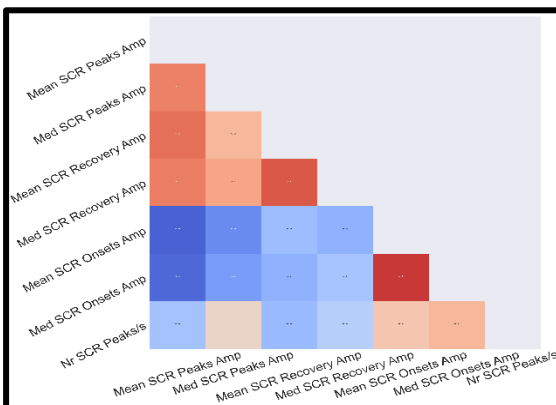


Fig. 3-19 - Correlation of the SCR features – Mean/Med peaks Amp, Mean/Med Recovery and Onsets Amp, and Nr SCR Peaks/s.

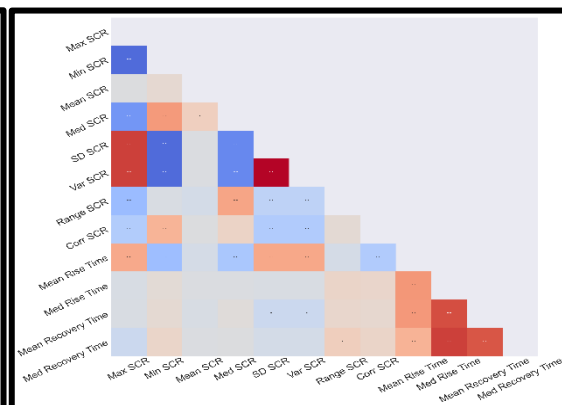


Fig. 3-20 - Correlation of the SCR features – Max SCR, Min SCR, Mean SCR, Med SCR, SD SCR, Var SCR, Range SCR, Corr SCR, Mean/Med Rise, and Recovery Time.

3.2.5. Anxiety and Affect Classes Characterization

The characterization of anxiety, positive and negative affect states, was performed by analysing if the features extracted from each signal, from the 15 individuals (before the data augmentation process), showed a significant difference between the respective classes. The Mann-Whitney U (MW) test, using the function “mannwhitneyu” [53] from the python package “SciPy” [51], appropriate for non-parametric features, was applied to determine which ones presented significant differences between anxiety and affect classes, with a confidence interval of 95%.

3.2.6. Pipeline Description

All the steps regarding the data normalization, use of supervised feature selection, data balancing, parameter tuning, and finally model evaluation were implemented using two different pipelines – one without data balancing techniques (Fig. 3-21) and the other for data balancing techniques (Fig. 3-22). The function “Pipeline” [54] from the “Imblearn” python package [55] allows us to apply all these steps while cross-validating (as described in Section 3.2.10).



Fig. 3-21 - Pipeline scheme with supervised feature selection techniques.

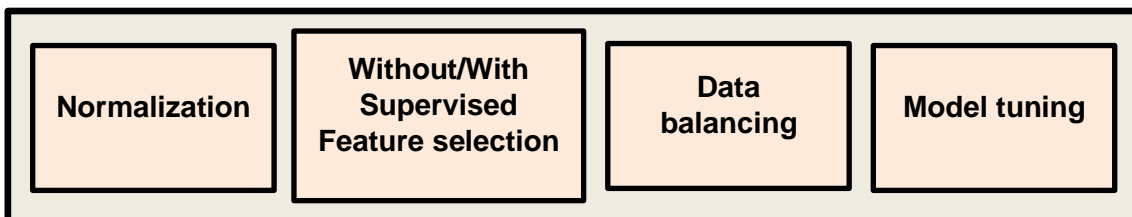


Fig. 3-22 - Pipeline scheme with data balance techniques.

3.2.7. Supervised Feature Selection – Filter and Wrapper Methods

Supervised feature selection selects features based on the target variable and is divided into three types of methods: filter, wrapper, and embedded methods. Concerning the supervised feature selection, two methods were applied and compared: filter – SelectKBest – and wrapper – RFECV – methods.

With filter methods, the selection of features is independent of any machine learning algorithms since features are selected based on their scores in various statistical tests for their relationship with the outcome variable [36].

The function “SelectKBest” [56] from the python package “Sklearn” [52] uses a statistical measurement technique to determine the influence and correlation of a feature, as opposed to that of the best feature subset (SelectKBest technique). In this case, the function “f_classif” [57] from the python package “Sklearn” [52] was chosen as a score function to select the best features, using an Analysis of Variance (ANOVA) F-test to determine the linear dependency of each feature with the target variable. The “SelectKBest” function [56] finds the best k-size subset of features, meaning the subset that achieves the best score with the classifier. The size of the subset was optimized inside the inner loop in the pipeline and was defined as $k = \{10,15,20,25,30\}$.

With wrapper methods, a subset of features is used to train a model, and then features are added or removed based on the inferences drawn from the model until a final subset is determined. Common examples of wrapper methods are forward feature selection, backward feature elimination, and recursive feature elimination (RFE) [36].

In this case, we decided to use Recursive feature elimination with cross-validation (RFECV) – through the function “RFECV” from the python package “Sklearn” [52]. “RFECV” [58] derives the best feature subset using the same process as RFE but unlike RFE, the number of features can be unspecified, and cross-validation is possible. This function uses by default 5-fold stratified cross-validation to automatically determine the optimal set of features. RFE [36] is a greedy optimization algorithm that aims to find the best-performing feature subset. It repeatedly creates models and keeps aside the best or the worst performing feature at each iteration. It constructs the next model with the left features until all the features are exhausted. It then ranks the features based on the order of their elimination. Therefore, the higher the number of features, the higher the number of computations, and the slower the method in terms of time complexity. The function “RFECV” [58] uses feature importance, which accesses the importance of the feature in the model, through the `coef_` or `feature_importances_` attributes of the estimator, and the one or few less important features are dropped in each iteration.

A minimum of ten features and a step equal to 1 (determining that one feature is eliminated in each computation) were defined in the function “RFECV” [58] parameters and the F1 was used as the scoring metric to choose the best set of features.

SVM presents a limitation when using the RFECV technique since it does not have a feature importance attribute on any kernel, except for Linear SVM, which is the one used in this study but may not be the most adequate for our data.

It is important to mention that when applying the different data balance techniques to our data while cross-validating, although the feature sets are selected by the supervised feature selection techniques before the data balancing, the best feature set is chosen considering the performance of the classifier after the data balance technique and classifier hyperparameter tuning. Then, for each classifier, supervised feature selection technique, and data balancing technique, the best combination of features may be different. In the case of the SelectKBest technique, since we are tuning the k parameter, not only the best feature set but also the best dimension of the features set can be different.

3.2.8. Feature Normalization

We did not eliminate outliers because they may represent important cases as they can embody more severe cases of anxiety and affect. When working with physiological signals, it is important to avoid subject-wise dependency (e.g., one person's sweating can be extraordinarily high).

Thus, the features were normalized, before the classification task, using the function "MinMaxScaler" from the python package "Sklearn" [52]. The Min-max normalization method follows the following formula to map the data to an $[0,1]$ interval: $X_n = (X - X_{min}) / (X_{max} - X_{min})$. By normalizing all the observations, not only do we reduce subject dependency but also guarantee that all features are on the same scale which is necessary for some classification algorithms.

3.2.9. Data Balance Techniques

As mentioned before, the dataset is imbalanced. Consequently, we chose several data balance techniques that change the data distribution. Under-sampling techniques are not an appropriate choice as we already have a small number of observations, and this technique would further reduce our data size. The imbalance ratio (IR) of anxiety is 2.75 (IR = 165/60). The imbalance ratio of negative affect and positive affect is 4 (IR = 180/45).

Four data balance techniques were applied – Random Oversampling (Over), SMOTE (SMO), ADASYN (ADA), and Borderline SMOTE 2 (Border) All these methods were settled to perform oversampling of the minority class and the parameters were set to guarantee a 1:1 class ratio. These data balance techniques were used in anxiety and negative affect classification. Positive affect classification did not present an imbalance problem since the positive class ("Low" PA) is also the majority class.

The “Imblearn” python package [55] implementation of the data balance techniques was used. The functions and parameters used for the data balance techniques are described in Table 6-4.

3.2.10. Classification of Anxiety and Affect levels

These tasks will be addressed as a binary classification. Based on the literature, we chose seven different machine learning models. All the models fall into the category of supervised learning algorithms. The algorithms tested in this research were: LR, LDA, DT, SVM, ADB, RF, and XGB. The classifiers were reported in the literature [18], [19] to attain good results and as suitable for the problem and the data available. Furthermore, the necessity to evaluate different machine learning algorithms that are easily reproducible, optimized, and do not take much computing power motivated the choices. The “Sklearn” python package [52] implementation of the classifiers was used. The python package’s functions used and the hyperparameters defined and optimized for each classifier are in Table 6-3.

In a classifier performance evaluation, a concern is that models may perform well when making predictions on data used during training, but poorly when classifying new data. This problem is referred to as overfitting and can be encouraged by the use of oversampling techniques [41]. To overcome this issue, we use cross-validation [41] which randomly splits the data many times, each time training the classifier from scratch using one portion of the data before measuring its performance on the remaining share of data. In the original dataset, the presence of severe skews, and instances of the minority class may not be present in the test set, so we use stratified cross-validation [41], where the original class distribution is preserved in each fold.

Nested stratified cross-validation was used to apply the supervised feature selection techniques, and the data balance techniques, and evaluate the models’ performance. By utilizing the same test set to both select the values of the parameters and evaluate, there is the risk of optimistically biasing our model evaluations. Therefore, if a test set is used to determine the model parameters, we need a different test set to get an evaluation of that model. The solution was nested cross-validation where inner cross-validation (inner loop) is used to tune the parameters and select the best model and outer cross-validation (outer loop) is used to evaluate the model selected [59]. The function “StratifiedKFold” from the python package “Sklearn” [52] was used to define the stratified cross-validation parameters for both loops, with no shuffling.

Through the use of the “Pipeline” function [54] from the “Imblearn” python package [55], the data from the training set was used to compute the normalization parameters (min and max value of each feature), which thereafter, were applied to normalize the data in both training and test sets. This function also guarantees that the data-balance techniques while nested stratified cross-validating, are only applied to the training set, maintaining the original class imbalance ratio in the test set.

The grid-search method, applied with the function “GridSearchCV” from the “Sklearn” python package [52], which consists of an exhaustive search over specified parameter values for an estimator, was used to optimize the model’s hyperparameters in the inner loop. The best hyperparameter combination is chosen using the F1 metric through 2-stratified-fold inner cross-validation.

The performance of the models was assessed through suitable metrics, as described in Chapter 2, Section 2.3.4: accuracy, F1, Roc-Auc, Precision, and Recall. For estimating the model’s performance, in the outer loop was used a 5-stratified fold outer cross-validation using the function “cross_validate” from the python package “Sklearn” [52]. The mean value of the performance metrics and their standard error (SE) were computed. Considering our problem, recall is more relevant than precision since misclassifying cases of high anxiety, low positive affect, and high negative affect is more prejudicial.

4. Results and Discussion

In this chapter, we will first discuss the results obtained on the characterization of anxiety, negative affect, and positive affect states. Next, we will examine the classification results, as well as the use of the different feature selection and data balance techniques. All the results presented in this chapter were attained using Python 3, version 3.8 [60].

4.1. Univariate Analysis

The features that showed a significant ability to discriminate between classes were used to describe each state of anxiety and negative and positive affect.

These features were selected through a univariate analysis, using an MW statistical test to assess if a feature shows significantly different medians between classes, which shows significant differences between anxiety and affect classes, with a confidence interval of 95%, for the features presented in Table 4-1, Table 4-2 and Table 4-3.

EMG	Med EMG
------------	---------

Table 4-1 Significant features for anxiety classification.

EDA	Max SCR, Min SCR, Med SCR, SD SCR, Var SCR, Nr of SCR peaks/s
EMG	Max EMG, Nr of Activ Peaks/s

Table 4-2 Significant features for negative affect classification.

ECG	RMSSD, MeanNN, SDNN, SDSD, CVNN, CVSD, MCVNN, MedianNN, pNN50, pNN20, TINN, MeanPP, MeanQQ, MeanRR, MeanSS, MeanTT, MedPP, MedQQ, MedRR, MedSS, MedTT, SDPP, SDQQ, SDRR, SDSS, SDTT, Min HR, Mean HR, Med HR, AUCC
EDA	Med EDA, SD EDA, Var EDA, Max SCL, Min SCL, Med SCL, SD SCL, Var SCL
EMG	Med Env

Table 4-3 Significant features for positive affect classification.

Nonetheless, this analysis only considers one variable and its effect/relationship on the target variable, disregarding the effect of relationships between the multiple variables and their combination.

Additionally, the accentuated imbalance of class distribution may affect the ability of the features to describe both classes since one class has fewer examples than the other.

4.1.1. Anxiety Characterization

The only feature that showed significant differences between anxiety classes was the Median amplitude of the EMG signal (Med EMG), which represents the median, a measure of centrality, of the amplitude of the EMG signal.

The EMG signal provides a useful characterization of the neuromuscular system, and it can be very effective in the context of anxiety characterization, since one of the manifestations of anxiety is increased muscular tension, such as in facial and trapezius muscles.

In the boxplot, in Fig. 4-1 it is possible to observe that the values of the feature are low and close to 0, as expected since no muscular activity is being induced. The “High” anxiety class shows higher median, maximum and minimum values than the “Low and moderate” anxiety class. Also, the “High” anxiety class displays a smaller interquartile range and a smaller spread than the “Low and moderate” anxiety class. None of the classes presented outliers. These findings may indicate that a higher muscular activation is present in the “High” anxiety class, as expected.

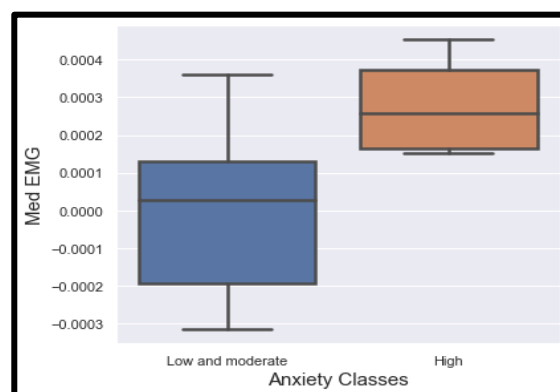


Fig. 4-1 - Boxplot of the EMG signal feature - Med EMG.

As reported in the literature [29], the amplitude of the EMG signal may be interpreted as a measure of the relative muscle tension. This measure can be taken directly from the EMG signal or after applying some signal transformations, for example,

after processing the EMG to obtain its linear envelope. The results are congruent with the literature, which has described a higher amplitude of the EMG signal during stress conditions than during rest conditions.

Neither EDA nor ECG extracted features showed significant differences between classes. We would expect measures from these two signals to discriminate between classes since in the literature these were found to be good indicators of stress and anxiety levels [37].

Nevertheless, the obtained results may be very influenced by the subject's position in the protocol of data collection. The fact that some participants were sitting, and others were standing may explain why some of the ECG and EDA features did not show differences between anxiety states [48].

4.1.2. Negative Affect Characterization

The characterization of negative affect states using the features described in Table 4-2 is now presented. Concerning the EMG signal, the Maximum amplitude of the EMG signal (Max EMG) and the Number of Activation peaks per second (Nr Activ Peaks/s) in the MW test showed significant differences between NA classes.

Regarding the Maximum amplitude of the EMG signal (Max EMG), when observing the boxplot (Fig. 4-2, on the left), the "High" NA class shows a lower median, minimum and maximum value, a smaller interquartile range, and a smaller spread. These findings indicate that a smaller maximum amplitude of the EMG signal is associated with higher levels of NA.

The Number of Activation peaks per second (Nr Activ Peaks/s), in Fig. 4-2 (on the right), represents the average number of times the amplitude of the signal was above a defined threshold and originates local maximums that represent the moment where the muscle contraction elevates.

When analysing the boxplot (Fig. 4-2, on the right), it shows a higher median value in the "High" NA class than in the "Low and moderate" NA class. In the "High" NA class, the interquartile range and the spread is smaller than in the "Low and Moderate" NA class.

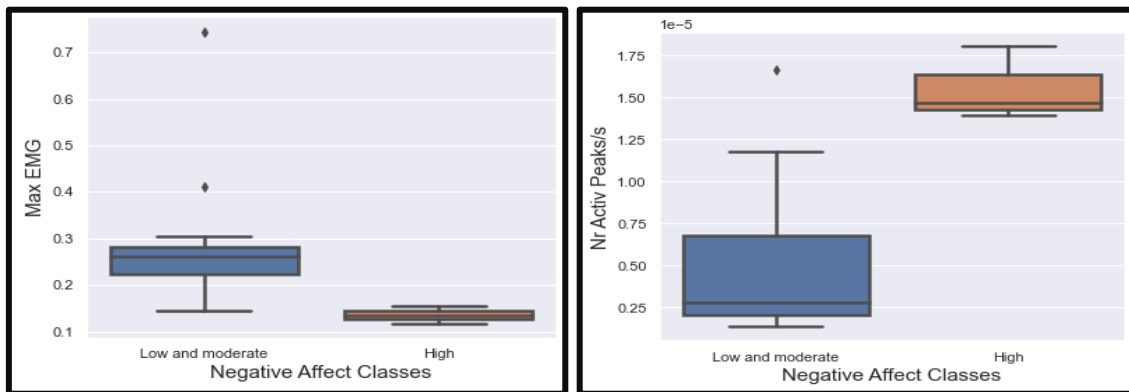


Fig. 4-2 - Boxplots of the EMG signal features - Max EMG (on the left) and Nr Activ Peaks/s (on the right).

These observations indicate that a higher frequency of activation peaks is associated with higher NA states. Literature calls emotional tone an involuntary contraction, permanent and moderate of muscles that results from an emotional state. This slight tension was reported as related to an expression of negative emotions such as fear or mental stress, showing that muscle activity increases during negative valence emotions [29]. Therefore, our findings are congruent with the literature since also indicate a higher muscular activation in higher NA states.

Concerning the EDA signal, the MW test's significant features were: Median SCR, Standard deviation SCR (SD SCR), Maximum SCR, Minimum SCR, Variance SCR, and Number of SCR peaks per second (Nr SCR Peaks/s). The boxplots of the EDA significant features: Median SCR, Maximum SCR, Minimum SCR, and Variance SCR are shown in the Attachments, in Fig. 6-11.

In Fig. 4-3, on the left boxplot, the Standard deviation of the SCR amplitude (SD SCR) shows a lower median value lower in the “High” NA class, as well as a lower interquartile range and spread. There are no outliers in either class. These observations seem to suggest that the SCR varies less in the “High” NA class.

In Fig. 4-3, on the right boxplot, it is possible to see that the Number of SCR Peaks per second (Nr SCR Peaks/s) has a higher median value in the “High” NA class. This class also shows a smaller interquartile range and spread. This is consistent with the literature as higher states of NA may boost skin activation as a result of increased sweat gland activity [61]. This increased physiological arousal reflects in the SCR component of the EDA and will increase the Number of SCR Peaks per second.

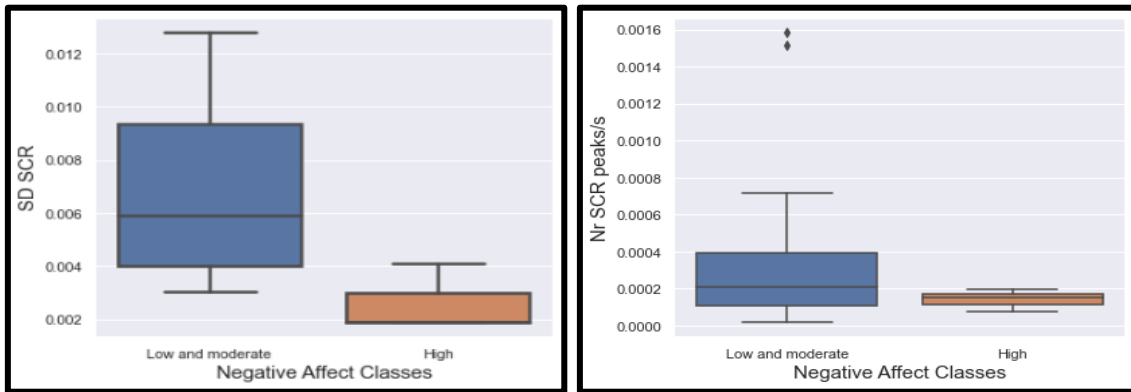


Fig. 4-3 - Boxplots of the EDA signal features - SD SCR (on the left) and Nr SCR peaks/s (on the right).

4.1.3. Positive Affect Characterization

The physiological characterization is still not entirely understood for both state and trait dimensions of positive affect [62], but there is converging evidence that positive affect activates the autonomic nervous system in meaningful ways [63].

The direction of the association of positive affect and cardiovascular function is determined by the type and intensity of the emotional state experienced. Positive affective states, such as joy, represent states of arousal, with a concomitant increase in heart rate and blood pressure, whereas others such as satisfaction with life, might be associated with a reduced reactivity.

Studies have found that, for example, the relationship between HRV and PA state is unclear since distinct types of positive affect can have different physiological manifestations. Positive emotionality, high or low, related to smooth, calm emotions shows a higher PNS activation whereas positive emotions related to excitement show higher activation of SNS [62]–[64].

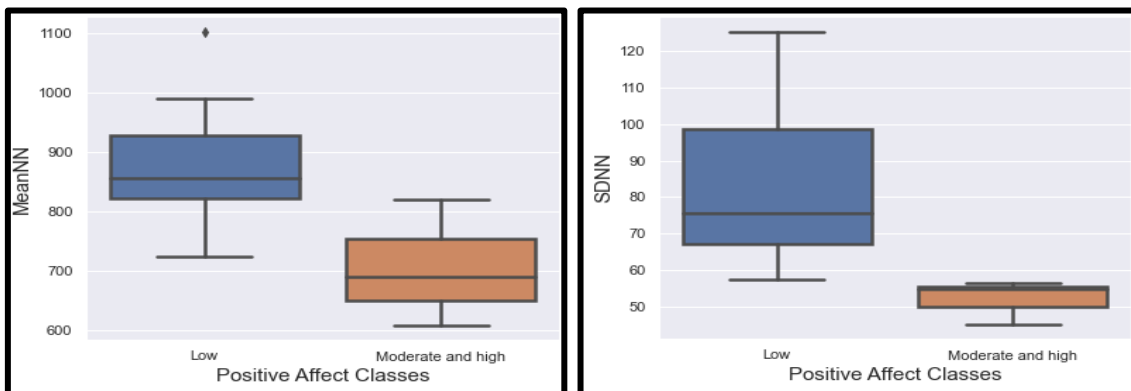


Fig. 4-4 - Boxplots of the HRV features – MedianNN (on the left) and SDNN (on the right).

The HRV time features that measure the interval between normalized R peaks, meaning the time interval between two adjacent heart beats – MeanNN (Fig. 4-4, on the left), MedianNN, MadNN, pNN50 and pNN20, TINN (Fig. 6-16) – and the features that describe the variation of those time intervals – CVNN, MCVNN, SDSD, SDNN, RMSSD (Fig. 6-13) – showed significant differences between PA classes.

Being HRV the fluctuation in the time intervals between adjacent heart beats and HR the number of heart beats per minute, it is straightforward that when the variation of these distances is higher this indicates the HRV is higher, and when the mean interval between R peaks is higher, the heart beats are longer and further apart and therefore the mean HR tends to be lower.

All the HRV time features that measure the length of the R-R intervals show to have similar behaviour and a clear pattern: the median values are significantly higher in the “Low” PA class and the values have a higher interquartile range and spread. This indicates that the time intervals between R peaks of the ECG signal are higher in the “Low” PA class, reflecting a lower HR. When observing the boxplot in Fig. 4-4, on the left, the median value of MeanNN is higher in the “Low” PA class, with both classes presenting a similar interquartile range.

Regarding all the HRV time features that measure the variation of the R-R intervals the same was verified. This means that the variability of the R-R time intervals is higher in the “Low” PA class, reflecting therefore a higher HRV. From the analysis of the boxplots, it is possible to observe that the “Low” PA class shows a higher HRV and lower HR, while the “Moderate and high” PA class show a lower HRV and higher HR.

These findings are consistent with the literature which states that in higher PA levels, related to activation and excitement, the activation of the SNS results in a higher HR and a possibly lower HRV [62]–[64]. In Fig. 4-4, on the right boxplot, the median value of SDNN is higher in lower PA levels. Also, the “Low” PA class presents a bigger interquartile range and spread.

The boxplots of all the standard deviation and mean/median distance between non-normalized cardiac cycle peaks (composed of the ECG peaks P, Q, R, S, and T), are shown in Attachments in Fig. 6-12, Fig. 6-14, and Fig. 6-15, present the same pattern.

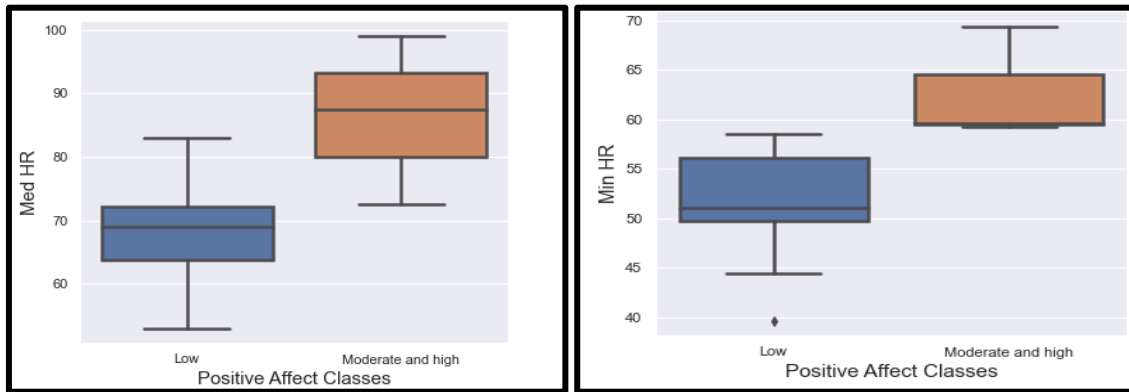


Fig. 4-5 - Boxplot of the heart rate features – Med HR (on the left) and Min HR (on the right).

The features Mean HR, AUCC (shown in Fig. 6-17, on the left and right, respectively), Median HR (Med HR), and Minimum HR (Min HR) showed significant differences between positive affect classes.

In Fig. 4-5, on the left boxplot, the Median HR (Med HR), has a lower median value in the “Low” PA class as well as a smaller interquartile range compared to the “Moderate and high” PA class. In Fig. 4-5, on the right boxplot, the Minimum HR (Min HR) shows a lower median value in the “Low” PA class. These findings suggest a higher HR corroborating the conclusions drawn that a lower HR is associated with lower PA levels, as a higher HR may be related to higher PA levels.

As mentioned before, the amplitude of the linear envelope of the EMG signal is a frequently used measure of the signal activation level and muscular tension. In Fig. 4-6, it is possible to observe that the median value of the median EMG linear envelope amplitude (Med Env) is higher in the “Low” PA class as well as its interquartile range and spread. These observations imply that there is a higher muscular tension associated with low PA states.

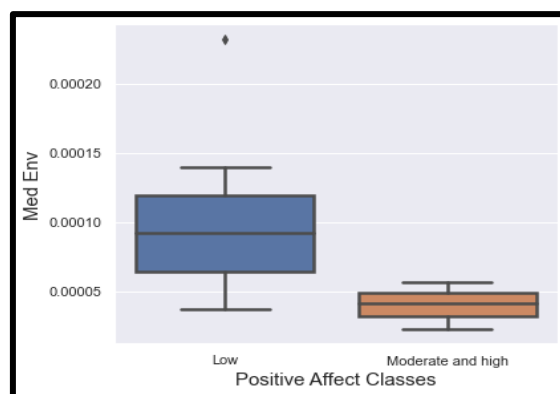


Fig. 4-6 - Boxplot of the EMG signal feature - Med Env.

Concerning the EDA signal, the features Standard deviation EDA, Variance EDA, Median EDA, Maximum SCL, Minimum SCL, Standard deviation SCL (SD SCL), Median SCL (Med SCL), and Variance SCL were able to characterize positive affect, as the MW showed significant differences between the median values of PA classes. The boxplots of the features Standard deviation EDA, Variance EDA, Median EDA, Maximum SCL, Minimum SCL, and Variance SCL are displayed in Fig. 6-18.

As observed in Fig. 4-7, on the left, the median value of the Median SCL amplitude (Med SCL) is lower in the “Low” PA class, but the interquartile range and spread are wider. In Fig. 4-7, on the right, it is possible to observe that the median value of the Standard deviation of the SCL amplitude (SD SCL) is higher in the “Low” PA class and its interquartile range and spread are considerably wider. The Standard deviation EDA, Variance EDA, Variance SCL, and Maximum SCL share the same behaviour as the Standard deviation SCL as seen in the Attachments, in Fig. 6-18.

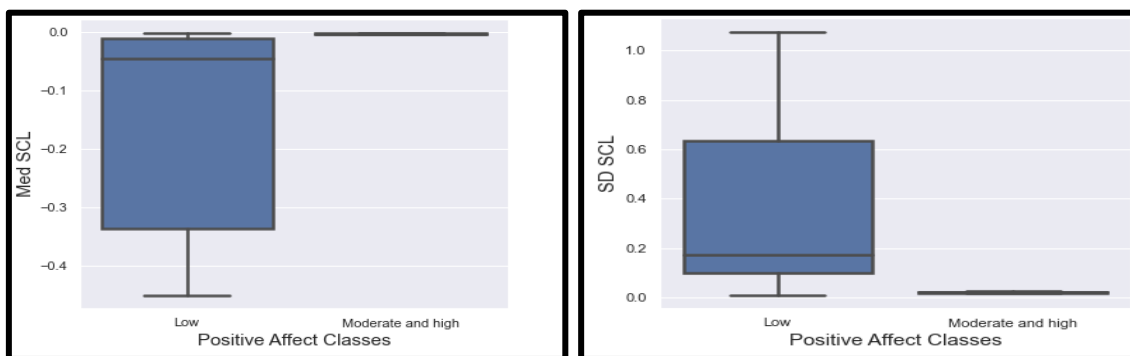


Fig. 4-7 - Boxplots of the EDA signal features - Med SCL (on the left) and SD SCL (on the right).

From the observations, we can state that lower PA states are linked with lower electrodermal activity, whereas high PA states reveal the opposite as visible in the EDA signal and its SCL component. None of the SCR extracted features showed significant differences between groups which may indicate that arousal events are not related to either of the PA classes in a neutral condition, whereas baseline arousal – a constantly higher level of arousal - is present in the "Low" PA class.

4.2. Unsupervised Feature Selection

Unsupervised feature selection was performed based on the pair-wise correlation method: in the case of two features with a correlation coefficient higher than 0.90, the feature with higher variance was selected.

The resultant feature set, with a total of 60 features is displayed in Table 4-4. To the presented feature set were then applied the supervised feature selection techniques – SelectKBest and RFECV – as shown in Section 4.3.

ECG	AUCC, Mean ECG, Med ECG, Max ECG, Min ECG, Var ECG, Range ECG, MeanNN, MadNN, TINN, pNN50, CVNN, LnHF, LF, VHF, LFn, SampEn, ApEn, SDTT, Max HR, Min HR, Var HR, Range HR,
EDA	Min EDA, SD EDA, Range EDA, Mean SCR, Med SCR, Max SCR, Min SCR, SD SCR, Range SCR, Corr SCR, Nr SCR Peaks/s, Mean and Med SCR Peaks Amp, Med SCR Onsets Amp, Med, and Mean SCR Recovery Amp, Mean and Med SCR Recovery Time and Rise Time, Max SCL, Med SCL, Range SCL, Corr SCL
EMG	Min EMG, Mean EMG, Med EMG, Range EMG, Perc 10 EMG, Min Env, Range Env, Perc 10 Env, Nr Pulse Onsets/s, Nr Activ Peaks/s, Mean and Med Amp Activ Peaks, Med Pulse Onsets Amp

Table 4-4 Features selected after the unsupervised feature selection and before the supervised feature selection.

4.3. Supervised Feature Selection and Classification Results

We compared the estimated performance of the classification models without using supervised feature selection and when using two different supervised feature selection methods – SelectKBest and RFECV - and four different data balance techniques – Random Oversampling, SMOTE, ADASYN, and Borderline SMOTE 2.

When classifying anxiety and negative affect, some of the classifiers showed, relatively lower measures of F1-score because of low recall or precision, while still maintaining good accuracy values. This may reflect the impact of data imbalance on the classification performance of the models. Therefore, some data-level balance techniques, described in Chapter 2, Section 2.3.3, were applied to improve the performance of the classifiers.

The commonly used “default” sampling rate of 1:1 might not always be the best option for every dataset and classifier. It would be beneficial to tune the sampling ratio on our dataset and consider each classifier, however, for comparison purposes between different balancing techniques and classifiers, the same parameters were chosen for each data balance technique and applied to all classifiers. Some classifiers may respond better to the parameters chosen, which can partially explain the differences encountered in the results for each data balance technique.

4.3.1. Anxiety Feature Selection

The supervised feature selection was implemented inside a pipeline with nested cross-validation. Thus, we retrieved for each model the best hyperparameters and the best set of features of the best estimator found through the grid-search method applied in the inner loop to the final features set described in Table 4-4.

Without data balancing, the frequencies of the features chosen at least three times by the supervised feature selection techniques are presented in Fig. 4-8. This figure displays the features that could be of greater importance for anxiety classification.

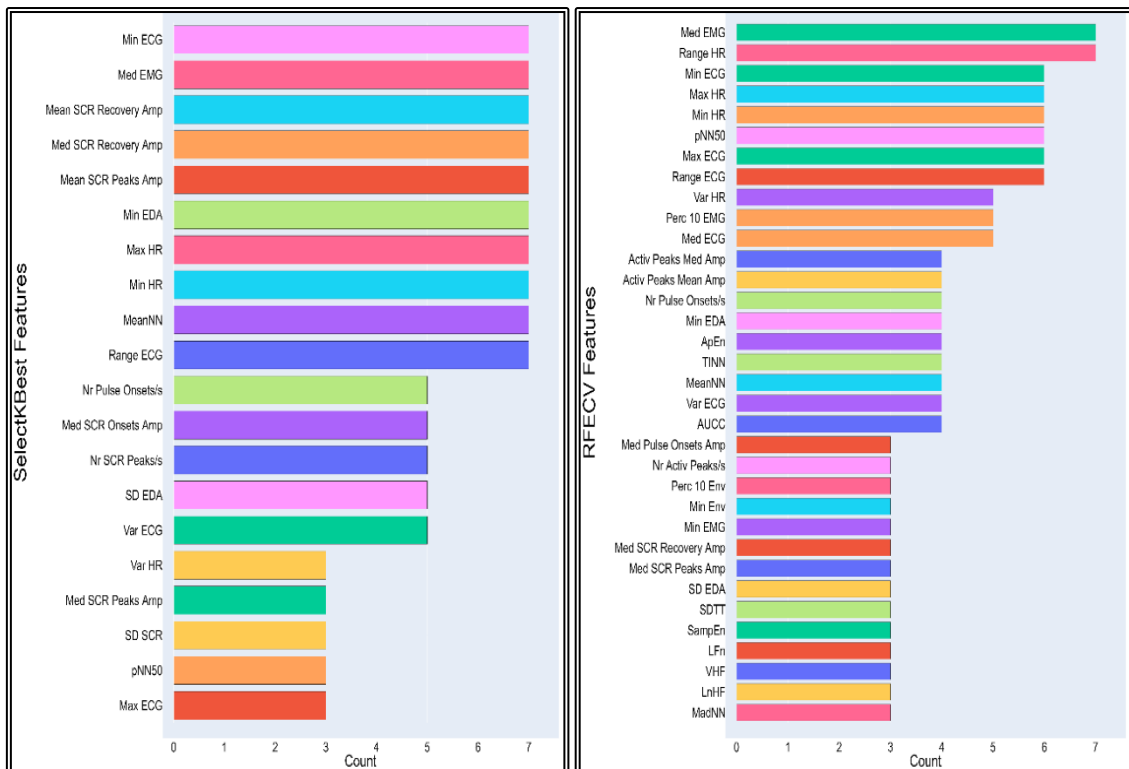


Fig. 4-8 - Features selected at least three times with SelectKBest (on the left) and with RFECV (on the right) for anxiety classification.

The results for the features selected with SelectKBest are shown in Fig. 4-8, on the left: 20 from the total of 25 different features selected in different feature groups were considered, with a possible minimum set size of ten features and a possible maximum set size of 30 features. The same group of ten features was present in all feature sets for all classifiers: Min ECG, Med EMG, Mean and Med SCR Recovery Amp, Mean SCR Peaks Amp, Min EDA, Max HR, Min HR, MeanNN, and Range ECG.

With RFECV, a total of 55 different features were selected, with a minimum defined feature set size of ten features. From Fig. 4-8, on the right, we can observe that Med EMG and Range ECG are present in the feature sets selected for all the classifiers,

while Range HR and Max HR are chosen for six of the seven classifiers. Perc 10 EMG, Min HR, pNN50, TINN, MeanNN, Var ECG, and Max ECG are present in five of the seven chosen feature sets.

Comparing both methods, we can say that the top 10 most frequently selected features have predominantly ECG-extracted features. Nonetheless, with RFECV, at the top 10, there are no features extracted from the EDA signal. This might suggest that when in combination with other physiological signals extracted features, the ECG and EMG features add more relevant information than EDA features.

Med EMG was the only EMG signal-extracted feature present in the top 10 in both supervised feature selection techniques and showed to be present in the feature sets for all classifiers. This feature was as well the only one used to characterize anxiety in Section 4.1.1, reinforcing its ability to distinguish between anxiety classes.

4.3.2. Anxiety Classification

The anxiety classification results for each classifier, with and without supervised feature selection, and data balance techniques, are presented in Table 6-9 to Table 6-15.

In classifying anxiety, the classifiers RF and XGB did not improve their performances after the supervised feature selection, as shown in Fig. 4-9, Fig. 4-12, and Fig. 4-13. RF did not improve its performance with SelectKBest and its F1-score with RFECV, and XGB did not achieve better results with either SelectKBest or RFECV.

These models are tree-based classifiers that already establish a feature importance ranking on themselves and therefore may not always benefit from feature selection. This ranking is determined through a measure of feature importance based on the mean decrease in impurity. Impurity is quantified by the splitting criterion of the decision trees which for RF is defined as the default setting Gini index and for XGB as the default setting information gain criterion, as shown in Table 6-3.

The Gini index measures the frequency at which any element of the dataset will be mislabelled when randomly labelled, ranging from 0 to 0.5. It is 0 when the node is considered pure, meaning that all its elements are of one unique class and therefore, this node will not be split again and reaches its maximum value - 0.5 - when the probability of the two classes is the same. The optimum split is chosen using the features with a lower Gini index [36].

The Information gain criterion, used by XGB, consists of a measure that indicates the disorder or uncertainty of features concerning the target and ranges from 0 to 1, getting its minimum value and maximum value the same way as the Gini index [36]. However, these methods can attribute high importance to features that may not be predictive of unseen data when the model is overfitting and therefore causing it to perform worst.

The RFECV technique uses the feature importance attribute of the classifiers, and therefore, if XGB is overfitting it can explain why this technique was not able to improve the XGB performance. Concerning the SelectKBest technique employed, it uses the function “f_classif” [57] from the python package “Sklearn” [52], which only detects linear dependencies between the features and the target variable, which may cause it to sometimes not choose the optimal features, especially for non-linear classifiers.

Nonetheless, an advantage of feature selection is that it reduces the score time of classifiers since it decreases the number of features. This makes these techniques helpful tools if the data dimension is high, and it is necessary to further reduce the score times of the classifiers.

Despite the described above, most classifiers improved their performance with supervised feature selection. The results for anxiety classification, without supervised feature selection techniques, are presented in Fig. 4-9.

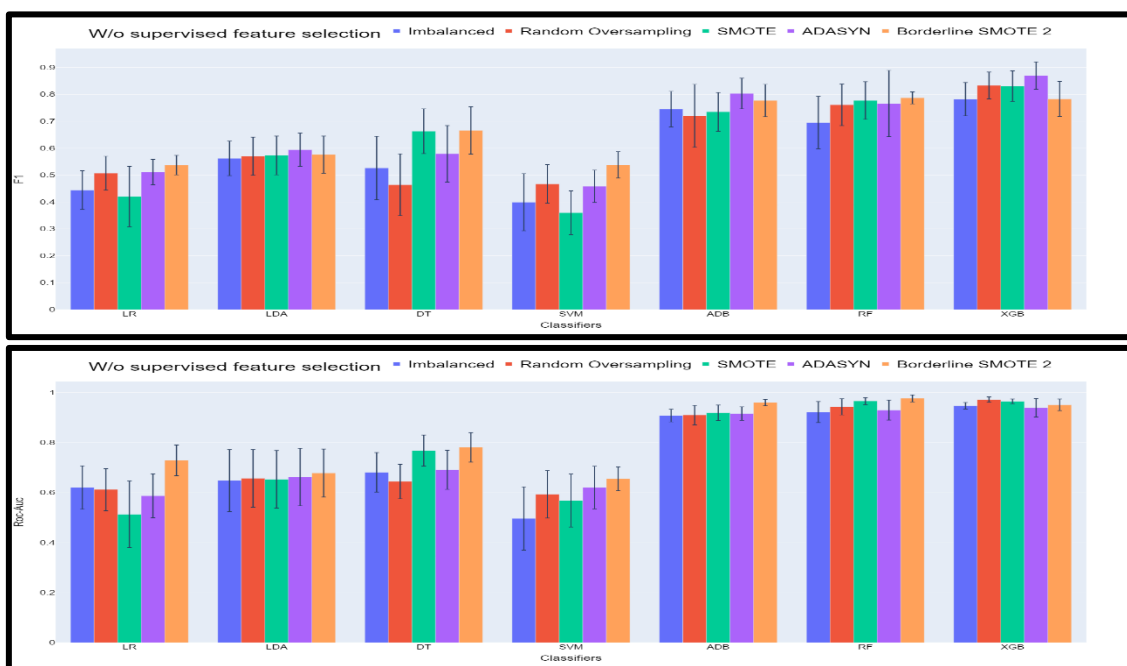


Fig. 4-9 - Anxiety classification results without supervised feature selection (top: F1-score, bottom: Roc-Auc).

Without using supervised feature selection techniques, concerning the F1 scores, Random Oversampling improved all classifiers' performance, except for DT and ADB, of which the scores declined. SMOTE increased all the classifiers' performance, excluding LR, and SVM. ADASYN and Borderline SMOTE 2 appeared to improve all the classifiers' scores except for ADB. Overall, the best F1 score was reached using ADASYN and XGB with a mean F1 score of 0.870 (SE=0.051). Regarding the Roc-Auc scores, Random Oversampling, SMOTE and ADASYN were not able to boost the performance of LR, while Random Oversampling also did not increase the score of DT and ADASYN of XGB. Borderline SMOTE 2 increased the scores of all classifiers. The best Roc-Auc score was achieved with RF and Borderline SMOTE 2, with a mean score of 0.977 (SE=0.014).

Overall, without supervised feature selection, Borderline SMOTE 2 was the most consistent technique and achieved the best results in both metrics in five classifiers - LR, DT, SVM and RF - and Roc-Auc with LDA and ADB. Borderline SMOTE 2 performed better in Roc-Auc which may reflect an introduction of noise or class overlap. Random Oversampling was the best performing technique in Roc-Auc with XGB. ADASYN achieved the highest scores in F1 with LDA, ADB, and XGB.

The success of Borderline SMOTE 2 can be explained by the way it synthesizes new instances. To achieve better predictions, most of the classification algorithms attempt to learn the borderline of each class as exactly as possible in the training process. The examples on the borderline and nearby are more likely to be misclassified than the ones far from it, and thus more important for classification.

Borderline SMOTE 2 differs from other oversampling techniques such as Random Oversampling and SMOTE, in which a random subset of the minority class is over-sampled and acts only on the examples on the decision boundary, which, however, depending on the data distribution, may on some occasions lead to a degree of class overlap. Borderline SMOTE 2 acts distinctly from Borderline SMOTE 1 since it synthesizes new minority examples not only from its minority neighbours but also from its majority neighbours.

To better understand the performances of the data balance techniques, we look at how the observations are distributed in the data space without the use of supervised feature selection in Fig. 4-10. It can be seen that there is a little overlap of classes where the instances might be easily misclassified and that there is no relevant noise or small disjuncts in the data.

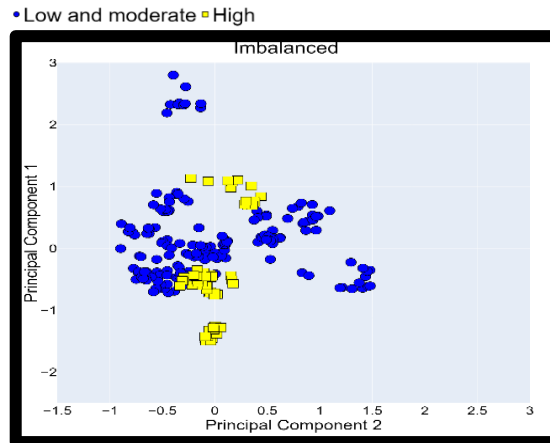


Fig. 4-10 - PCA analysis of the anxiety classes.

Contrary to Random Oversampling and SMOTE, which are random methods, ADASYN also detects border areas, although less hard boundaries than Borderline SMOTE. This may explain why, overall, ADASYN performed better than SMOTE and Random Oversampling. As we can see from the PCA analysis in Fig. 4-11, even if the class overlap persists and is even a little increased by the data balance techniques, a better definition of the decision boundary is provided, especially with Borderline SMOTE 2 and ADASYN.

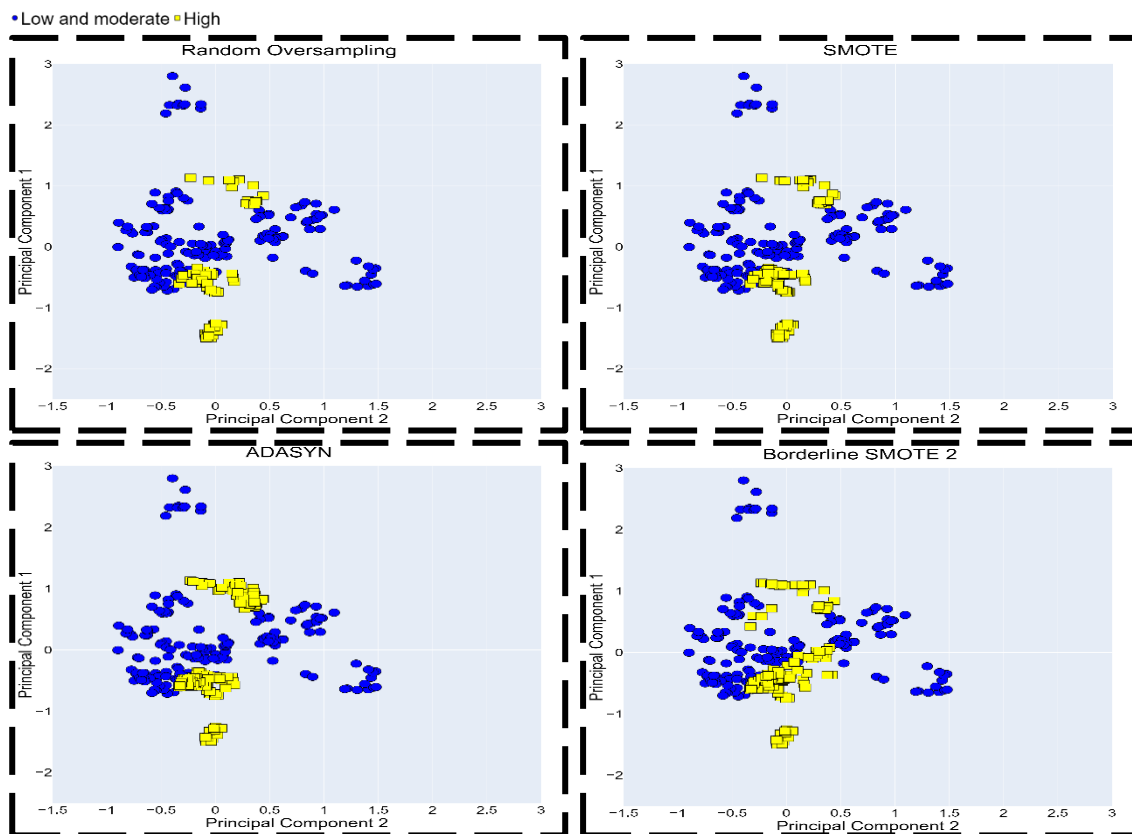


Fig. 4-11 - PCA analysis of the data balance techniques applied to the anxiety classes.

A comparison study [40] that used Random Oversampling, SMOTE, ADASYN, and Borderline SMOTE 1 and 2 supports our results. As a whole, for both true positive rate (TPR) and F1, Borderline SMOTE 1 and Borderline SMOTE 2 presented better performances than the other data balance techniques.

As explained in Chapter 3, Section 3.2.7, the evaluation scores are obtained for the best combination of supervised feature selection technique, data balance technique, and classifier. The combinations for the highest scores found for each classifier are described for SelectKBest in Table 6-5 and RFECV in Table 6-6. Fig. 4-12 exhibits all the scores obtained with SelectKBest for anxiety classification.

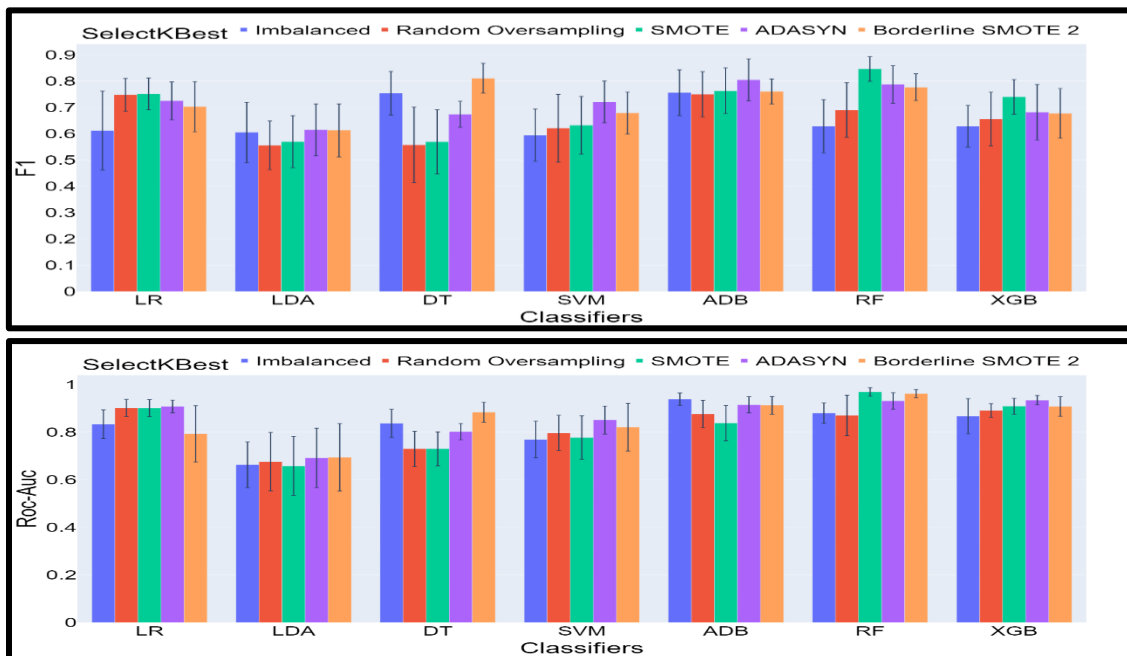


Fig. 4-12 - Anxiety classification results using SelectKBest (top: F1-score, bottom: Roc-Auc).

Using SelectKBest, when analysing the results obtained with the data balance techniques, we were able to draw the following observations. About F1, Random Oversampling showed improvements with all classifiers aside from LDA, DT, and ADB. SMOTE increased the performance for all classifiers except LDA and DT. ADASYN did not improve DT's performance, but on the other hand, Borderline SMOTE 2 improved the performance of all classifiers. The best F1 score was achieved with SMOTE and RF with a mean score of 0.846 (SE=0.047). Concerning Roc-Auc, Random Oversampling showed improvements with all classifiers but DT, ADB, and RF. SMOTE showed no enhancements in the performance of LDA, DT, and ADB. ADASYN did not improve the DT's and ADB's scores and Borderline SMOTE 2 did not show improvements with LR

and ADB. The best Roc-Auc result was attained with SMOTE and RF, with a mean score of 0.969 (SE=0.018). Using SelectKBest, the combinations that reached the best scores for each classifier are presented in Table 6-5.

When looking at the results attained using RFECV and data balance techniques shown in Fig. 4-13, we can draw the following observations.

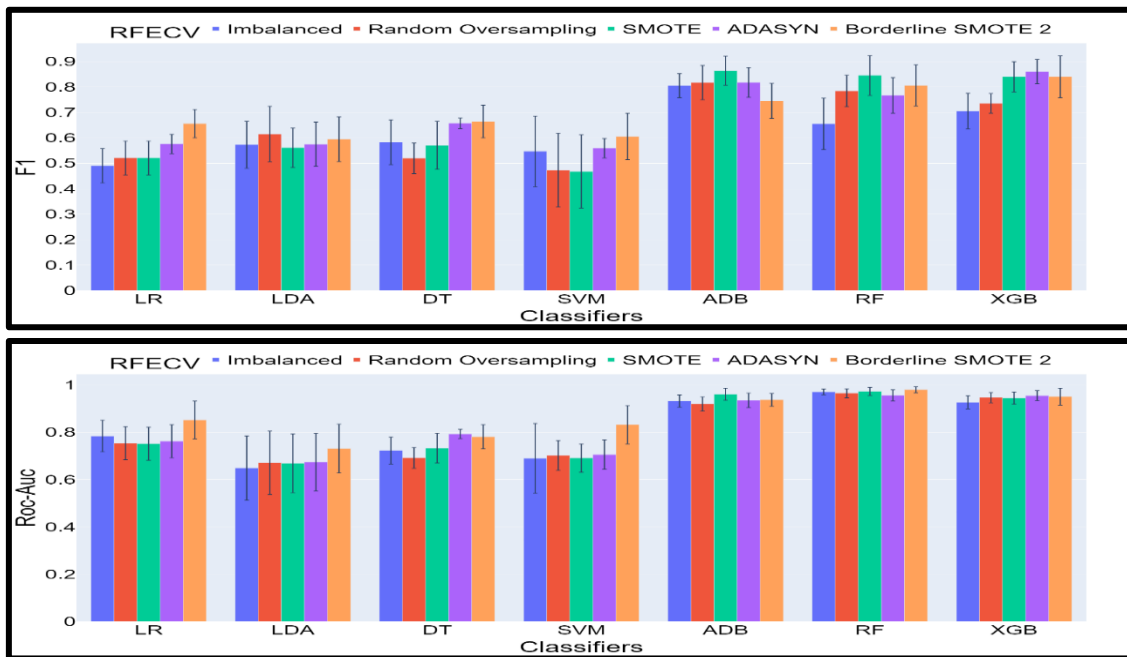


Fig. 4-13 - Anxiety classification results using RFECV (top: F1-score, bottom: Roc-Auc).

While applying RFECV, respecting the F1 metric, Random Oversampling improved all classifiers' performance except DT and SVM, SMOTE showed improvements in all classifiers' performance aside from LDA, DT, and SVM, while ADASYN appeared to increase the performance of all classifiers. Borderline SMOTE 2 did not improve the score of ADB. The best F1 result was achieved using SMOTE and ADB, with a mean score of 0.864 (SE=0.057). Regarding Roc-Auc, Random Oversampling showed improved scores for LDA, SVM, and XGB, while SMOTE increased the evaluation scores of all classifiers except LR. ADASYN improved all models' performance apart from LR and RF, and Borderline SMOTE 2 boosted the performance of all classifiers. The best Roc-Auc score was gotten with Borderline SMOTE 2 and RF with a mean score of 0.980 (SE=0.017). Using RFECV, the combinations that obtained the best scores for each classifier are shown in Table 6-8.

When using SelectKBest and RFECV, for both metrics, a pattern can be distinguished in the global efficacy of the balance techniques which reveals a better

performance of the combination of the supervised feature selection techniques with Borderline SMOTE 2 but also with ADASYN, and the different classifiers.

As presented in Table 6-7, with SelectKBest, Borderline SMOTE 2 achieved the best results in both metrics with one classifier - DT and in Roc-Auc with LDA. SMOTE was the best technique in both metrics with RF and F1 with LR. ADASYN outperformed the other techniques, in both metrics, in SVM, ADB, and XGB, in Roc-Auc with LR, and in F1 with LDA. Random Oversampling was the least efficient technique in both metrics. Ranking the combinations of SelectKBest and the data balance techniques, considering the number of classifiers in which the highest scores were achieved, we obtain the following order: ADASYN, Borderline SMOTE 2 together with SMOTE, and Random Oversampling.

In Table 6-8 it can be seen that when using RFECV, Borderline SMOTE 2 attained the highest results in both metrics with two classifiers – LR, and SVM, in F1 with DT, and Roc-Auc with LDA and RF. SMOTE was the best technique with two classifiers – in both metrics with ADB and F1 with RF. ADASYN outperformed the other techniques in XGB, and in Roc-Auc with DT. Random Oversampling was again the least efficient technique in both metrics achieving only the best results in F1 with LDA. Ordering the combination of RFECV with the data balance techniques, considering the number of classifiers in which the best results were obtained, we get the following sequence: Borderline SMOTE 2, ADASYN together with SMOTE, and Random Oversampling.

In some cases, the classifiers did not increase any of the evaluation metrics: without supervised feature selection, this is verified for SMOTE with LR and Random Oversampling with DT; with SelectKBest, this was reported for Random Oversampling with DT, and ADB, for SMOTE with LDA and DT and for ADASYN with DT; employing RFECV, this was noted for Random Oversampling with DT.

In these cases, the explanation may be due to the introduction of noise or class overlap resulting from the data balance technique change in the data distribution. Also, overfitting may be an explanation for this worsening performance: LR, LDA, DT, and ADB were the models where neither of the metrics was improved. Globally, the best results using the data balance techniques were reached by the combinations detailed next:

a) Concerning F1, without supervised feature selection techniques, ADASYN and XGB achieved the best performance with a mean score of 0.870 (SE=0.051); with SelectKBest, SMOTE and RF obtained an F1 mean score of 0.846 (SE=0.047); lastly,

the combination of RFECV, SMOTE, and ADB achieved an F1 mean score of 0.864 (SE=0.057). The best global combination was without supervised feature selection techniques, ADASYN and XGB.

b) Concerning Roc-Auc, without supervised feature selection, RF and Borderline SMOTE 2 achieved a mean score of 0.977 (SE=0.014). With SelectKBest, SMOTE and RF presented a mean score of 0.969 (SE=0.018). With RFECV, Borderline SMOTE 2 and RF obtained a mean score of 0.980 (SE=0.013). The best overall combination was with RFECV, Borderline SMOTE 2, and RF.

4.3.3. Negative Affect Feature Selection

Similarly, to what was made for anxiety, we analyzed, without data balancing, the features selected at least three times, with the supervised feature selection techniques, which are shown in Fig. 4-14.

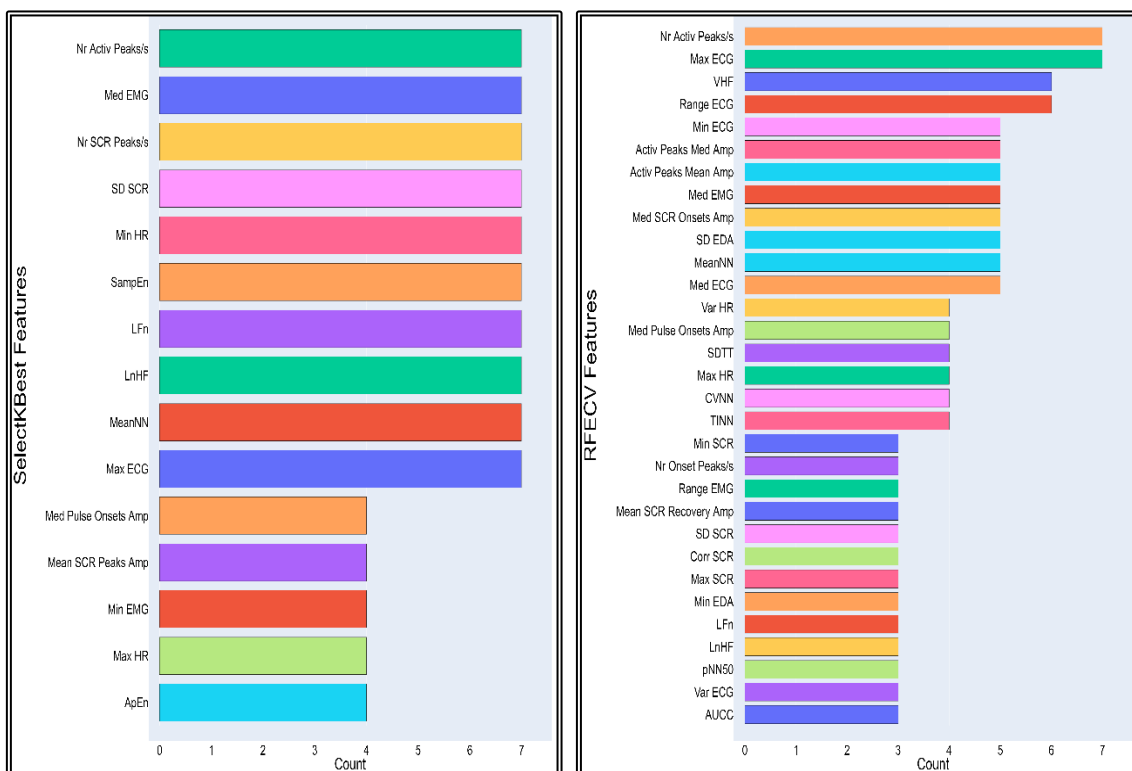


Fig. 4-14 - Features selected at least three times with SelectKBest (on the left) and with RFECV (on the right) for negative affect classification.

About the features selected for negative affect classification, using SelectKBest shown in Fig. 4-14, on the left, 16 from a total of 25 different features were considered (with a minimum possible set size of ten and a maximum possible set size of 30), and ten features were used in all the seven classifiers. This group is composed of the

following features: Nr of Activ Peaks/s, Med EMG, Nr of SCR Peaks/s, SD SCR, Min HR, SampEn, LFn, LnHF, MeanNN, and Max ECG. We can notice that this set includes features from all physiological signals. Most features were extracted from ECG, followed by EMG and EDA.

When using RFECV, with a possible minimum set size of ten features, 57 different features were selected in different feature sets. As shown in Fig. 4-14, on the right, two features were present for the seven classifiers – Nr of Activ Peaks/s and Max ECG. VHF and Range ECG were selected for six classifiers. Min ECG, Activ Peaks Med and Mean Amp, Med EMG, Med SCR Onset Amp, SD EDA, MeanNN, and Med ECG, which were extracted from the three signals, were selected for five classifiers.

Through SelectKBest, the top 10 most frequently selected features include features from all signals, with two EMG features, two EDA features, and six ECG features. With RFECV, the two features selected for all classifiers were an EMG-extracted feature and one ECG-extracted feature. These results reinforce the efficacy of combining different physiological signals. In total, most of the features were extracted from the ECG signal, and consequently, we can state that ECG was the most important signal for negative affect classification when using supervised feature selection.

The only two features chosen for all the classifiers by the two feature selection techniques were the Nr of Activ Peaks/s and Max ECG. This converges a bit with the negative affect characterization done before and described in Section 4.1.2 since the Nr of Activ Peaks/s was one of the discriminant EMG extracted features.

4.3.4. Negative Affect Classification

The negative affect classification scores for each classifier with and without supervised feature selection, and data balance techniques, are presented in Table 6-16 to Table 6-22. For negative affect classification, ADB did not improve its performance either with SelectKBest or RFECV and RF did not improve its performance with SelectKBest, as seen when comparing Fig. 4-15, Fig. 4-18 and Fig. 4-19.

Both these classifiers are tree-based algorithms that use the Gini index [36] importance measure to establish a feature importance ranking when classifying. As explained, if in the case of negative affect classification, ADB is overfitting, RFECV may not choose the most valuable features with unseen data and thus worsen the classifier's performance. Furthermore, the "f_classif" function used with "SelectKBest" [57], may work better for linear classifiers than for non-linear ones, as stated before.

Notwithstanding, the classifiers' score times were reduced with the supervised feature selection, and the majority of classifiers still showed performance improvements with the supervised feature selection techniques. The results for negative affect classification, without supervised feature selection, are presented in Fig. 4-15.

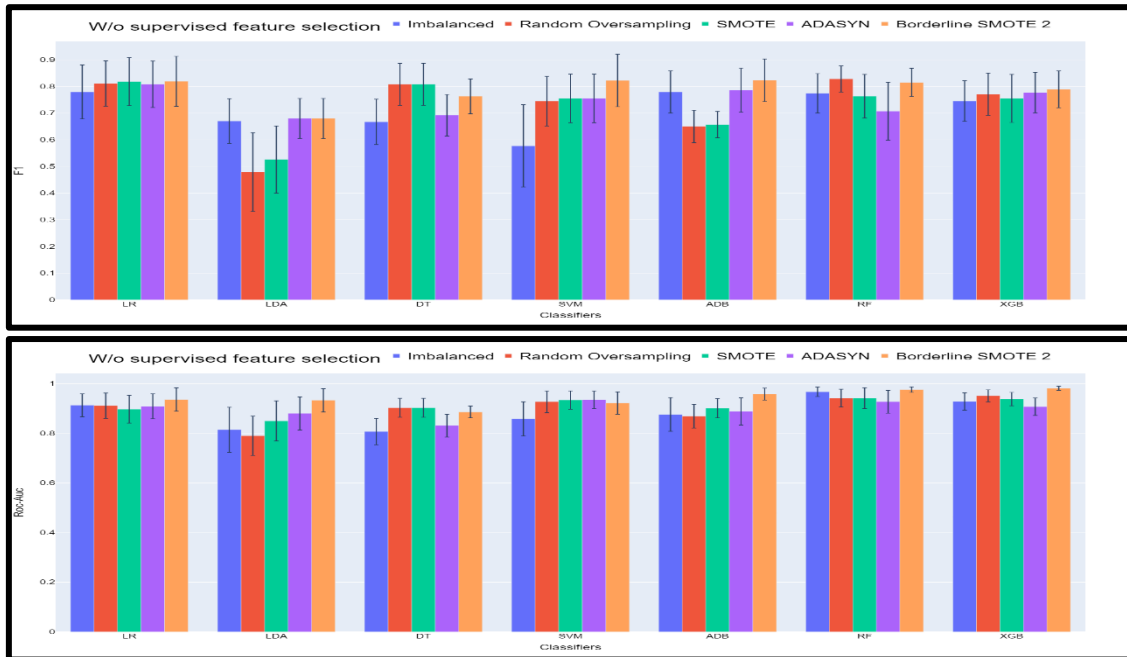


Fig. 4-15 - Negative affect classification results without supervised feature selection (top: F1-score, bottom: Roc-Auc).

Without supervised feature selection, for the F1 scores, Random Oversampling and SMOTE improved the performance of all classifiers, except LDA and ADB. SMOTE also decreased the F1 score of RF. ADASYN increased the scores for all classifiers apart from RF, while Borderline SMOTE 2 improved all classifiers' performance. The best F1 was achieved by RF and Random Oversampling with a mean score of 0.828 (SE=0.049). Regarding the Roc-Auc scores, Random Oversampling did not improve the evaluation scores of LR, LDA, ADB, and RF, SMOTE decreased the performance of LR and RF, while ADASYN did not improve LR, RF, and XGB performance. Borderline SMOTE 2 displayed improvements with all classifiers. The best Roc-Auc score was achieved by XGB with Borderline SMOTE 2 with a mean score of 0.981 (SE=0.008).

Borderline SMOTE 2 was again very constant and, for both metrics, achieved the best performances of all techniques with LR, LDA, ADB, and XGB; the highest Roc-Auc score with RF, and the highest F1 with SVM. Contemplating the times in which the data balance techniques achieved the best scores in a classifier, they can be arranged in descending order as Borderline SMOTE 2, Random Oversampling, and SMOTE together with ADASYN.

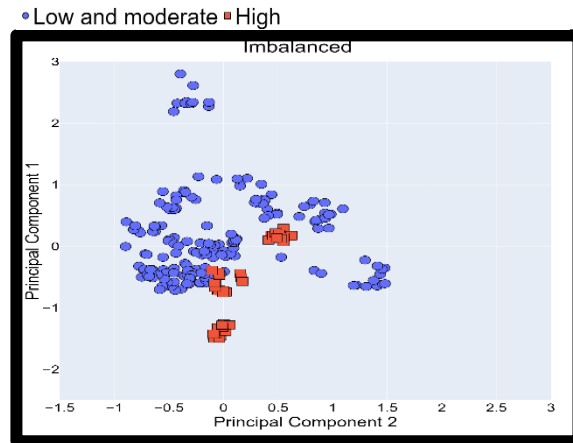


Fig. 4-16 - PCA analysis of the negative affect classes.

From the distribution of the data without supervised feature selection, on the PCA analysis in Fig. 4-16, we can see that the class overlap is small, and there is not much noise or small disjunctions. It can also be observed that the imbalance of classes is more accentuated than the imbalance between the anxiety classes and that the class overlap is a little enhanced by the data balance techniques (as noticed in Fig. 4-17). It can be perceived that the decision boundary becomes more clearly defined, especially with Borderline SMOTE 2, as shown in Fig. 4-17.

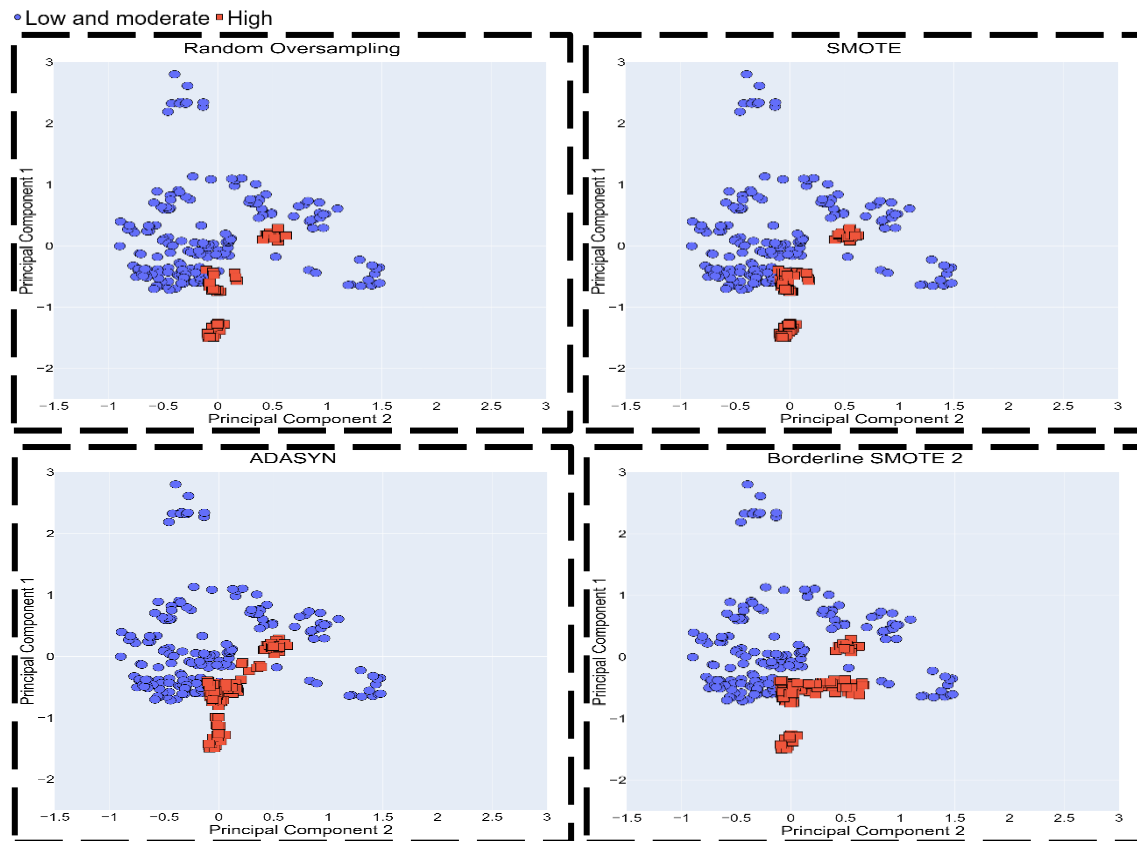


Fig. 4-17 - PCA analysis of the data balance techniques applied to the negative affect classes.

Regarding the supervised feature selection techniques, the combinations for the best scores reached are presented for SelectKBest in Table 6-7 and RFECV in Table 6-8.

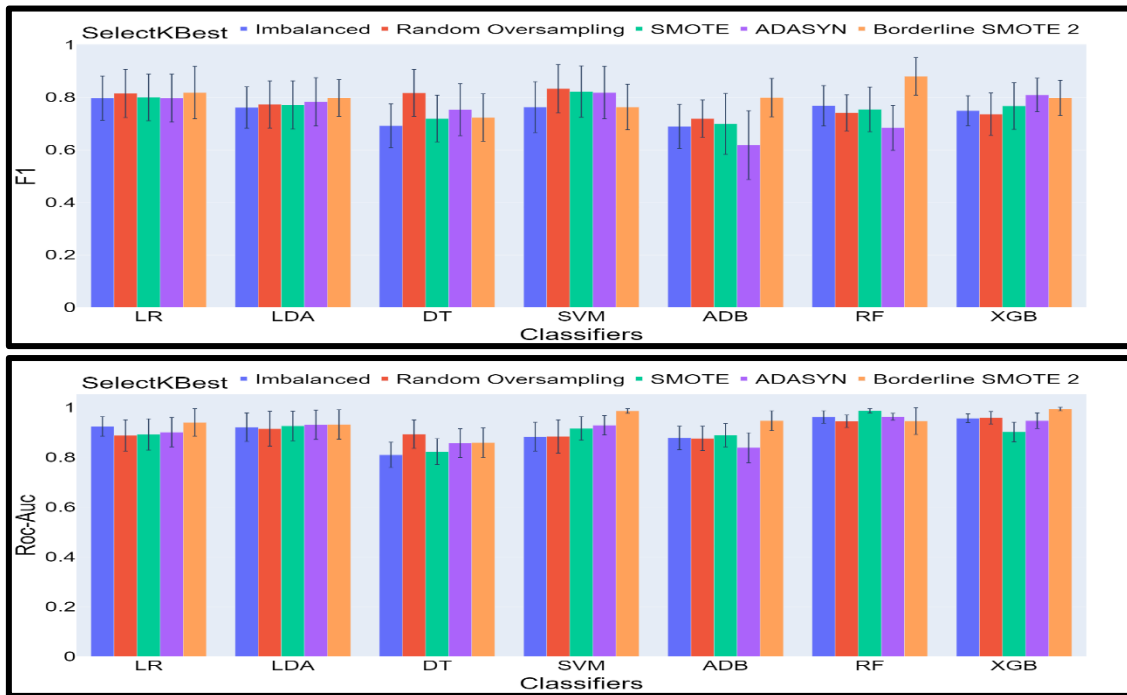


Fig. 4-18 - Negative affect classification results using SelectKBest (top: F1-score, bottom: Roc-Auc).

Analysing the results attained using with SelectKBest, in Fig. 4-18, for the F1 scores, Random Oversampling did not improve the performance of RF and XGB, SMOTE did not show to increase the performance of RF, ADASYN did not improve LR (where equal mean score with a higher standard error was achieved), RF and ADB evaluation results, and Borderline SMOTE 2 improved the performance of all classifiers except SVM (where an equal mean score with a lower standard error was obtained). The best F1 result was achieved by RF and Borderline SMOTE 2, with a mean score of 0.880 (SE=0.072). Regarding Roc-Auc, Random Oversampling did not improve the performance of LR, LDA, ADB, and RF, SMOTE did not increase the performance of LR and XGB, while ADASYN showed to not improve the performance of LR, ADB, and XGB, and Borderline SMOTE 2 showed improvements with all classifiers except with RF. The best Roc-Auc result was achieved by XGB, and Borderline SMOTE 2 with a mean of 0.993 (SE=0.007).

From these results, we can see that Borderline SMOTE 2 was the most consistent technique since it improved both metrics in all classifiers. Borderline SMOTE 2 reached the best results in both metrics with LR, LDA, and ADB, the highest Roc-Auc score with SVM and XGB, and the best F1 score with RF. SMOTE attained the highest

Roc-Auc for RF. ADASYN was the best technique with XGB in F1. Random Oversampling was the best performing technique with DT in both metrics and with SVM in F1.

Ranking the combination of SelectKBest and the data balance techniques, based on the number of classifiers that achieved the best results in a classifier, we obtain the following descending order: Borderline SMOTE 2, Random Oversampling, and SMOTE together with ADASYN. The combinations that attained the best scores for each classifier are presented in Table 6-7.

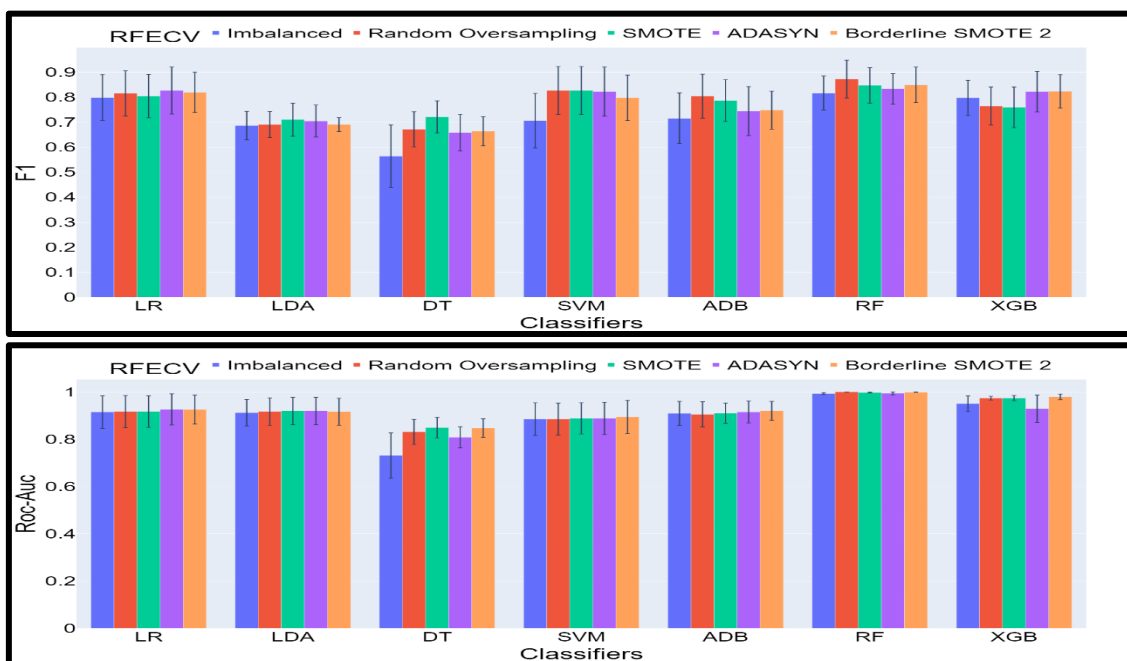


Fig. 4-19 - Negative affect classification results using RFECV (top: F1-score, bottom: Roc-Auc).

When using RFECV, and concerning F1, as shown in Fig. 4-19, ADASYN and Borderline SMOTE 2 showed improvements with all classifiers. Random Oversampling and SMOTE did not increase F1 with XGB. The highest F1 results were reached with RF and Random Oversampling with a mean of 0.872 (SE=0.076). Regarding Roc-Auc, Random Oversampling did not show enhancements in the performance of SVM obtaining an equal mean score with a smaller standard error (0.885 (SE=0.068)), and with ADB, attaining a decreased score. ADASYN did not demonstrate improvements in the evaluation scores with XGB. SMOTE and Borderline SMOTE 2 showed improvements with all classifiers. The best Roc-Auc results were with RF and Random Oversampling with a mean of 1.000 (SE=0.000).

Overall, SMOTE and Borderline SMOTE 2, were the techniques with a more consistent score improvement. Listing the combination of RFECV and the data balance techniques, accounting for the total number of times they reached the highest scores, we got: SMOTE, Borderline SMOTE 2 together with Random Oversampling, and lastly ADASYN (the best combinations are shown in Table 6-8).

It was also noticed that some data balance techniques did not improve either Roc-Auc or F1: without the use of supervised feature selection techniques, this was the case for Random Oversampling with LDA, ADB, and, for SMOTE and ADASYN with RF; using SelectKBest, this happened for Random Oversampling with RF, as well as for ADASYN with LR and ADB; through RFECV, with all classifiers, at least one metric was improved with Random Oversampling, SMOTE, ADASYN, and Borderline SMOTE 2.

The situations in which only one or neither metric was increased, may be justified by the introduction of noise, class overlap, or even overfitting of the classifiers.

In general, the best F1 scores were achieved with the following combinations: without supervised feature selection, with RF and Random Oversampling with a mean score of 0.828 (SE=0.049); with SelectKBest with RF and Borderline SMOTE 2 with a mean score of 0.880 (SE=0.072); using RFECV with RF and Random Oversampling with mean 0.872 (SE=0.076). The top global combination was SelectKBest with RF and Random Oversampling.

Globally, the best Roc-Auc scores were achieved with the next combinations: without supervised feature selection, with XGB and Borderline SMOTE 2, with a mean score of 0.981 (SE=0.008); using SelectKBest, with XGB and Borderline SMOTE 2, with a mean score of 0.993 (SE=0.007); with RFECV, with RF and Random Oversampling, with a mean score of 1.000 (SE=0.000). The best overall combination was RFECV with RF and Random Oversampling.

4.3.5. Positive Affect Feature Selection

For positive affect, without any data balance technique, the features selected at least three times, with the supervised feature selection techniques – SelectKBest and RFECV - were considered and are shown in Fig. 4-20, together with their corresponding frequencies.

This analysis allows us to visualize which features could be of greater importance for classifying distinct levels of positive affect, especially low states of positive affect.

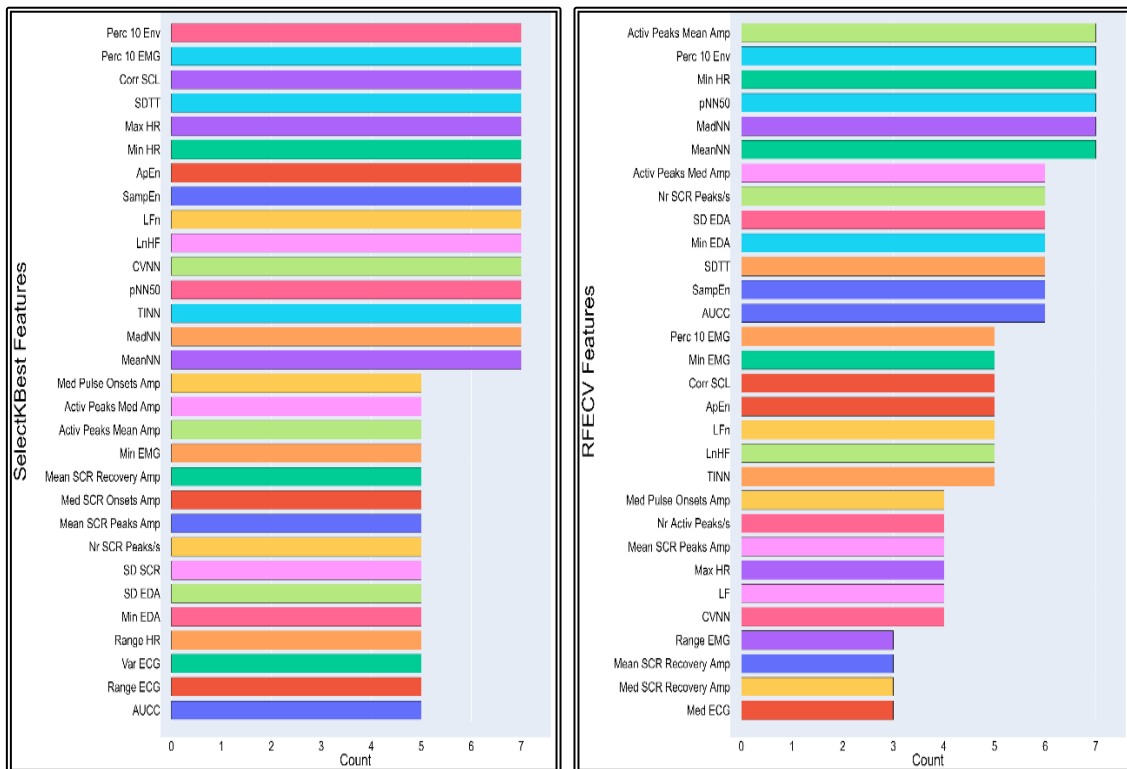


Fig. 4-20 - Features selected at least three times with SelectKBest (on the left) and with RFECV (on the right) for positive affect classification.

From Fig. 4-20, on the left, we can observe the frequency of each of the 30 different selected features with SelectKBest (with a possible minimum of ten features and a possible maximum of 30). A set of 15 features - Perc 10 Env, Perc 10 EMG, Corr SCL, SDTT, Max HR, Min HR, ApEn, SampEn, LFn, LnHF, CVNN, pNN50, TINN, MadNN, and MeanNN - are present in all feature sets of all the classifiers. This shows that EDA has little relevance for this classification problem.

From Fig. 4-20, on the right, we can observe the frequency of each of the 51 different selected features with RFECV (with a possible minimum set size of ten features). The features selected for all the classifiers are Mean Activation Peaks Amp, Perc 10 Env, Min HR, pNN50, MadNN, and MeanNN. For six classifiers the features Activ Peaks Med Amp, Nr SCR Peaks/s, SD EDA, Min EDA, SDTT, SampEn, and AUCC were chosen. For five out of the seven classifiers, Perc 10 EMG, Min EMG, ApEn, LFn, LnHF, and TINN were selected. The most selected features by RFECV reinforce the smaller influence of EDA-extracted features in positive affect classification.

In total, ECG was the most frequent signal used when classifying positive affect, stressing the importance of ECG features for positive affect classification, in alignment

with the results attained in positive affect characterization. EMG was the second most important signal followed by EDA.

4.3.6. Positive Affect Classification

The positive affect classification results with and without supervised feature selection are presented in Table 6-23. Similar to what happened in both previously presented classification tasks, three classifiers did not show a better performance with the supervised feature selection techniques when classifying positive affect.

The results for the classification of positive affect without supervised feature selection techniques, with SelectKBest, and RFECV are presented in Fig. 4-21.

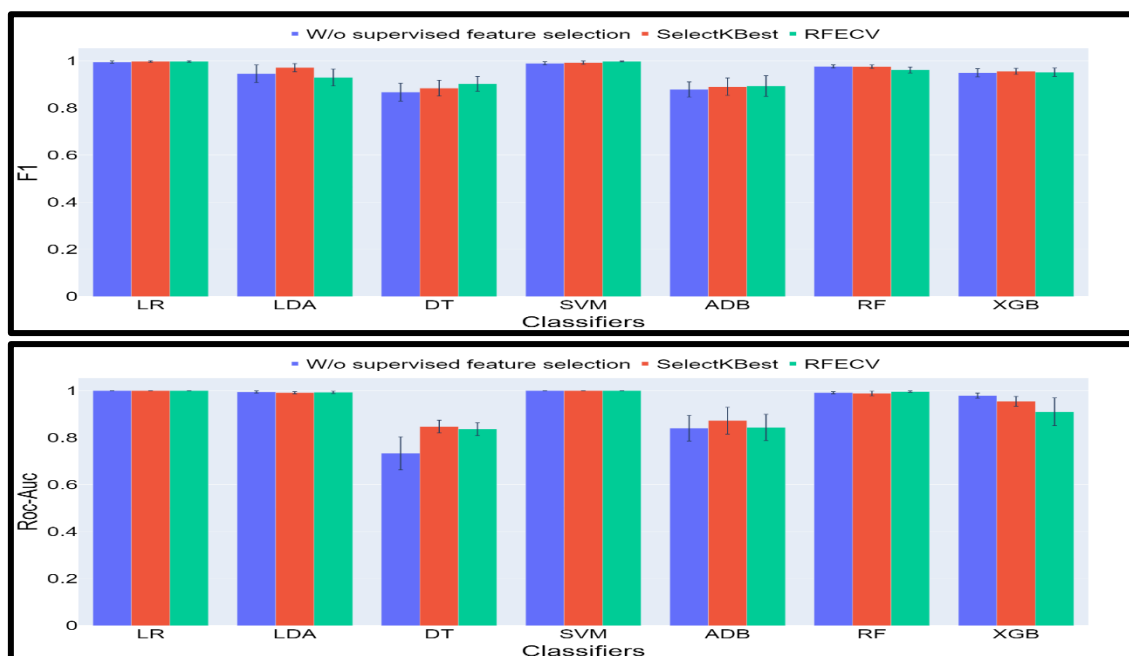


Fig. 4-21 - Positive affect classification results without supervised feature selection, with SelectKBest and RFECV (top: F1-score, bottom: Roc-Auc).

For positive affect classification, LDA, RF, and XGB did not improve their performance with SelectKBest and RFECV.

Comparing the results without supervised feature selection with using SelectKBest or RFECV, the F1 of LR and SVM improved, while Roc-Auc of LR presented the same value (1) and SVM achieved similar values. ADB and DT improved both metrics with SelectKBest and RFECV. LDA with SelectKBest improved F1 but not Roc-Auc and with RFECV it did not improve the F1 or Roc-Auc. RF with SelectKBest did not improve F1 nor Roc-Auc, while after employing RFECV, the Roc-Auc improved but not F1. With

XGB, using SelectKBest and RFECV, the F1 was improved but the Roc-Auc score decreased.

When classifying positive affect, the majority class is the “Low” PA class which is our target class. Therefore, it does not exist imbalance problem and no data balance techniques are required. Nonetheless, different supervised feature selection techniques were applied to assess how different methods can affect the models’ performance for this classification problem. Compared to without supervised feature selection methods, better results were attained when using supervised feature selection with some classifiers while others obtained the worst results.

Without a supervised feature selection technique, the three classifiers that achieved the best Roc-Auc scores were LR, LDA, and SVM. The best Roc-Auc score was achieved by LR and SVM with a mean score of 1.000 (SE=0.000). The three classifiers that achieved the best F1 scores were LR, SVM, and RF. The best F1 was achieved by LR with a mean score of 0.994 (SE=0.005).

Using SelectKBest, concerning the Roc-Auc scores, LR, LDA, and SVM, performed the best, with LR and SVM achieving the highest mean score of 1.000 (SE=0.000). Regarding F1, LR, SVM, and RF performed the best, with the highest mean score being 0.997 (SE=0.003) obtained by LR.

Applying RFECV, the three classifiers that attained the best Roc-Auc scores were LR, SVM, and RF. The best Roc-Auc score was achieved by LR and SVM with a mean score of 1.000 (SE=0.000). Concerning F1, LR, SVM, and RF achieved the best scores. The best score was achieved by LR and SVM with a mean score of 0.997 (SE=0.003).

Overall, regarding both metrics, LR and SVM were the classifiers that achieved the best scores in both supervised feature selection methods, although better performance would be expected by ensemble classifiers (RF, ADB, or XGB).

The linear models – LR and SVM – outperformed all the others. Linear models can overperform tree ensemble models if the data is linearly separable. Concerning the ensemble models, RF outperformed XGB and both performed better than ADB which can indicate the overfitting of these last two classifiers. The DT classifier showed similar scores to the ADB classifier.

4.3.7. Summary of the Classification Results

The main conclusions for anxiety classification are described next:

About the performance of the supervised feature selection techniques:

- SelectKBest only did not improve the results of RF and XGB.
- RFECV only did not improve the F1 of RF and both metrics of XGB.
- Nonetheless, the score times decreased when using supervised feature selection techniques

Concerning the performance of the combination with data balance techniques:

- Overall, Borderline SMOTE 2 was the most consistent technique.
- Without supervised feature selection, Borderline SMOTE 2 was the most successful technique - achieving more times the highest scores. With supervised feature selection the combinations with Borderline SMOTE 2 and with ADASYN, were the most successful.

The main conclusions for negative affect classification are the following:

Concerning the performance of the supervised feature selection techniques:

- SelectKBest only did not improve the results of ADB and RF.
- RFECV only did not improve the results of ADB.
- The score times decreased with the supervised feature selection techniques

Concerning the performance of the data balance techniques:

- Overall, Borderline SMOTE 2 was the most consistent technique.
- Without supervised feature selection and SelectKBest, Borderline SMOTE 2 was the most successful. With RFECV, the combination with SMOTE was the most successful.

The main conclusions for positive affect classification are:

- With supervised feature selection, only one metric of LDA, RF, and XGB improved. With and without supervised feature selection, the classifiers that performed better were:

- Roc-Auc: LR and SVM with and without SelectKBest and RFECV.
- F1: LR with and without SelectKBest and RFECV
- The linear models – LR and SVM – outperformed the others despite better results being expected from the ensemble classifiers.

5. Conclusion

Our study used the WESAD dataset and the combination of three physiological signals, ECG, EDA, and EMG, to characterize and classify different states of anxiety and affect, with special importance on more severe and less frequent states of these mental conditions. The physiological signals were processed, and relevant features were extracted to characterize different states of anxiety and affect. Two different supervised feature selection algorithms, seven machine learning algorithms, and four data-balance techniques were used to classify anxiety and negative and positive affect.

5.1. Contributions

Anxiety, negative affect, and positive affect were addressed as three independent characterization and classification tasks and we were able to draw some main conclusions regarding the different states of these mental conditions.

Concerning the characterization of anxiety, the study revealed that only the Median EMG was able to characterize anxiety. The Median amplitude of the EMG signal showed that higher states of anxiety are related to higher median values of the Median amplitude of the EMG signal.

When characterizing negative affect, we concluded that it can be characterized using the EMG signal, through the Maximum EMG and the Number of Activation Peaks per second. We were able to observe that higher levels of negative affect are associated with a smaller maximum amplitude of the EMG signal and a higher frequency of activation peaks. Additionally, negative affect can be described using statistical and peak features from the SCR component. The Number of SCR Peaks per second showed increased values in higher negative affect levels, indicating higher electrodermal response activity.

When characterizing different states of positive affect, we determined that it can be distinguished by the features that measure the time interval between normalized R peaks and the features that describe the variation of those time intervals. From those features, we concluded that in lower positive affect levels, the intervals between R peaks are higher, reflecting a lower HR and that the variability of time intervals is higher, suggesting a higher HRV. Positive affect is also characterized by heart rate features which corroborate a lower HR associated with lower levels of positive affect. Furthermore, the median EMG linear envelope also indicated a higher muscular tension

associated with lower positive affect levels and concerning the EDA signal, the SCL component showed that lower positive affect states are associated with decreased electrodermal activity.

Concerning the classification tasks, we were able to classify the different states of anxiety and affect.

Regarding the features chosen through the application of the supervised feature selection techniques for the three classification tasks, we concluded that for all three classification tasks, all physiological signals had importance, especially features extracted from the ECG signal. The supervised feature selection techniques improved in general the performance of most classifiers in all three classification tasks.

Respecting the precision and recall scores, for all three classification tasks, regardless of feature selection, the recall scores were, in general, higher than the precision scores. For the three classification tasks, most recall and precision scores are improved when the supervised feature selection techniques were applied.

The balance techniques have also been shown to improve, in general, the classification results. Moreover, in some cases, the addition of class overlaps, noise, and even overfitting of the classifiers, prevented the enhancement of the results. Borderline SMOTE 2 showed to achieve the most consistent results in both anxiety and negative affect classification. Analysing the precision and recall when using data balancing, without supervised feature selection and with SelectKBest, is notable an overall pattern of increased precision and recall, and with RFECV of decreased precision and increased recall. Nonetheless in cases where the F1 metric was not improved by the data balancing, either both metrics were decreased or in general precision worsened considerably.

5.2. Limitations and Future work

Two main shortcomings of this dissertation are the reduced number of participants, which restricts the representativeness, and the neutral condition of this protocol, which was not designed specifically for this type of study.

The latest limitation verifies since while being already difficult to achieve a neutral condition itself, the fact that the participants knew they would face different stimuli after the neutral condition may lead them to feel anxious, particularly the ones that were exposed to the stress condition right after the neutral condition. Additionally, the fact that

half the participants were sitting down, and half were standing up during the neutral condition can influence the physiological signals, especially the EMG signal.

Further, we used data collected from a neutral condition, where the predominance of low levels of positive affect and negative affect is due to the absence of affective stimuli [48]. However, PANAS carries some limits when it comes to measuring low arousal states of affect [49], like the ones that may be experienced during a neutral condition. PANAS includes a relatively large number of high-arousal positive and negative items that may not adequately capture low arousal affective states.

Despite the limitations presented, our findings allowed us to successfully characterize and classify anxiety and affect in a neutral condition, providing a deeper understanding of these conditions. We conclude through our outcomes that it is viable to use classification models in applications connected with wearable devices, like watches or smartphones, which can help detect and manage anxiety and affect.

In future work, there are some topics to be considered further in this line of research. At first by considering a protocol specifically designed for attaining neutral conditions and engaging more participants. Also, by addressing anxiety and affect classification tasks from a multiclassification perspective or as ordinal classification tasks, the last would require more data.

Concerning the feature selection techniques used, supervised feature selection such as sequential feature selection or different filter methods may be an option as well.

Regarding the data balance techniques, when facing a data imbalance problem, the use of techniques that combine over/under-sampling methods and the integration of those into machine learning algorithms, are also worth trying. Lastly, the further tuning of the classification balance hyperparameters together with the classifier hyperparameters may further improve the results.

References

- [1] P. Schmidt, A. Reiss, R. Duerichen, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," *ICMI 2018 - Proc. 2018 Int. Conf. Multimodal Interact.*, 2018, doi: 10.1145/3242969.3242985.
- [2] Health and Safety Executive and H. and S. Executive, "Work-related stress, anxiety or depression statistics in Great Britain 2019," *Annu. Stat.*, pp. 1–9, 2019, [Online]. Available: <http://www.hse.gov.uk/statistics/lfs/index.htm>.
- [3] "Stress vs. Anxiety – Knowing the Difference Is Critical to Your Health - Mental Health First Aid." <https://www.mentalhealthfirstaid.org/external/2018/06/stress-vs-anxiety/> (accessed Aug. 28, 2021).
- [4] "Health Status." https://stats.oecd.org/Index.aspx?DatasetCode=HEALTH_STAT (accessed Aug. 16, 2021).
- [5] J. Hyde, K. M. Ryan, and A. M. Waters, "Psychophysiological Markers of Fear and Anxiety," *Curr. Psychiatry Rep.*, vol. 21, no. 7, 2019, doi: 10.1007/s11920-019-1036-x.
- [6] E. Henje Blom, E. M. Olsson, E. Serlachius, M. Ericson, and M. Ingvar, "Heart rate variability (HRV) in adolescent females with anxiety disorders and major depressive disorder," *Acta Paediatr. Int. J. Paediatr.*, 2010, doi: 10.1111/j.1651-2227.2009.01657.x.
- [7] L. M. Shin and I. Liberzon, "The neurocircuitry of fear, stress, and anxiety disorders," *Neuropsychopharmacology*, vol. 35, no. 1, pp. 169–191, 2010, doi: 10.1038/npp.2009.83.
- [8] P. J. Tully, N. J. Harrison, P. Cheung, and S. Cosh, "Anxiety and Cardiovascular Disease Risk: a Review," *Curr. Cardiol. Rep.*, vol. 18, no. 12, 2016, doi: 10.1007/s11886-016-0800-3.
- [9] R. C. Thurston, M. Rewak, and L. D. Kubzansky, "An Anxious Heart: Anxiety and the onset of cardiovascular diseases," *Prog. Cardiovasc. Dis.*, 2013, doi: 10.1016/j.pcad.2013.03.007.
- [10] C. M. M. Licht, E. J. C. De Geus, D. A. Richard Van, and B. W. J. H. Penninx, "Association between anxiety disorders and heart rate variability in the netherlands study of depression and anxiety (NESDA)," *Psychosom. Med.*, 2009,

doi: 10.1097/PSY.0b013e3181a292a6.

- [11] K. G. van der Kooy, H. P. J. van Hout, H. W. J. van Marwijk, M. de Haan, C. D. A. Stehouwer, and A. T. F. Beekman, "Differences in heart rate variability between depressed and non-depressed elderly," *Int. J. Geriatr. Psychiatry*, vol. 21, no. 2, pp. 147–150, 2006, doi: 10.1002/gps.1439.
- [12] R. Sioni and L. Chittaro, "Stress Detection Using Physiological Sensors," 2015.
- [13] P. Schmidt, A. Reiss, R. Dürichen, and K. Van Laerhoven, "Wearable-Based Affect Recognition—A Review," *Sensors 2019, Vol. 19, Page 407*, Sep. 2019, doi: 10.3390/S19194079.
- [14] G. Pinto, J. M. Carvalho, F. Barros, S. C. Soares, A. J. Pinho, and S. Brás, "Multimodal emotion evaluation: A physiological model for cost-effective emotion classification," *Sensors (Switzerland)*, vol. 20, no. 12, pp. 1–13, 2020, doi: 10.3390/s20123510.
- [15] E. Garcia-Ceja, M. Riegler, T. Nordgreen, P. Jakobsen, K. J. Oedegaard, and J. Tørresen, "Mental health monitoring with multimodal sensing and machine learning: A survey," *Pervasive Mob. Comput.*, 2018, doi: 10.1016/j.pmcj.2018.09.003.
- [16] P. Branco, L. Torgo, and R. P. Ribeiro, "A survey of predictive modeling on imbalanced domains," *ACM Comput. Surv.*, 2016, doi: 10.1145/2907070.
- [17] Y. Liu and S. Du, "Psychological stress level detection based on electrodermal activity," *Behav. Brain Res.*, vol. 341, no. November 2017, pp. 50–53, 2018, doi: 10.1016/j.bbr.2017.12.021.
- [18] N. Sharma and T. Gedeon, "Objective measures, sensors and computational techniques for stress recognition and classification: A survey," *Comput. Methods Programs Biomed.*, 2012, doi: 10.1016/j.cmpb.2012.07.003.
- [19] S. S. Panicker and P. Gayathri, "A survey of machine learning techniques in physiology based mental stress detection systems," *Biocybern. Biomed. Eng.*, 2019, doi: 10.1016/j.bbe.2019.01.004.
- [20] B. S. McEwen, "Neurobiological and Systemic Effects of Chronic Stress," *Chronic Stress*, vol. 1, 2017, doi: 10.1177/2470547017692328.
- [21] "What's the difference between stress and anxiety?"

- <https://www.apa.org/topics/stress/anxiety-difference> (accessed Aug. 28, 2021).
- [22] J. Šalkevičius, R. Damaševičius, R. Maskeliunas, and I. Laukienė, “Anxiety level recognition for virtual reality therapy system using physiological signals,” *Electron.*, 2019, doi: 10.3390/electronics8091039.
- [23] 2017 Hyochol Ahn, et al, “Digital Technologies in the Treatment of Anxiety: Recent Innovations and Future Directions,” *Physiol. Behav.*, vol. 176, no. 10, 2017, doi: 10.1007/s11920-018-0910-2.Digital.
- [24] G. Ernst, “Heart-Rate variability — More than Heart Beats?,” vol. 5, no. September, pp. 1–12, 2017, doi: 10.3389/fpubh.2017.00240.
- [25] 2017 Hyochol Ahn, et al, “ECG signatures of psychological stress,” *Physiol. Behav.*, vol. 176, no. 10, pp. 139–148, 2017, doi: 10.1016/j.jelectrocard.2015.08.005.ECG.
- [26] L. J. Julian, “Measures of Anxiety,” 2014, doi: 10.1002/acr.20561.Measures.
- [27] R. Sanmartín *et al.*, “Positive and negative affect schedule-short form: Factorial invariance and optimistic and pessimistic affective profiles in Spanish children,” *Front. Psychol.*, vol. 0, no. MAR, p. 392, Mar. 2018, doi: 10.3389/FPSYG.2018.00392/PDF.
- [28] R. A. Calvo and S. D’Mello, “Affect detection: An interdisciplinary review of models, methods, and their applications,” *IEEE Trans. Affect. Comput.*, 2010, doi: 10.1109/T-AFFC.2010.1.
- [29] K. Gouizi, C. Maaoui, and F. Bereksi Reguig, “Negative emotion detection using EMG signal,” *Proc. - 2014 Int. Conf. Control. Decis. Inf. Technol. CoDIT 2014*, 2014, doi: 10.1109/CoDIT.2014.6996980.
- [30] D. Watson and L. A. Clark, “Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales,” *Oxford Handb. Posit. Psychol. 3rd Ed.*, 1988, doi: 10.1093/oxfordhb/9780199396511.013.69.
- [31] Crawford JR and Henry JD, “The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample,” *Br. J. Clin. Psychol.*, pp. 245–265, 2004.
- [32] T. Ambrona and B. López-Pérez, “A Longitudinal Analysis of the Relationship between Positive and Negative Affect and Health,” *Psychology*, vol. 05, no. 08,

- pp. 859–863, 2014, doi: 10.4236/psych.2014.58097.
- [33] D. Watson, “Intraindividual and Interindividual Analyses of Positive and Negative Affect: Their Relation to Health Complaints, Perceived Stress, and Daily Activities,” *J. Pers. Soc. Psychol.*, 1988, doi: 10.1037/0022-3514.54.6.1020.
- [34] Y. S. Can, B. Arnrich, and C. Ersoy, “Stress detection in daily life scenarios using smart phones and wearable sensors: A survey,” *J. Biomed. Inform.*, 2019, doi: 10.1016/j.jbi.2019.103139.
- [35] L. Sornmo and P. Laguna, *Bioelectrical Signal Processing in cardiac and neurological applications*.
- [36] T. K. Todorova, M. W. Schreiber, and M. Fontecave, *Data, Machine Learning: The Art and Science of Algorithms that Make Sense of*, vol. 10, no. 3. 2020.
- [37] A. A, A. A, and B. A, “Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review,” *J. Biomed. Inform.*, Feb. 2016, doi: 10.1016/J.JBI.2015.11.007.
- [38] Y. S. Can, N. Chalabianloo, D. Ekiz, and C. Ersoy, “Continuous stress detection using wearable sensors in real life: Algorithmic programming contest case study,” *Sensors (Switzerland)*, vol. 19, no. 8, 2019, doi: 10.3390/s19081849.
- [39] “XGBoost Documentation — xgboost 1.5.0-dev documentation.” <https://xgboost.readthedocs.io/en/latest/> (accessed Jun. 28, 2021).
- [40] S. J. Dattagupta and S. Jayanta Dattagupta, “A performance comparison of oversampling methods for data generation in imbalanced learning tasks,” *Nov. Inf. Manag. Sch.*, p. 28, 2018, Accessed: Sep. 15, 2021. [Online]. Available: <https://docs.google.com/viewer?url=https%3A%2F%2Frun.unl.pt%2Fbitstream%2F10362%2F31307%2F1%2FTEGI0396.pdf>.
- [41] G. Douzas, F. Bacao, and F. Last, “Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE,” *Inf. Sci. (Ny)*, 2018, doi: 10.1016/j.ins.2018.06.056.
- [42] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” *J. Artif. Intell. Res.*, 2002, doi: 10.1613/jair.953.
- [43] S. He, H., Bai, Y., Garcia, E., & Li, “ADASYN: Adaptive synthetic sampling

- approach for imbalanced learning. In IEEE International Joint Conference on Neural Networks, 2008,” *IJCNN 2008.(IEEE World Congr. Comput. Intell. (pp. 1322– 1328)*, 2008.
- [44] H. M. Nguyen, E. W. Cooper, and K. Kamei, “Borderline over-sampling for imbalanced data classification,” *Int. J. Knowl. Eng. Soft Data Paradig.*, 2011, doi: 10.1504/ijkesdp.2011.039875.
- [45] G. Lee, B. Choi, H. Jebelli, C. R. Ahn, and S. Lee, “Reference Signal-Based Method to Remove Respiration Noise in Electrodermal Activity (EDA) Collected from the Field,” *Comput. Civ. Eng. 2019 Data, Sensing, Anal. - Sel. Pap. from ASCE Int. Conf. Comput. Civ. Eng. 2019*, pp. 17–25, 2019, doi: 10.1061/9780784482438.003.
- [46] É. de M. R. Ferreira *et al.*, “Stress, anxiety, self-efficacy, and the meanings that physical therapy students attribute to their experience with an objective structured clinical examination,” *BMC Med. Educ. 2020 201*, vol. 20, no. 1, pp. 1–9, Sep. 2020, doi: 10.1186/S12909-020-02202-5.
- [47] D. Makowski *et al.*, “NeuroKit2: A Python toolbox for neurophysiological signal processing,” *Behav. Res. Methods*, vol. 53, no. 4, pp. 1689–1696, Aug. 2021, doi: 10.3758/S13428-020-01516-Y.
- [48] F. Shaffer and J. P. Ginsberg, “An Overview of Heart Rate Variability Metrics and Norms,” *Front. Public Heal.*, 2017, doi: 10.3389/fpubh.2017.00258.
- [49] J. D. Hunter, “Matplotlib: A 2D graphics environment,” *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, 2007, doi: 10.1109/MCSE.2007.55.
- [50] “statistics — Mathematical statistics functions — Python 3.10.5 documentation.” <https://docs.python.org/3/library/statistics.html> (accessed Jul. 17, 2022).
- [51] “SciPy API — SciPy v1.8.1 Manual.” <https://docs.scipy.org/doc/scipy/reference/> (accessed Jul. 17, 2022).
- [52] “1. Supervised learning — scikit-learn 1.1.1 documentation.” https://scikit-learn.org/stable/supervised_learning.html#supervised-learning (accessed Jul. 15, 2022).
- [53] “scipy.stats.mannwhitneyu — SciPy v1.8.1 Manual.” <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.mannwhitneyu.html?highlight=stats> (accessed Jul. 17, 2022).

- [54] “Pipeline — Version 0.9.1.” <https://imbalanced-learn.org/stable/references/generated/imblearn.pipeline.Pipeline.html> (accessed Jun. 11, 2022).
- [55] “imbalanced-learn · PyPI.” <https://pypi.org/project/imbalanced-learn/> (accessed Jul. 17, 2022).
- [56] “sklearn.feature_selection.SelectKBest — scikit-learn 1.0.2 documentation.” https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html?msclkid=6bb4f8b1ae3011ec8037e0b069e7094a (accessed Mar. 28, 2022).
- [57] “sklearn.feature_selection.f_classif — scikit-learn 1.1.1 documentation.” https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.f_classif.html#sklearn.feature_selection.f_classif (accessed Jul. 17, 2022).
- [58] “sklearn.feature_selection.RFECV — scikit-learn 1.0.2 documentation.” https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFECV.html?msclkid=90a2afc1ae3011ec8a4ab8a1dcf666f4 (accessed Mar. 28, 2022).
- [59] G. C. Cawley and N. L. C. Talbot, “On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation,” *J. Mach. Learn. Res.*, vol. 11, pp. 2079–2107, 2010.
- [60] “Python Release Python 3.8.0 | Python.org.” <https://www.python.org/downloads/release/python-380/> (accessed Sep. 29, 2022).
- [61] J. Kraus, R. Roman, L. Lacinová, M. Lamoš, M. Brázdil, and M. Fredrikson, “Imagery-induced negative affect, social touch and frontal EEG power band activity,” *Scand. J. Psychol.*, Dec. 2020, doi: 10.1111/SJOP.12661.
- [62] D. J and P.-G. J., “Positive affect and parasympathetic activity: Evidence for a quadratic relationship between feeling safe and content and heart rate variability,” *Psychiatry Res.*, Nov. 2017, doi: 10.1016/J.PSYCHRES.2017.07.077.
- [63] S. Dockray and A. Steptoe, “Positive affect and psychobiological processes,” 2010, doi: 10.1016/j.neubiorev.2010.01.006.
- [64] I. Määttänen *et al.*, “Positive affect state is a good predictor of movement and

stress: combining data from ESM/EMA, mobile HRV measurements and trait questionnaires,” *Heliyon*, vol. 7, no. 2, p. e06243, Feb. 2021, doi: 10.1016/J.HELIYON.2021.E06243.

6. Attachments

6.1. Features Description

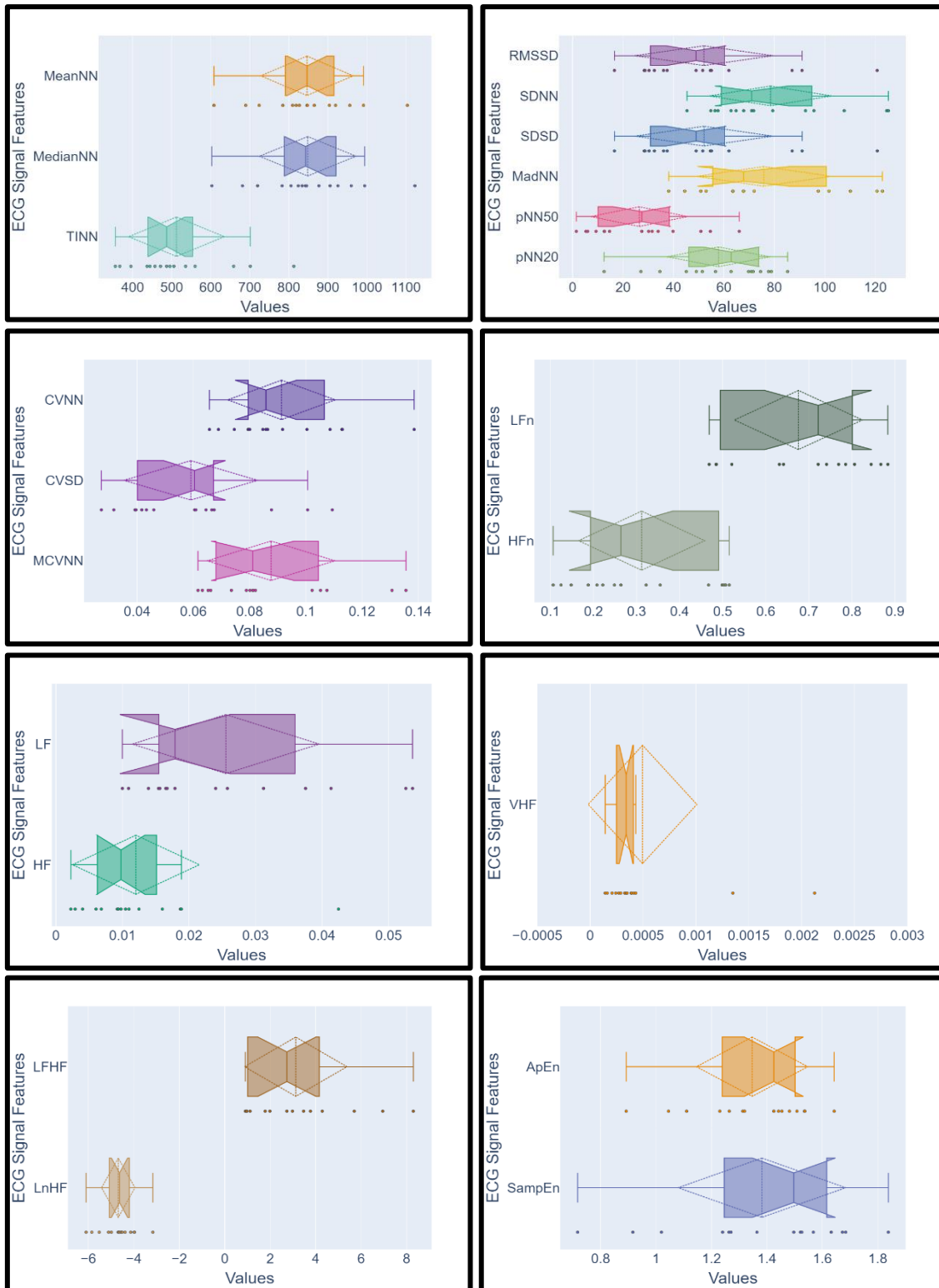


Fig. 6-1 - ECG signal extracted features: HRV time, frequency, and non-linear features.

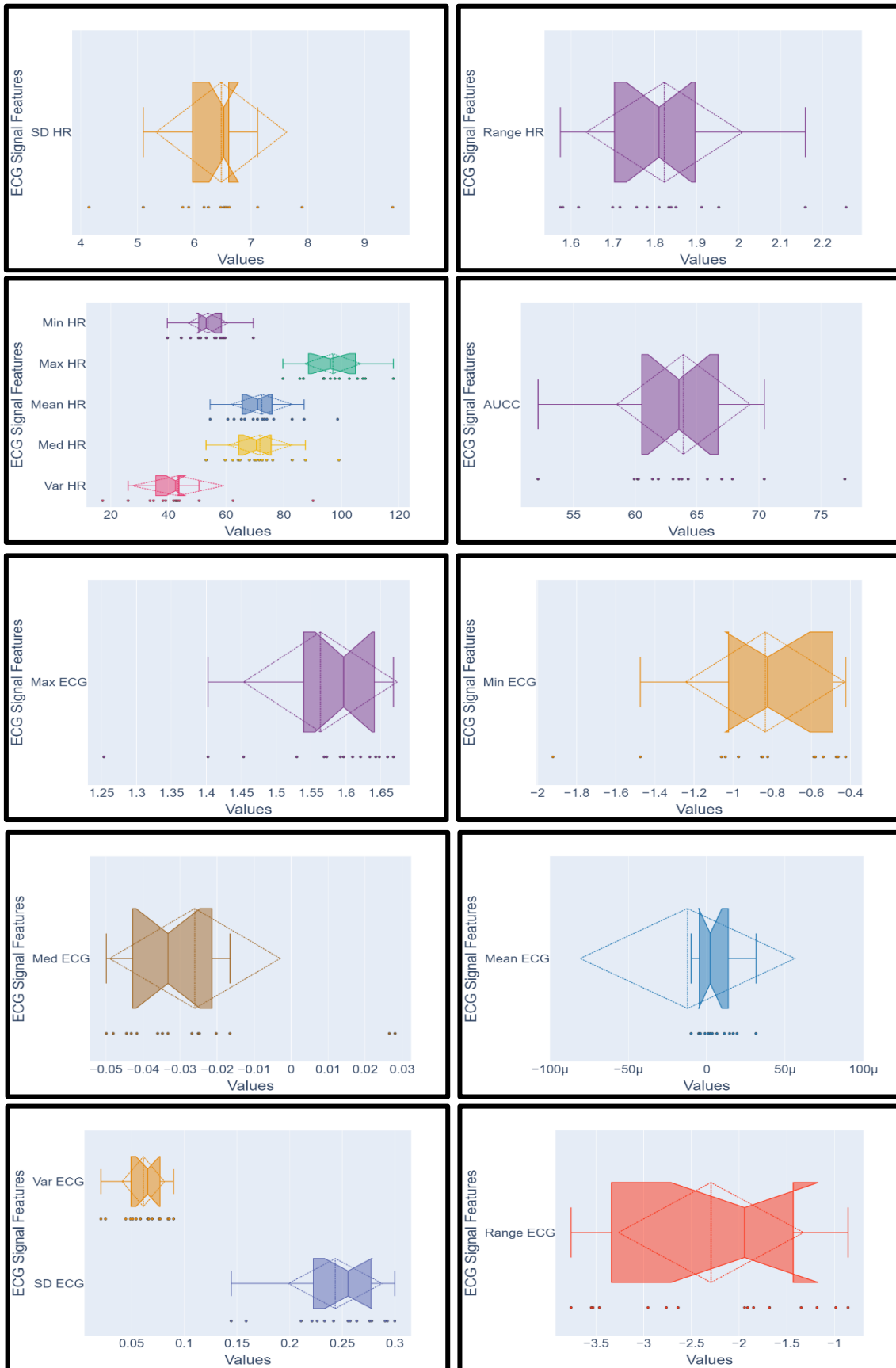


Fig. 6-2 - ECG signal extracted features: HR and ECG signal statistical features.

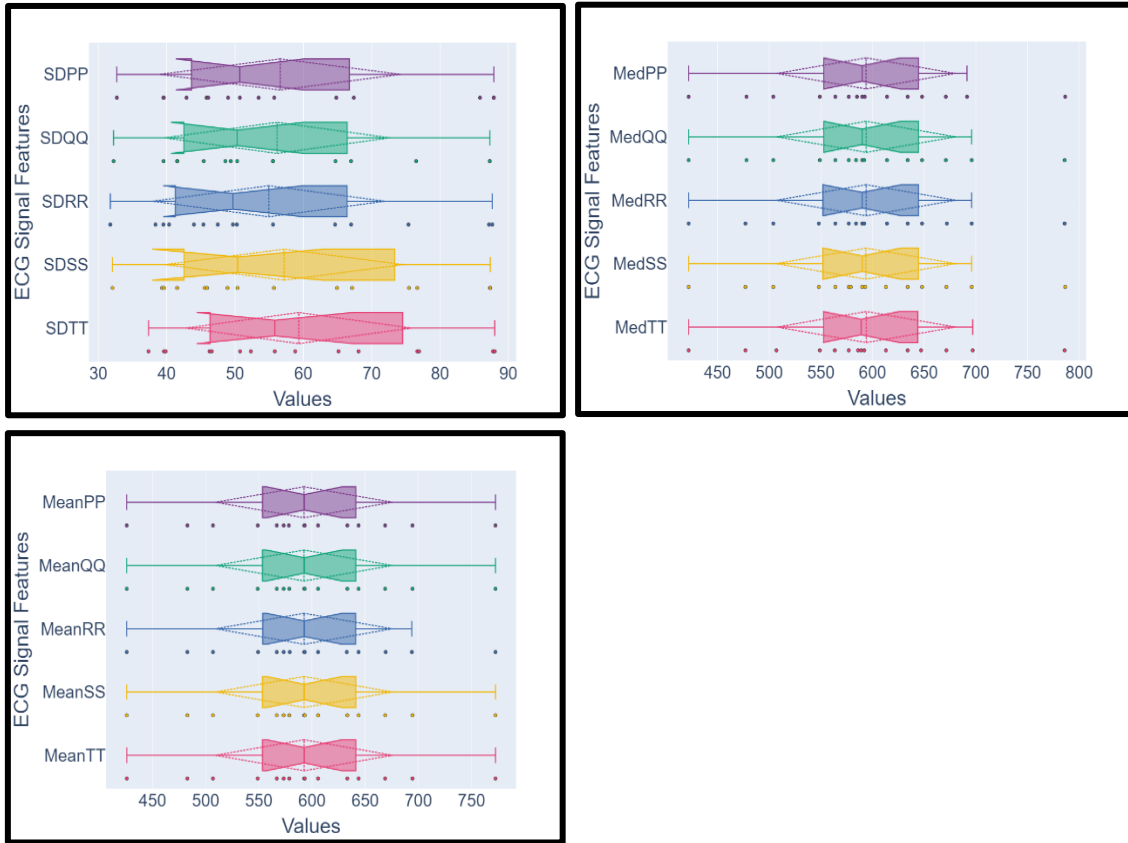


Fig. 6-3 - ECG signal extracted features: non-normalized peaks' features.

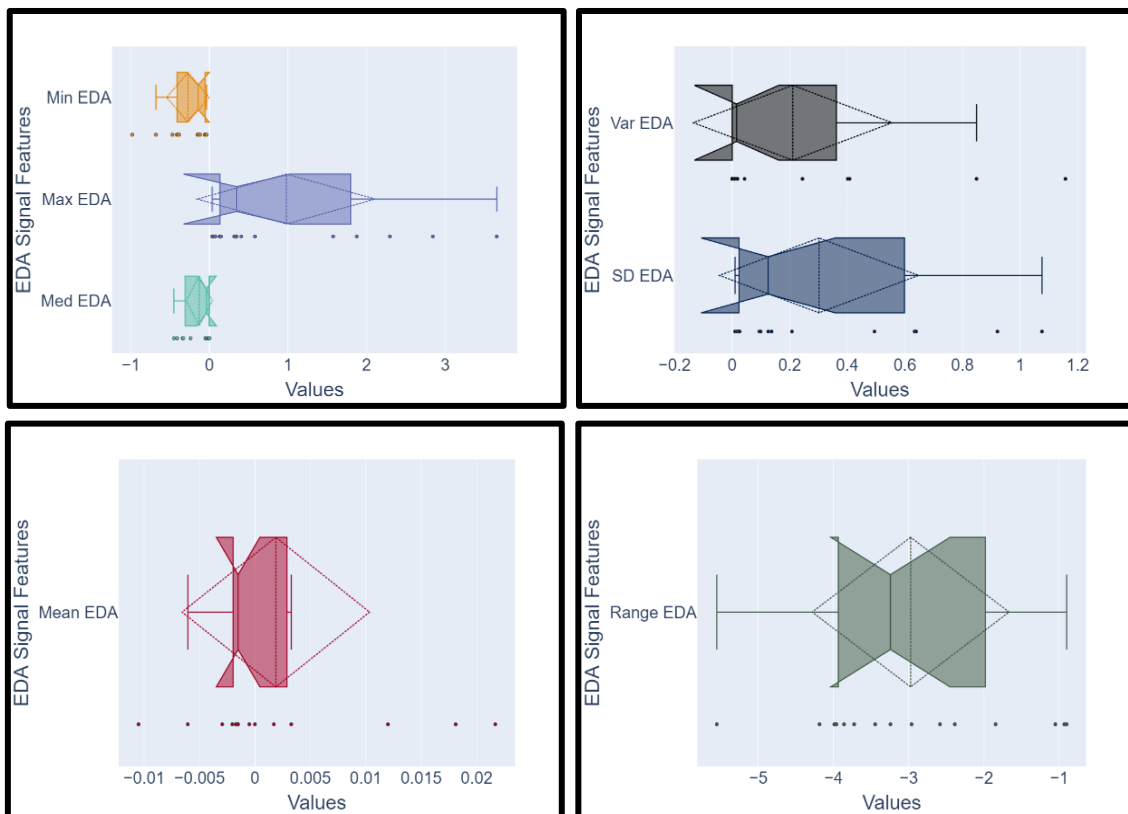


Fig. 6-4 - EDA signal extracted features: EDA signal statistical features.

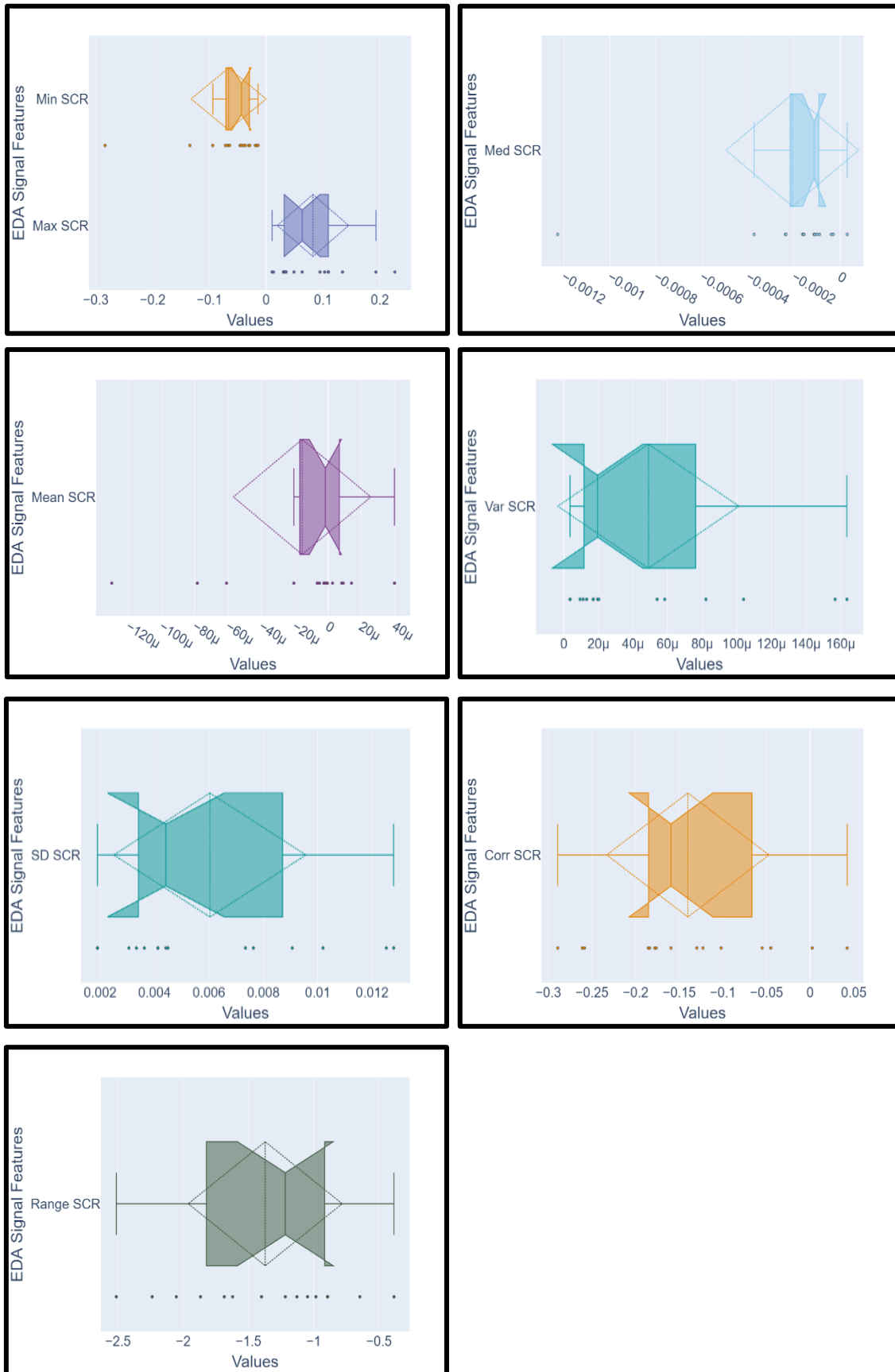


Fig. 6-5 - EDA signal extracted features: SCR statistical features.

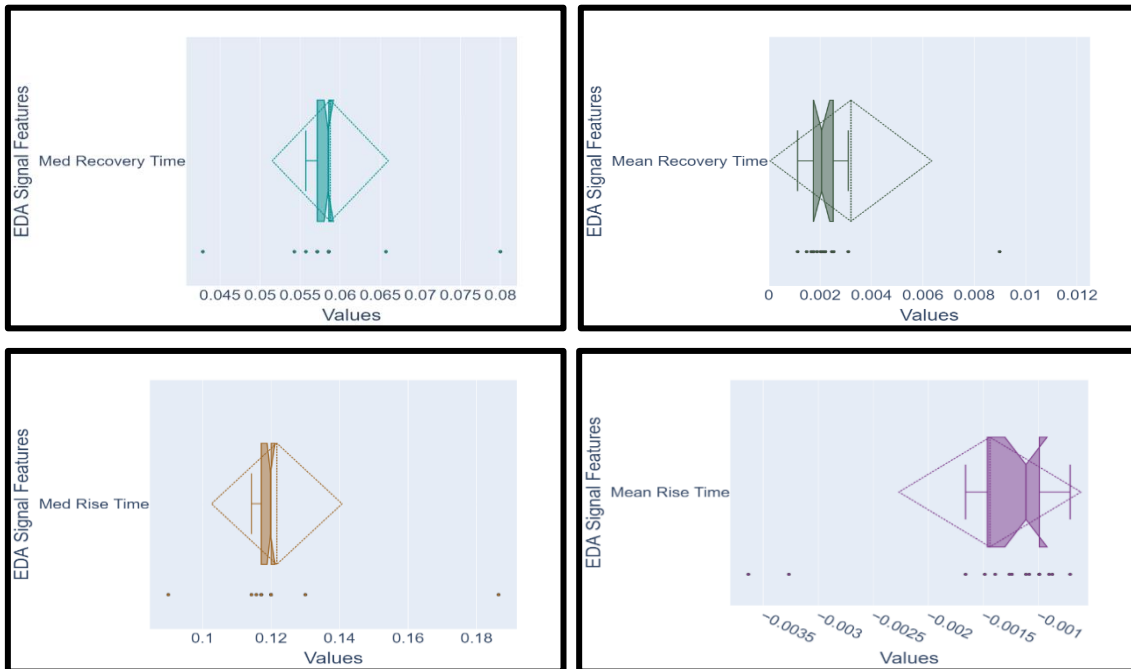


Fig. 6-6 - EDA signal extracted features: mean and median time of the rise and recovery of the SCR peaks.

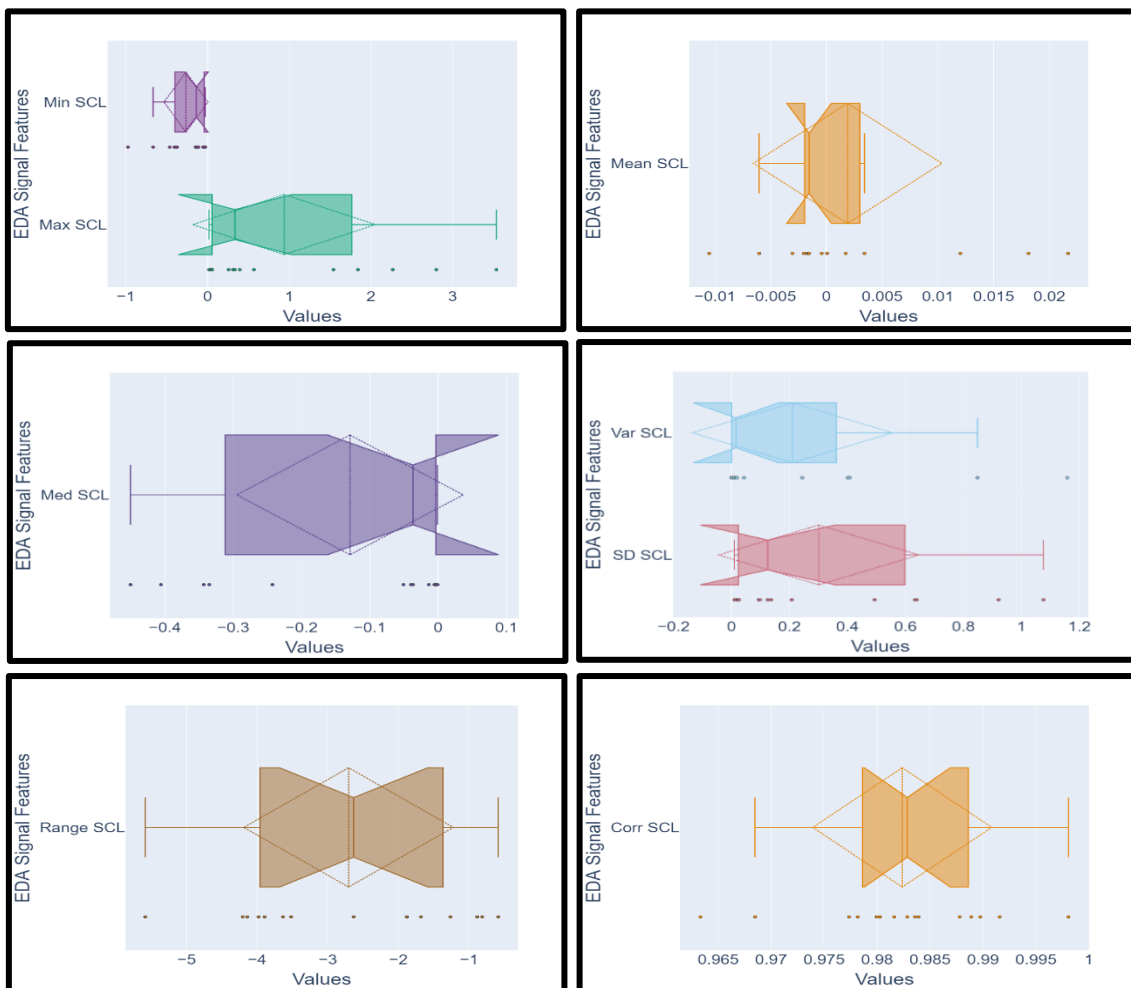


Fig. 6-7 - EDA signal extracted features: SCL statistical features.

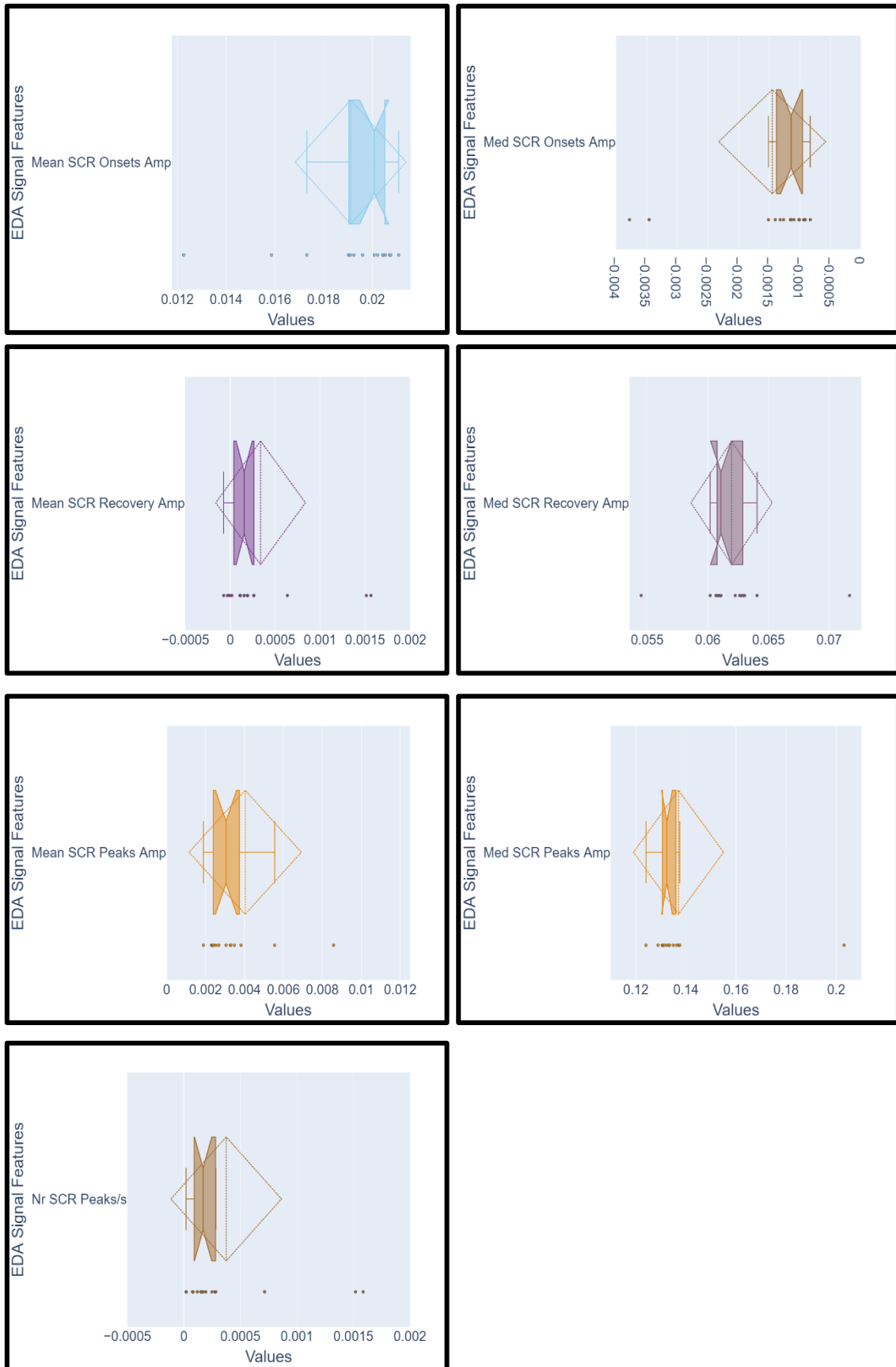


Fig. 6-8 - EDA signal extracted features: SCR peaks' features.

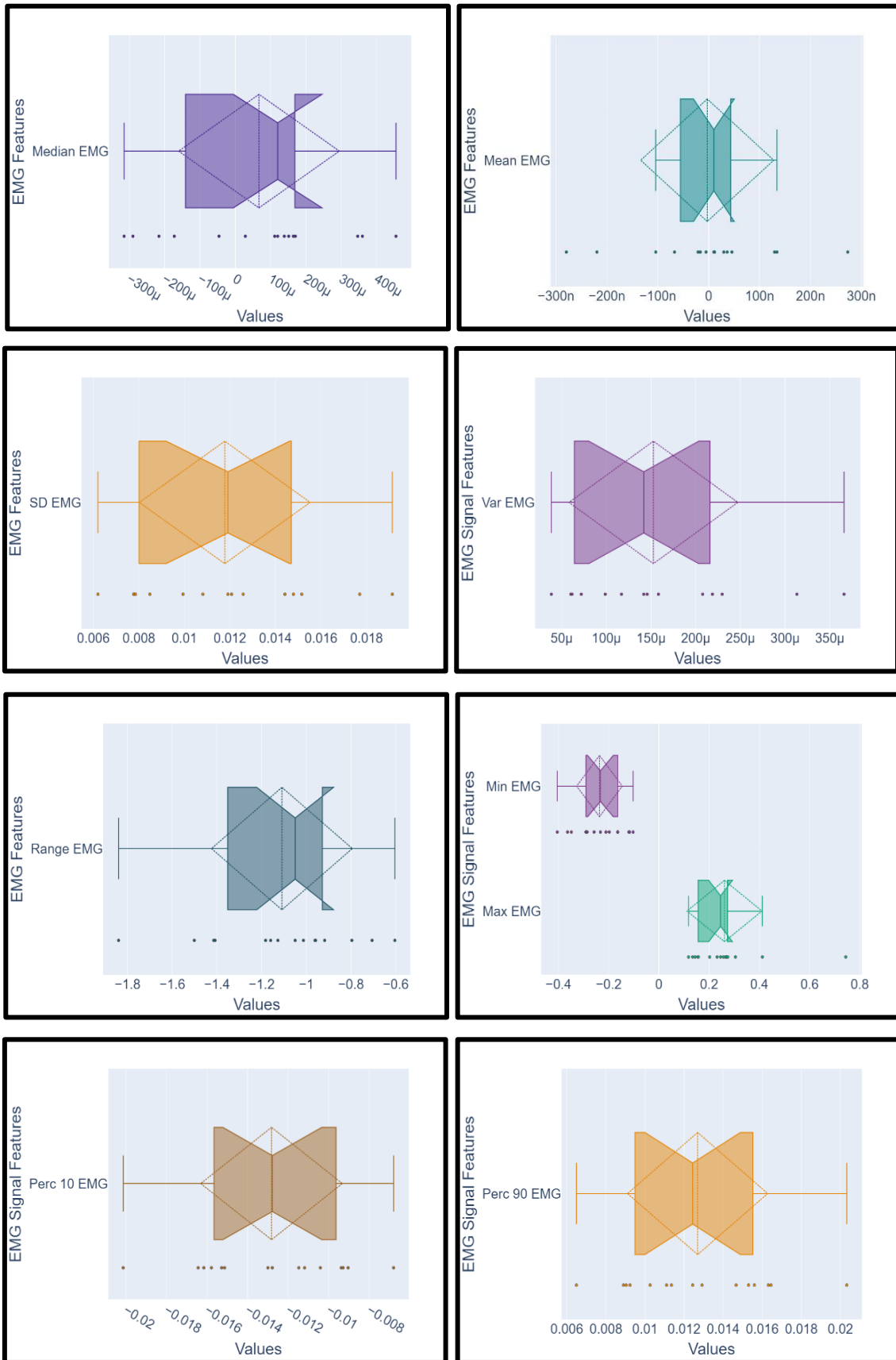


Fig. 6-9 - EMG signal extracted features: statistical features.

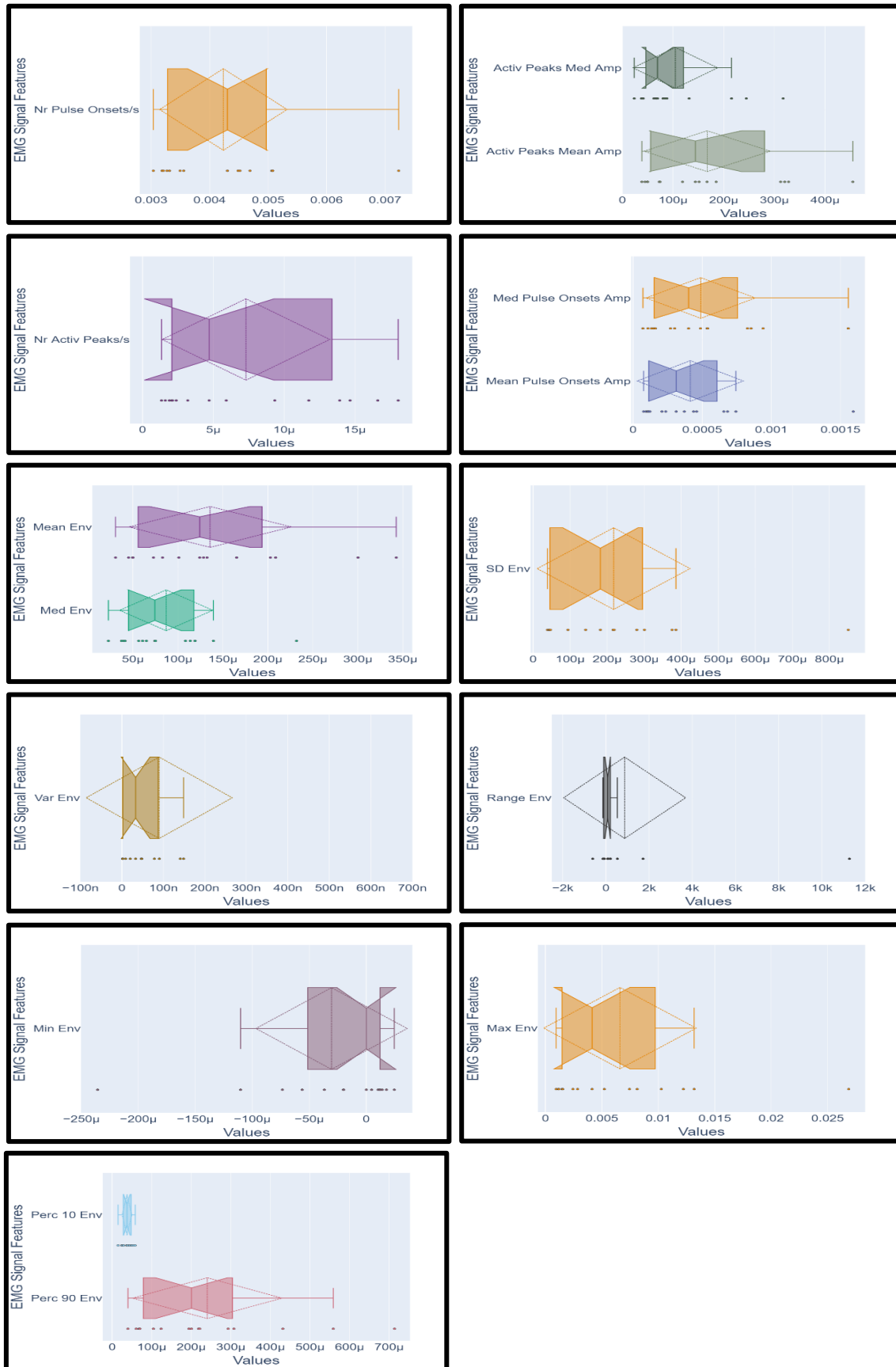


Fig. 6-10 - EMG signal extracted features: EMG envelope statistical features.

6.2. Discriminant negative and positive affect Features

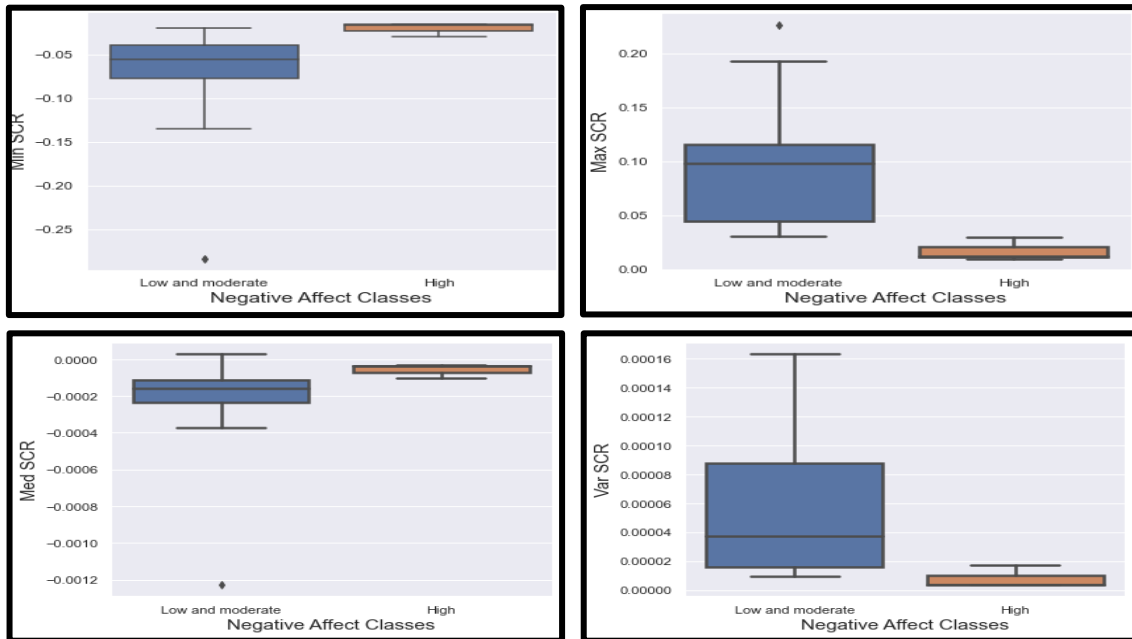


Fig. 6-11 - Boxplots of the EDA signal significant features for negative affect.

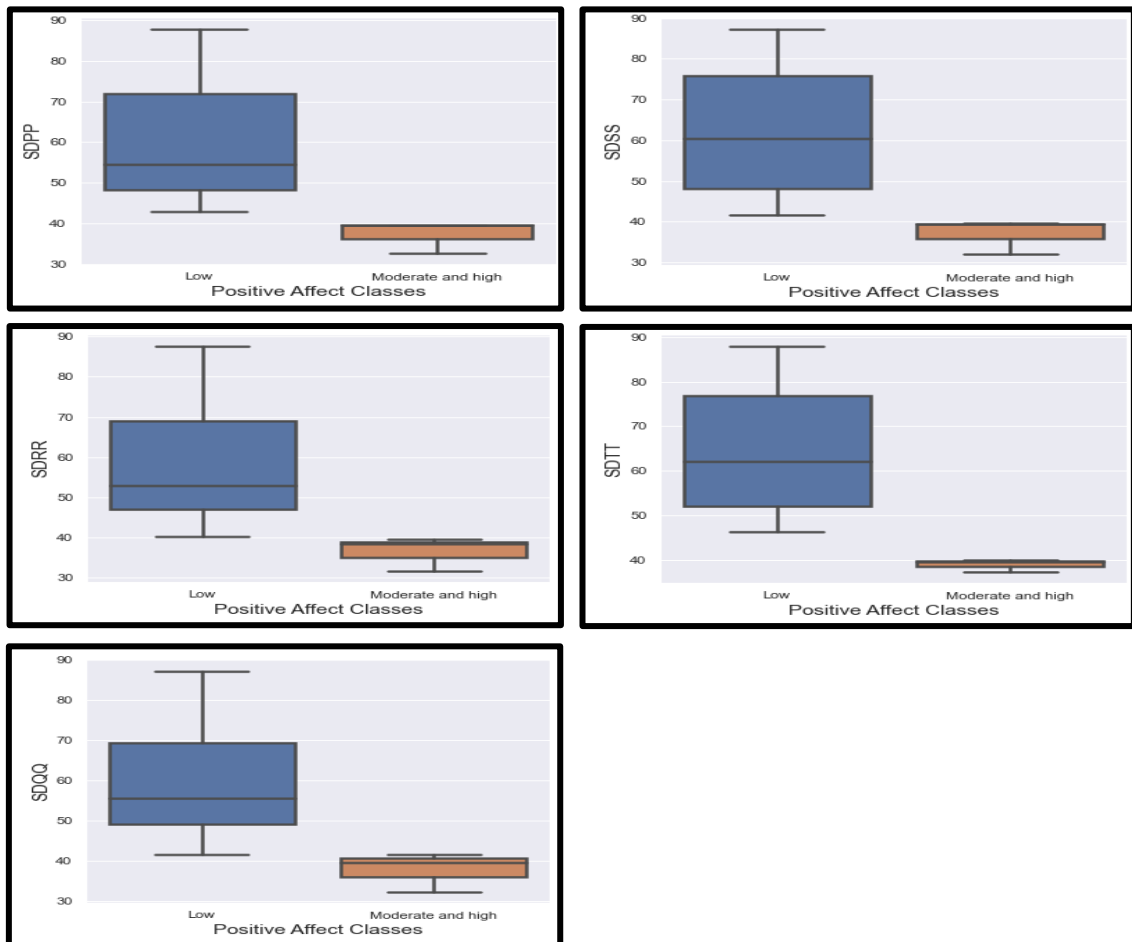


Fig. 6-12 - Boxplots of the Standard deviation of the time interval between ECG peaks, significant for positive affect.

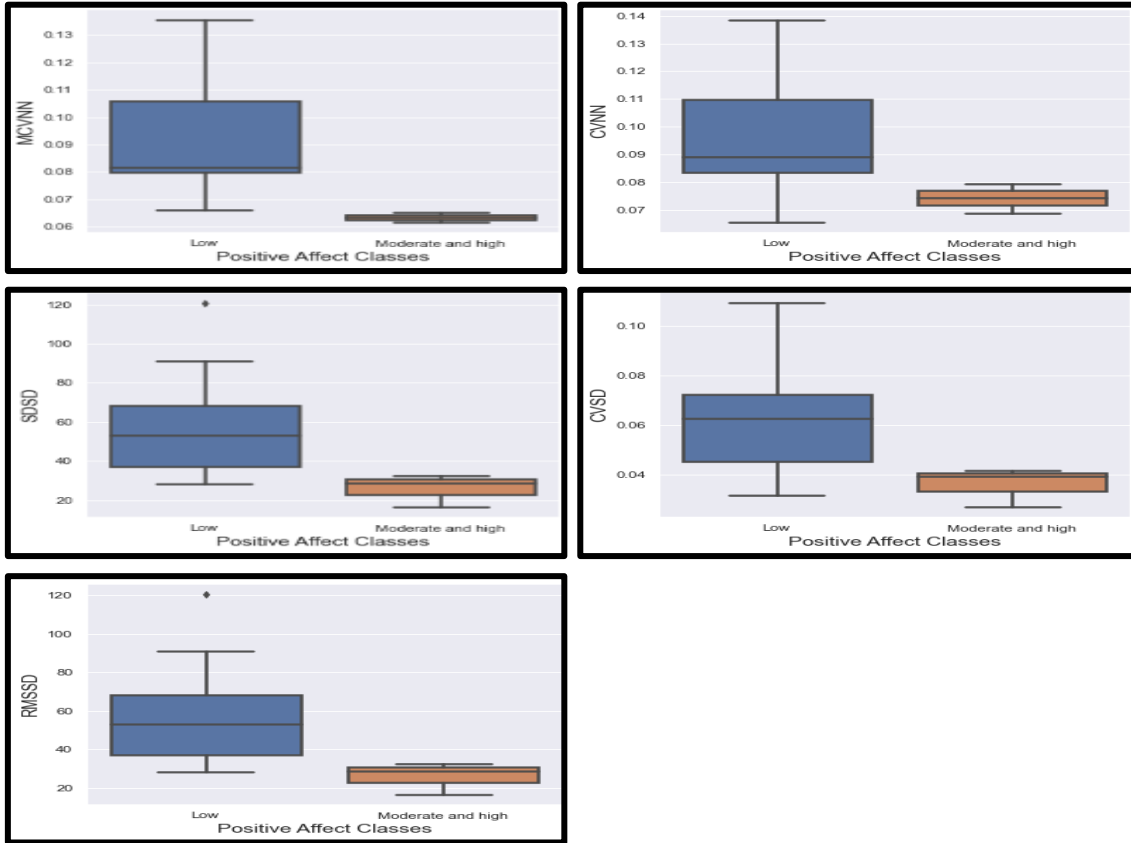


Fig. 6-13 - Boxplots of positive affect significant features that measure the variation of the time interval between R peaks.

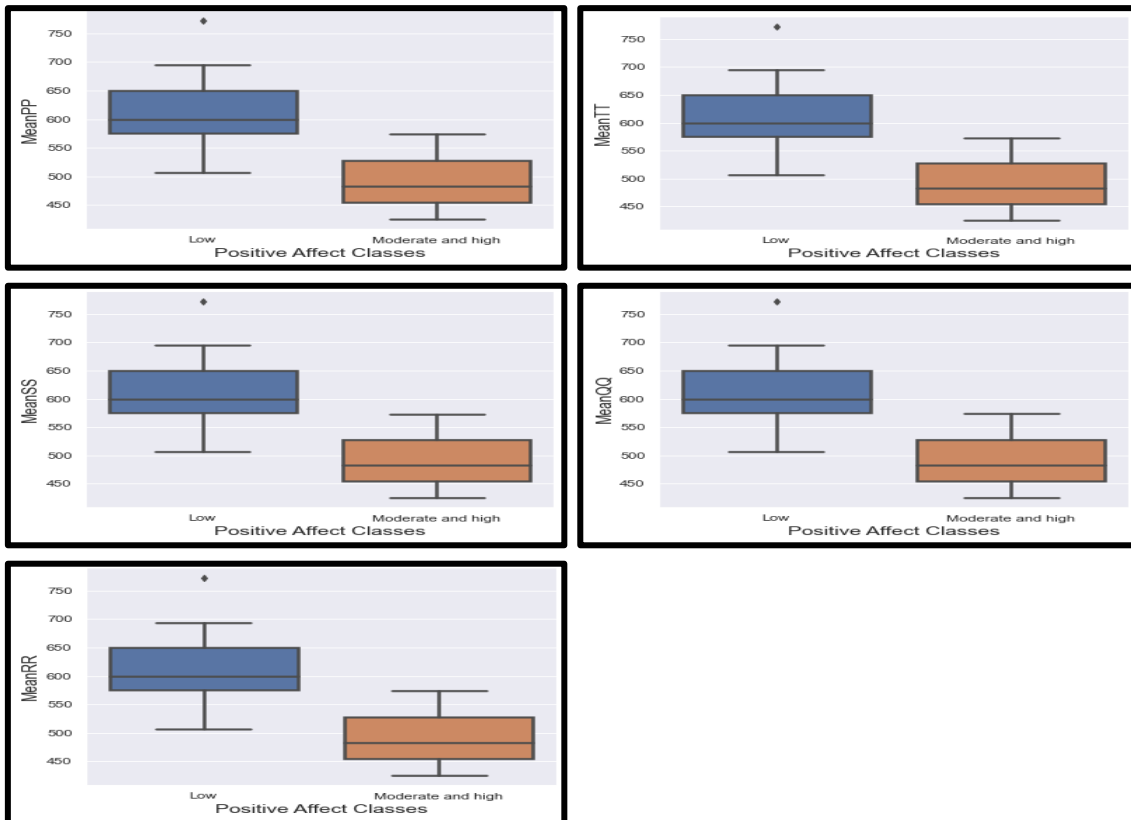


Fig. 6-14 - Boxplots of the features mean interval between ECG peaks, significant for positive affect.

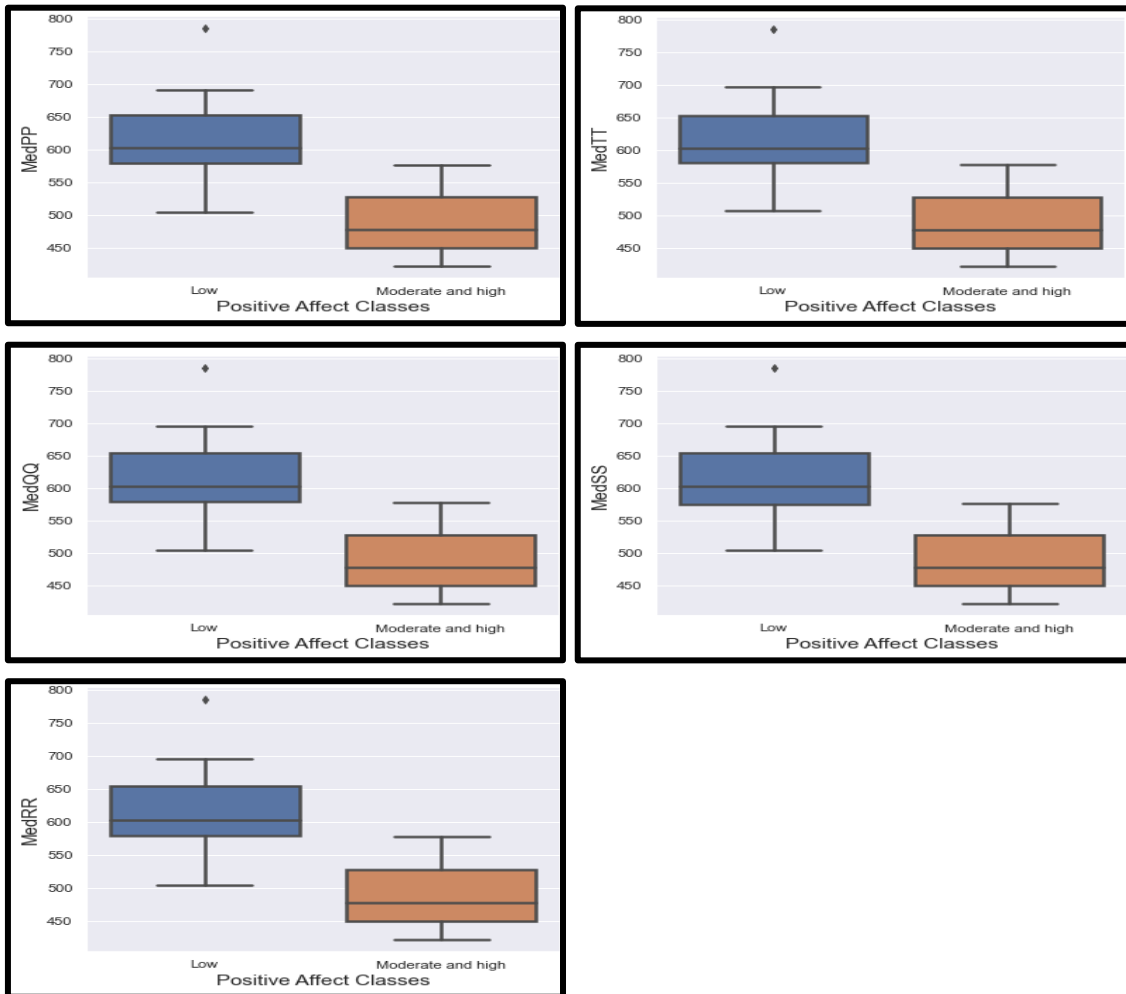


Fig. 6-15 - Boxplots of the median intervals between ECG peaks, significant for positive affect.

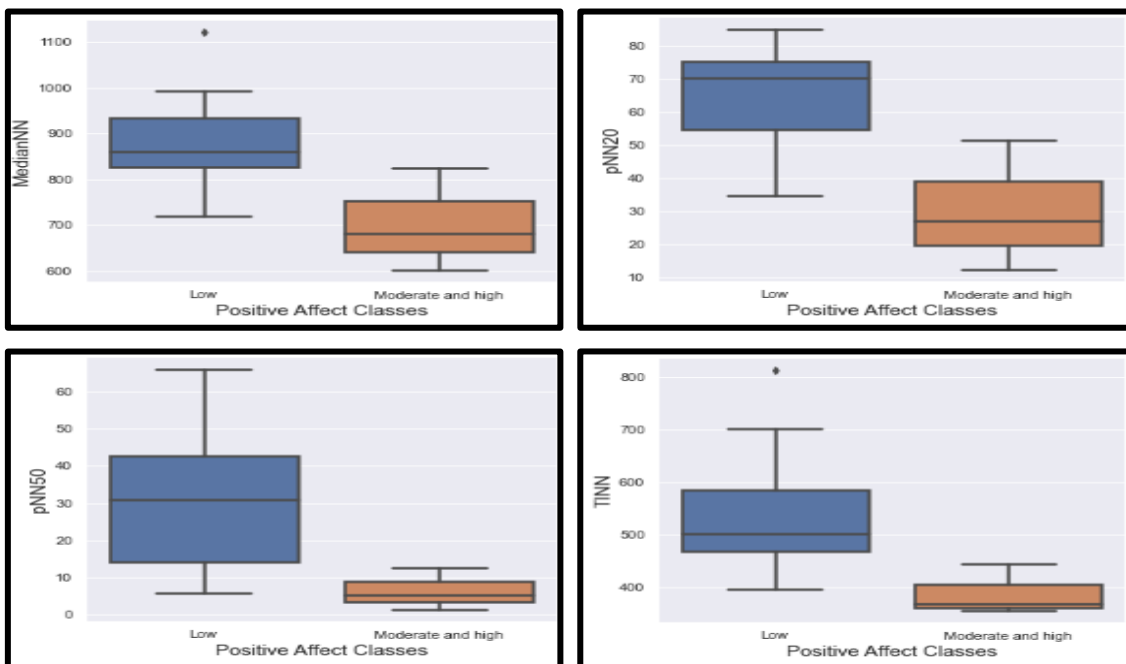


Fig. 6-16 - Boxplots of positive affect significant features that measure the interval between R peaks.

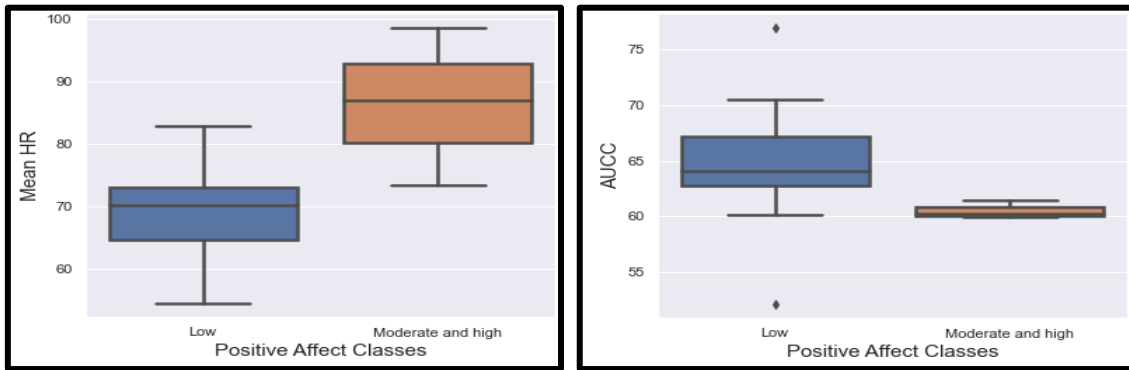


Fig. 6-17 - Boxplots of the significant features Mean HR and AUCC, for positive affect.

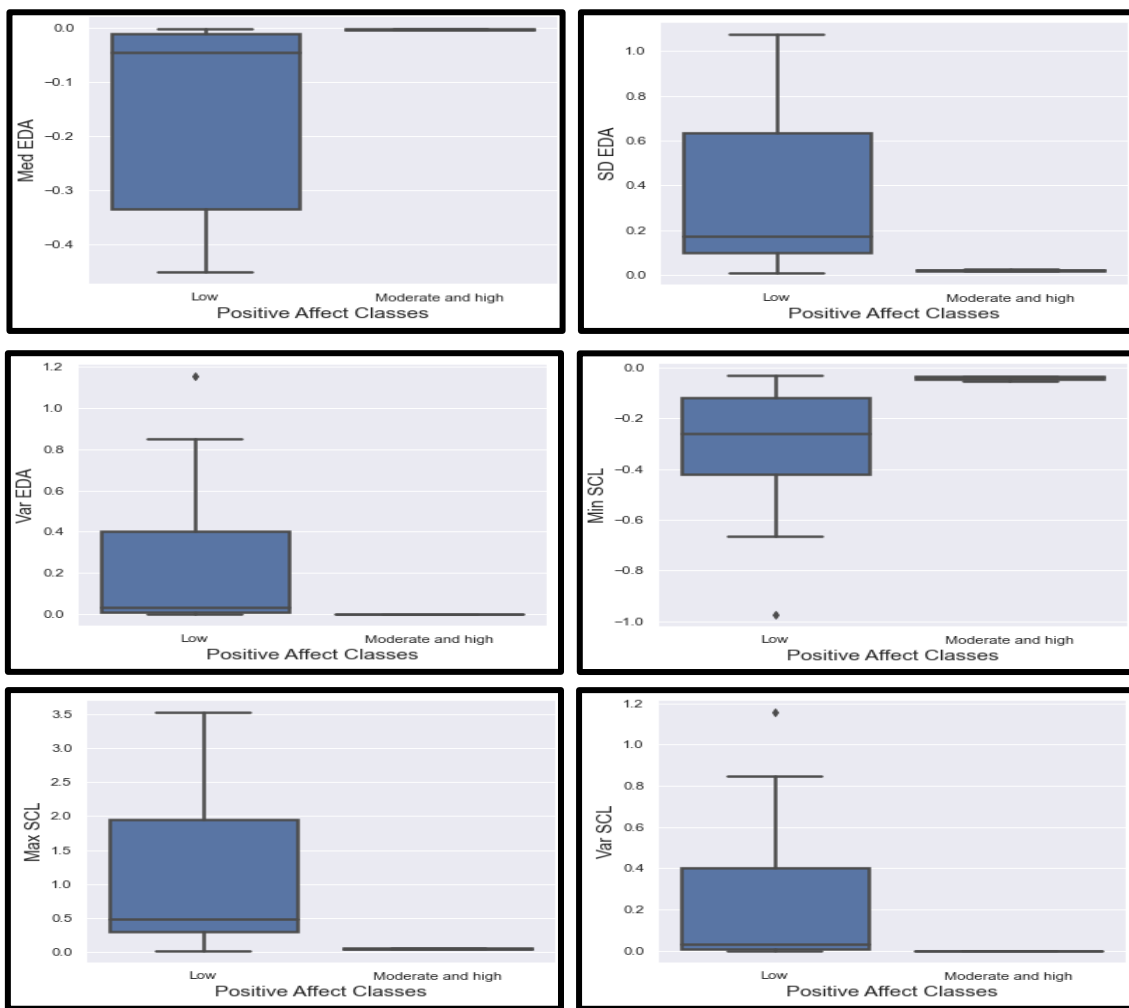


Fig. 6-18 - Boxplots of the EDA signal significant features for positive affect.

6.3. Questionnaires' description

1 - Not at all 2 – Somewhat 3 - Moderately so 4 - Very much so
I feel at ease
I feel nervous
I am jittery
I am relaxed
I am worried
I feel pleasant

Table 6-1 STAI questionnaire items.

Negative Affect Scale	Positive Affect Scale
1 - Not at all 2 - A little bit 3 – Somewhat 4 - Very much 5 - Extremely	
Irritable	Active
Distressed	Enthusiastic
Ashamed	Interested
Alert	Inspired
Annoyed	Strong
Nervous	Determined
Guilty	Attentive
Scared	Jittery
Hostile	Excited
Afraid	Proud

Table 6-2 PANAS questionnaire items.

6.4. Models Parameters and Parameters Optimization

	Functions	Hyperparameters' Value
LR	"LogisticRegression"	<ul style="list-style-type: none"> • C: {1, 3, 10, 30, 100, 300, 1000} • max_iter: {100, 200, 300, 400} • penalty: {"l2", "none"} • solver: "lbfgs"
LDA	"LinearDiscriminantAnalysis"	<ul style="list-style-type: none"> • solver: {"lsqr", "eigen", "svd"} • shrinkage: None
DT	"DecisionTreeClassifier"	<ul style="list-style-type: none"> • max_features: {"sqrt", "log2"} • criterion: {"gini", "entropy"} • splitter: {"best", "random"} • max_depth: None • max_features: None • max_leaf_nodes: None • min_impurity_decrease: 0.0 • min_samples_leaf: 1 • min_samples_split: 2
SVM	"SVC"	<ul style="list-style-type: none"> • C: {0.01, 0.1, 1, 10, 100, 1000} • gamma: {"scale", "auto"} • kernel: "linear" • decision_function_shape: "ovr" • degree: 3
ADB	"AdaBoostClassifier"	<ul style="list-style-type: none"> • learning_rate: {0.5, 1.0, 1.5} • n_estimator: {10, 20, 30} • algorithm: "SAMME.R" • base_estimator: DecisionTreeClassifier • max_depth: 1
RF	"RandomForestClassifier"	<ul style="list-style-type: none"> • max_features: {"sqrt", "log2"} • n_estimators: {10, 20, 30} • bootstrap: True • ccp_alpha: 0.0 • criterion: "gini" • max_depth: 1 • max_leaf_nodes: None

		<ul style="list-style-type: none"> • max_samples: None • min_impurity_decrease: 0.0 • min_samples_leaf: 1 • min_samples_split: 2
XGB	"XGBClassifier"	<ul style="list-style-type: none"> • learning rate: 1 • subsample: {0.5, 0.8, 1} • n_estimators: {10, 20, 30} • base_score: 0.5 • booster: "gbtree" • gamma: 0 • importance_type: "gain" • max_depth: 6 • num_parallel_tree: 1 • tree_method: "exact"

Table 6-3 Models' hyperparameters and optimization values.

6.5. Data Balancing Techniques Parameters

Techniques	Functions	Parameters
Over	"RandomOverSampler"	sampling_strategy = 1
SMO	"SMOTE"	sampling_strategy = 1, K = 5
ADA	"ADASYN"	sampling_strategy = 1, K = 17
Border	"BorderlineSMOTE"	sampling_strategy = 1, k = 5, m = 10, kind = 'Borderline-2'

Table 6-4 Data balance techniques functions and the selected parameters.

6.6. Combination of Features and Data Balancing Technique

SelectKBest			F1	Roc-Auc
LR	SMO	ADA	AUCC, Range ECG, Max ECG, Var ECG, MeanNN, pNN50, LnHF, VHF, LFn, SampEn, Min HR, Max HR, SDTT, Min EDA, SD EDA, Med SCR, Max SCR, SD SCR, Nr SCR Peaks/s, Mean SCR Peaks Amp, Med SCR Peaks Amp, Med SCR Onsets Amp, Med SCR Recovery Amp, Mean SCR Recovery Amp, Med EMG,	

			Range EMG, Nr Pulse Onsets/s, Med Pulse Onsets Amp, Min ECG, Var HR
LDA	ADA	Border	Range ECG, Max ECG, Var ECG, MeanNN, pNN50, LnHF, VHF, LFn, SampEn, Min HR, Max HR, Min EDA, SD EDA, SD SCR, Nr SCR Peaks/s, Mean SCR Peaks Amp, Med SCR Peaks Amp, Med SCR Recovery Amp, Mean Recovery Amp, Med EMG, Range EMG, Nr Pulse Onsets/s, Min ECG, Var HR
DT	Border	Border	Range ECG, Max ECG, Var ECG, MeanNN, pNN50, Min HR Max HR, Min EDA, SD EDA, SD SCR, Nr SCR Peaks/s, Mean SCR Peaks Amp, Med SCR Peaks Amp, Med SCR Onsets Amp, Med SCR Recovery Amp, Mean SCR Recovery Amp, Med EMG, Nr Pulse Onsets/s, Min ECG, Var HR
SVM	ADA	ADA	Range ECG, Max ECG, Var ECG, MeanNN, pNN50, LnHF, VHF, LFn, SampEn, Min HR Max HR, Min EDA, SD EDA, SD SCR, Number SCR Peaks/s, Mean SCR Peaks Amp, Med SCR Peaks Amp, Med SCR Onsets Amp, Med SCR Recovery Amp, Mean SCR Recovery Amp, Med EMG, Range EMG, Nr Pulse Onsets/s, Min ECG, Var HR
ADB	ADA	ADA	Range ECG, Var ECG, MeanNN, Min HR, Max HR, Min EDA, SD EDA, Nr SCR Peaks/s, Mean SCR Peaks Amp, Med SCR Onsets Amp, Med SCR Recovery Amp, Mean SCR Recovery Amp, Med EMG, Nr Onsets Peaks/s, Min ECG
RF	SMO	SMO	Range ECG, MeanNN, Min HR, Max HR, Min EDA, Mean SCR Peaks Amp, Med SCR Recovery Amp, Mean SCR Recovery Amp, Med EMG, Min ECG
XGB	ADA	ADA	Range ECG, Max ECG, Var ECG, MeanNN, pNN50, LnHF, VHF, LFn, SampEn, Min HR, Max HR, Min EDA, SD EDA, SD SCR, Nr SCR Peaks/s, Mean SCR Peaks Amp, Med SCR Peaks Amp, Med SCR Onsets Amp, Med SCR Recovery Amp, Mean SCR Recovery Amp, Med EMG, Range EMG, Nr Pulse Onsets/s, Min ECG, Var HR

Table 6-5 Best combinations for anxiety classification using SelectKBest.

RFECV			F1	Roc-Auc
LR	Border	Border	AUCC, Range ECG, Max ECG, Med ECG, ApEn, Min HR, Max HR, Range HR, Min EDA, SD EDA, Mean SCR Peaks Amp, Med EMG, Nr Pulse Onsets/s, Min ECG, Var HR	
LDA	Over	Border	Max ECG, MeanNN, TINN, pNN50, SampEn, Min HR, Range HR, Min EDA, Med SCR, Med SCR Peaks Amp, Med SCR Recovery Amp, Med Recovery time, Max SCL, Med SCL, Med EMG, Min Env, Activ Peaks Mean Amp, Activ Peaks Med Amp, Min ECG, Mean ECG	
DT	Border	ADA	Range ECG, Max ECG, Var ECG, Med ECG, MeanNN, MadNN, TINN, pNN50, CVNN, LnHF, LF, VHF, LFn, SampEn, ApEn, Min HR, Max HR, Range HR, SDTT, Corr SCL, Med EMG, Min EMG, Range EMG, Perc 10 EMG, Min Env, Perc 10 Env, Nr Pulse Onsets/s, Nr Activ Peaks/s, Activ Peaks Mean Amp, Activ Peaks Mean Amp, Med Pulse Onsets Amp, Min ECG, Mean ECG, Var HR, Range EDA, Min SCR, Mean SCR, Range SCL	
SVM	Border	Border	Range ECG, Med ECG, pNN50, ApEn, Min HR Max HR, Range HR, SD EDA, Corr SCR, Med EMG, Min EMG, Perc 10 EMG, Min Env, Nr Pulse Onsets/s, Min ECG, Var HR	
ADB	SMO	SMO	AUCC, Range ECG, Max ECG, Var ECG, Median ECG, MeanNN, MadNN, TINN, pNN50, LnHF, VHF, LFn, Min HR, Max HR, Range HR, SDTT, Min EDA, Med SCR, Max SCR, Corr SCR, Nr SCR Peaks/s, Med SCR Peaks Amp, Med SCR Onsets Amp, Med SCR Recovery Amp, Mean SCR Recovery Amp, Med Recovery Time, Mean Recovery Time, Max SCL, Med SCL, Med EMG, Perc 10 EMG, Perc 10 Env, Nr Pulse Onsets/s, Nr Activ Peaks/s, Activ Peaks Mean Amp, Activ Peaks Med Amp, Pulse Onsets Med Amp, Min ECG, Var HR	
RF	SMO	Border	AUCC, Range ECG, Max ECG, Var ECG, Med ECG, MeanNN, MadNN, TINN, pNN50, CVNN, LnHF, VHF, LFn, SampEn, ApEn, Min HR, Max HR, Range HR, SDTT, Min EDA, SD EDA, Max SCR, SD SCR, Nr SCR Peaks/s, Mean SCR Peaks Amp, Med SCR Peaks Amp, Med SCR Onsets Amp, Med SCR Recovery Amp, Mean SCR Recovery Amp, Mean Recovery time, Med	

			EMG, Perc 10 EMG, Perc 10 Env, Activ Peaks Mean Amp, Activ Peaks Med Amp, Med Pulse Onsets Amp, Min ECG, Var HR
XGB	ADA	ADA	AUCC, Range ECG, Max ECG, Var ECG, pNN50, Max HR, Range HR, Med EMG, Perc 10 EMG, Nr Activ Peaks/s

Table 6-6 Best combinations for anxiety classification using RFECV.

SelectKBest			F1	Roc-Auc
LR	Border	Border	Max ECG, MeanNN, LnHF, LFn, SampEn, ApEn, Min HR, Max HR, SD SCR, Nr SCR Peaks/s, Med EMG, Min EMG, Nr Activ peaks/s, Activ Peaks Mean Amp, Med Pulse Onsets Amp	
LDA	Border	Border	Max ECG, MeanNN, LnHF, LFn, SampEn, ApEn, Min HR, Max HR, SD SCR, Nr SCR Peaks/s, Med EMG, Min EMG, Nr Activ peaks/s, Activ Peaks Mean Amp, Med Pulse Onsets Amp	
DT	Over	Over	Range ECG, Max ECG, Var ECG, MeanNN, TINN, pNN50, LnHF, LFn, SampEn, ApEn, Min HR, Max HR, Min EDA, SD EDA, Med SCR, Max SCR, Corr SCR, SD SCR, Nr SCR Peaks/s, Mean SCR Peaks Amp, Med SCR Onsets Amp, Mean SCR Recovery Amp, Med EMG, Min EMG, Range EMG, Nr Activ peaks/s, Activ Peaks Mean Amp, Med Pulse Onset Amp, Min SCR	
SVM	Over	Border	Max ECG, MeanNN, LnHF, LFn, SampEn, Min HR, SD SCR, Nr SCR Peaks/s, Med EMG, Nr Activ Peaks/s	
ADB	Border	Border	Max ECG, MeanNN, LnHF, LFn, SampEn, Min HR, SD SCR, Nr SCR Peaks/s, Med EMG, Nr Activ Peaks/s	
RF	Border	SMOTE	Max ECG, MeanNN, pNN50, LnHF, LFn, SampEn, ApEn, Min HR, Max HR, Min EDA, Med SCR, Max SCR, SD SCR, Nr SCR Peaks/s, Med EMG, Min EMG, Range EMG, Nr Activ Peaks/s, Mean Activ Peaks Amp, Med Pulse Onsets Amp	
XGB	ADA	Border	Max ECG, MeanNN, LnHF, LFn, SampEn, Min HR, SD SCR, Nr SCR Peaks/s, Med EMG, Nr Activ Peaks/s	

Table 6-7 Best combinations for negative affect classification using SelectKBest.

RFECV			F1	Roc-Auc
LR	ADA	ADA	Range ECG, Max ECG, Med ECG, MeanNN, MadNN, TINN, pNN50, CVNN, LnHF, VHF, LFn, SampEn, ApEn, Min HR, Max HR, SDTT, Min EDA, SD EDA, Max SCR, Range SCR, Corr SCR, SD SCR, Nr SCR Peaks/s, Mean SCR Peaks Amp, Med SCR Peaks Amp, Med SCR Onsets Amp, Mean SCR Recovery Am, Med Recovery Time, Corr SCL, Med EMG, Min EMG, Range EMG, Perc 10 EMG, Min Env, Nr Pulse Onsets/s, Nr Activ Peaks/s, Mean Activ Peaks Amp, Med Pulse Onsets Amp, Min ECG, Var HR, Min SCR	
LDA	SMO	SMO/ ADA	Range ECG, Max ECG, MeanNN, TINN, CVNN, VHF, Min HR, Max HR, Range HR, SDTT, Min EDA, SD EDA, Median SCR Recovery Amp, Mean Recovery Time, Mean Rise Time, Max SCL, Median SCL, Median EMG, Number Activation Peaks/s, Mean Activation Peaks Amplitude, Min ECG, Mean ECG	
DT	SMO	SMO	Range ECG, Max ECG, Var ECG, Med ECG, MeanNN, MadNN, TINN, pNN50, CVNN, LnHF, LF, Nr Activ Peaks/s, Activ Peaks Mean Amp, Activ Peaks Med Amp, Min ECG, Mean ECG, Var HR, Range EDA, Min SCR, Mean SCR, Range SCL	
SVM	Over/ SMOTE	Border	AUCC, Max ECG, Med ECG, TINN, CVNN, VHF, LFn, Max HR, Min EDA, SD EDA, Max SCR, Corr SCR, Med SCR Onsets Amp, Med EMG, Range EMG, Perc 10 EMG, Nr Pulse Onsets/s, Nr Activ Peaks/s, Activ Peaks Med Amp, Med Pulse Onsets Amp, Min ECG, Var HR	
ADB	Over	Border	AUCC, Range ECG, Max ECG, Med ECG, VHF, SDTT, SD EDA, SD SCR, Med SCR Onsets Amp, Mean SCR Recovery Amp, Med EMG, Nr Pulse Onsets/s, Nr Activ Peaks/s, Activ Peaks Med Amp	
RF	Over	Over	AUCC, Range ECG, Max ECG, Var ECG, Median ECG, MeanNN, pNN50, LnHF, VHF, LFn, SampEn, Max HR, SDTT, SD EDA, Med SCR, Max SCR, Corr SCR, SD SCR, Nr SCR Peaks/s, Mean SCR Peaks Amp, Med SCR Onsets Amp, Mean SCR Recovery Amp, Med EMG, Range EMG, Perc 10 Env, Nr	

			Activ Peaks/s, Mean Activ Peaks Amp, Med Activ Peaks Amp, Med Pulse Onsets Amp, Min ECG, Var HR, Min SCR
XGB	Border	Border	Range ECG, Max ECG, Var ECG, MeanNN, VHF, Med SCR Onsets Amp, Med SCR Recovery Amp, Nr Activ Peaks/s, Activ Peaks Mean Amp, Activ Peaks Med Amp

Table 6-8 Best combinations for negative affect classification using RFECV.

6.7. Anxiety Classification Results

	Accuracy	Roc-Auc	F1	Precision	Recall	
	0.631 (0.054)	0.620 (0.087)	0.444 (0.072)	0.406 (0.087)	0.567 (0.109)	-
Over	0.604 (0.059)	0.612 (0.084)	0.507 (0.063)	0.394 (0.057)	0.750 (0.100)	
SMO	0.529 (0.098)	0.513 (0.133)	0.420 (0.112)	0.325 (0.091)	0.617 (0.164)	
ADA	0.542 (0.078)	0.587 (0.088)	0.511 (0.047)	0.380 (0.052)	0.833 (0.047)	
Border	0.578 (0.084)	0.729 (0.061)	0.537 (0.036)	0.443 (0.076)	0.850 (0.080)	
	0.809 (0.060)	0.833 (0.060)	0.612 (0.150)	0.563 (0.152)	0.717 (0.166)	S
Over	0.827 (0.054)	0.902 (0.036)	0.748 (0.062)	0.720 (0.107)	0.867 (0.069)	
SMO	0.831 (0.051)	0.901 (0.036)	0.751 (0.061)	0.724 (0.105)	0.867 (0.069)	
ADA	0.782 (0.074)	0.908 (0.027)	0.725 (0.072)	0.687 (0.122)	0.900 (0.072)	
Border	0.716 (0.113)	0.793 (0.118)	0.703 (0.095)	0.613 (0.122)	0.933 (0.028)	
	0.707 (0.061)	0.784 (0.067)	0.490 (0.067)	0.522 (0.186)	0.467 (0.159)	R
Over	0.671 (0.069)	0.754 (0.070)	0.521 (0.067)	0.472 (0.075)	0.650 (0.101)	
SMO	0.671 (0.069)	0.752 (0.070)	0.521 (0.067)	0.472 (0.075)	0.650 (0.101)	
ADA	0.676 (0.067)	0.762 (0.070)	0.576 (0.038)	0.492 (0.056)	0.783 (0.087)	
Border	0.711 (0.076)	0.852 (0.080)	0.656 (0.056)	0.527 (0.065)	0.933 (0.060)	

Table 6-9 LR anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).

	Accuracy	Roc-Auc	F1	Precision	Recall	
	0.640 (0.091)	0.648 (0.124)	0.562 (0.064)	0.460 (0.069)	0.783 (0.080)	-
Over	0.636 (0.092)	0.657 (0.115)	0.570 (0.071)	0.453 (0.070)	0.817 (0.083)	
SMO	0.644 (0.094)	0.653 (0.115)	0.573 (0.072)	0.465 (0.073)	0.800 (0.087)	
ADA	0.649 (0.086)	0.662 (0.115)	0.594 (0.062)	0.467 (0.066)	0.867 (0.056)	
Border	0.636 (0.081)	0.678 (0.096)	0.576 (0.069)	0.447 (0.068)	0.850 (0.068)	
	0.760 (0.085)	0.663 (0.096)	0.605 (0.115)	0.474 (0.109)	0.600 (0.114)	S
Over	0.609 (0.102)	0.676 (0.123)	0.556 (0.093)	0.449 (0.101)	0.783 (0.061)	
SMO	0.622 (0.106)	0.658 (0.124)	0.570 (0.099)	0.471 (0.112)	0.783 (0.061)	
ADA	0.640 (0.107)	0.692 (0.125)	0.615 (0.099)	0.504 (0.120)	0.883 (0.056)	
Border	0.671 (0.101)	0.694 (0.142)	0.620 (0.101)	0.517 (0.117)	0.833 (0.094)	
	0.658 (0.107)	0.649 (0.136)	0.573 (0.092)	0.521 (0.114)	0.717 (0.077)	R
Over	0.671 (0.114)	0.671 (0.134)	0.615 (0.109)	0.531 (0.125)	0.800 (0.087)	
SMO	0.631 (0.095)	0.669 (0.124)	0.561 (0.077)	0.443 (0.071)	0.800 (0.084)	

ADA	0.636 (0.100)	0.674 (0.122)	0.575 (0.087)	0.450 (0.079)	0.833 (0.094)
Border	0.627 (0.102)	0.731 (0.104)	0.595 (0.088)	0.466 (0.095)	0.883 (0.065)

Table 6-10 LDA anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).

	Accuracy	Roc-Auc	F1	Precision	Recall	
	0.729 (0.070)	0.681 (0.079)	0.526 (0.118)	0.513 (0.109)	0.583 (0.133)	
Over	0.680 (0.052)	0.644 (0.069)	0.464 (0.115)	0.424 (0.127)	0.567 (0.138)	
SMO	0.707 (0.088)	0.768 (0.062)	0.663 (0.083)	0.560 (0.107)	0.560 (0.107)	-
ADA	0.702 (0.085)	0.691 (0.078)	0.579 (0.105)	0.548 (0.129)	0.667 (0.085)	
Border	0.764 (0.065)	0.781 (0.059)	0.666 (0.088)	0.609 (0.110)	0.817 (0.106)	
	0.862(0.048)	0.837 (0.059)	0.754 (0.083)	0.736 (0.083)	0.783 (0.087)	
Over	0.813(0.046)	0.730 (0.074)	0.557(0.144)	0.576 (0.149)	0.550 (0.145)	
SMO	0.791 (0.051)	0.730 (0.071)	0.569 (0.122)	0.620 (0.089)	0.600 (0.140)	S
ADA	0.764 (0.044)	0.802 (0.034)	0.674 (0.049)	0.585 (0.069)	0.883 (0.087)	
Border	0.884 (0.041)	0.884 (0.042)	0.811 (0.057)	0.781 (0.082)	0.883 (0.065)	
	0.711 (0.061)	0.723 (0.057)	0.583 (0.088)	0.533 (0.106)	0.750 (0.118)	
Over	0.711 (0.028)	0.692 (0.044)	0.520 (0.060)	0.478 (0.034)	0.650 (0.128)	
SMO	0.756 (0.036)	0.733 (0.063)	0.570 (0.094)	0.512 (0.072)	0.683 (0.138)	R
ADA	0.751 (0.030)	0.793 (0.020)	0.657 (0.021)	0.554 (0.054)	0.883 (0.069)	
Border	0.733 (0.075)	0.781 (0.051)	0.664 (0.064)	0.567 (0.080)	0.883 (0.073)	

Table 6-11 DT anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).

	Accuracy	Roc-Auc	F1	Precision	Recall	
	0.649 (0.070)	0.496 (0.127)	0.399 (0.106)	0.399 (0.150)	0.500 (0.131)	
Over	0.596 (0.062)	0.593 (0.095)	0.467 (0.072)	0.402 (0.089)	0.667 (0.131)	
SMO	0.529 (0.073)	0.568 (0.106)	0.360 (0.081)	0.307 (0.082)	0.483 (0.121)	-
ADA	0.524 (0.083)	0.620 (0.085)	0.458 (0.060)	0.365 (0.067)	0.733 (0.121)	
Border	0.573 (0.092)	0.655 (0.048)	0.538 (0.049)	0.482 (0.121)	0.833 (0.085)	
	0.760 (0.064)	0.769 (0.077)	0.595 (0.099)	0.649 (0.123)	0.683 (0.136)	
Over	0.778 (0.072)	0.797 (0.074)	0.621 (0.129)	0.618 (0.145)	0.717 (0.157)	
SMO	0.782 (0.068)	0.777 (0.091)	0.632 (0.110)	0.649 (0.092)	0.750 (0.149)	S
ADA	0.764 (0.087)	0.851 (0.059)	0.721 (0.079)	0.679 (0.127)	0.917 (0.075)	
Border	0.707 (0.096)	0.821 (0.100)	0.679 (0.080)	0.575 (0.112)	0.933 (0.015)	
	0.693 (0.091)	0.690 (0.148)	0.547 (0.139)	0.485 (0.133)	0.683 (0.174)	
Over	0.702 (0.096)	0.702 (0.063)	0.473 (0.145)	0.537 (0.160)	0.550 (0.179)	
SMO	0.667 (0.081)	0.691 (0.060)	0.468 (0.144)	0.523 (0.158)	0.550 (0.179)	R
ADA	0.622 (0.062)	0.706 (0.062)	0.559 (0.038)	0.462 (0.103)	0.817 (0.089)	
Border	0.627 (0.111)	0.832 (0.081)	0.605 (0.091)	0.520 (0.118)	0.883 (0.087)	

Table 6-12 SVM anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).

	Accuracy	Roc-Auc	F1	Precision	Recall	
	0.867 (0.043)	0.908 (0.025)	0.745 (0.066)	0.759 (0.094)	0.817 (0.043)	
Over	0.893 (0.037)	0.910 (0.039)	0.720 (0.117)	0.903 (0.059)	0.650 (0.136)	
SMO	0.827 (0.051)	0.919 (0.031)	0.735 (0.072)	0.706 (0.109)	0.833 (0.078)	-
ADA	0.898 (0.029)	0.916 (0.027)	0.803 (0.057)	0.884 (0.075)	0.800 (0.090)	
Border	0.831 (0.058)	0.960 (0.013)	0.777 (0.060)	0.694 (0.093)	0.933 (0.028)	
	0.871 (0.044)	0.939 (0.026)	0.756 (0.087)	0.762 (0.087)	0.767 (0.098)	
Over	0.836 (0.060)	0.876 (0.058)	0.750 (0.086)	0.716 (0.107)	0.833 (0.078)	
SMO	0.827 (0.070)	0.838 (0.074)	0.763 (0.087)	0.725 (0.113)	0.900 (0.089)	S
ADA	0.893 (0.043)	0.915 (0.034)	0.805 (0.080)	0.802 (0.085)	0.817 (0.080)	
Border	0.831(0.048)	0.913 (0.037)	0.761 (0.047)	0.697 (0.089)	0.900 (0.037)	
	0.867 (0.052)	0.932 (0.026)	0.805 (0.047)	0.801 (0.110)	0.833 (0.033)	
Over	0.902 (0.036)	0.920 (0.030)	0.817 (0.067)	0.878 (0.079)	0.833 (0.097)	
SMO	0.920 (0.038)	0.961 (0.026)	0.864 (0.057)	0.875 (0.082)	0.867 (0.038)	R
ADA	0.884 (0.043)	0.935 (0.031)	0.818 (0.059)	0.820 (0.099)	0.883 (0.069)	
Border	0.818 (0.064)	0.937 (0.027)	0.745 (0.069)	0.712 (0.1109)	0.867 (0.073)	

Table 6-13 ADB anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).

	Accuracy	Roc-Auc	F1	Precision	Recall	
	0.867 (0.030)	0.922 (0.042)	0.695 (0.097)	0.853 (0.081)	0.650 (0.121)	
Over	0.876 (0.045)	0.943 (0.032)	0.761 (0.077)	0.896 (0.093)	0.750 (0.111)	
SMO	0.858 (0.062)	0.966 (0.014)	0.777 (0.070)	0.836 (0.105)	0.800 (0.084)	-
ADA	0.889 (0.056)	0.930 (0.040)	0.766 (0.123)	0.774 (0.107)	0.767 (0.136)	
Border	0.876 (0.026)	0.977 (0.014)	0.787 (0.022)	0.835 (0.091)	0.817(0.068)	
	0.853 (0.030)	0.880 (0.043)	0.628 (0.101)	0.887 (0.066)	0.567 (0.132)	
Over	0.884 (0.032)	0.870 (0.085)	0.690 (0.104)	0.975 (0.022)	0.583 (0.120)	
SMO	0.929 (0.021)	0.969 (0.018)	0.846 (0.047)	0.953 (0.026)	0.767 (0.064)	S
ADA	0.902 (0.031)	0.931 (0.035)	0.787 (0.071)	0.900 (0.060)	0.717 (0.084)	
Border	0.867 (0.034)	0.962 (0.017)	0.777 (0.051)	0.762 (0.079)	0.833 (0.067)	
	0.844 (0.039)	0.971 (0.012)	0.655 (0.101)	0.844 (0.0809)	0.650 (0.142)	
Over	0.898 (0.027)	0.965 (0.019)	0.784 (0.062)	0.901 (0.064)	0.733 (0.123)	
SMO	0.893 (0.061)	0.973 (0.017)	0.845 (0.078)	0.826 (0.096)	0.917 (0.075)	R
ADA	0.840 (0.058)	0.956 (0.024)	0.767 (0.070)	0.733 (0.104)	0.867 (0.069)	
Border	0.858 (0.073)	0.980 (0.013)	0.806 (0.081)	0.760 (0.097)	0.917 (0.075)	

Table 6-14 RF anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).

	Accuracy	Roc-Auc	F1	Precision	Recall	
	0.893 (0.028)	0.947 (0.013)	0.782 (0.062)	0.870 (0.072)	0.767 (0.098)	
Over	0.898 (0.034)	0.972 (0.011)	0.833 (0.050)	0.803 (0.079)	0.900 (0.055)	-
SMO	0.898 (0.038)	0.965 (0.009)	0.831 (0.057)	0.814 (0.081)	0.883 (0.065)	

ADA	0.920 (0.032)	0.939 (0.037)	0.870 (0.051)	0.830 (0.072)	0.933 (0.043)	
Border	0.840 (0.068)	0.951 (0.023)	0.783 (0.065)	0.775 (0.111)	0.883 (0.065)	
	0.813 (0.040)	0.867 (0.074)	0.628 (0.079)	0.767 (0.104)	0.650 (0.134)	S
Over	0.773 (0.072)	0.891 (0.029)	0.656 (0.103)	0.604 (0.108)	0.767 (0.119)	
SMO	0.822 (0.062)	0.909 (0.033)	0.740 (0.066)	0.717 (0.102)	0.833 (0.071)	
ADA	0.831 (0.050)	0.934 (0.020)	0.682 (0.105)	0.764 (0.093)	0.767 (0.140)	
Border	0.764 (0.079)	0.908 (0.041)	0.677 (0.094)	0.649 (0.130)	0.783 (0.080)	
	0.809 (0.044)	0.927 (0.028)	0.705 (0.070)	0.697 (0.101)	0.717 (0.115)	R
Over	0.822 (0.046)	0.947 (0.022)	0.735 (0.039)	0.739 (0.106)	0.850 (0.083)	
SMO	0.893 (0.047)	0.945 (0.026)	0.840 (0.059)	0.806 (0.089)	0.917 (0.047)	
ADA	0.916 (0.031)	0.955 (0.021)	0.860 (0.048)	0.830 (0.072)	0.917 (0.047)	
Border	0.907 (0.057)	0.951 (0.036)	0.840 (0.082)	0.805 (0.104)	0.933 (0.060)	

Table 6-15 XGB anxiety classification scores without (-) and with SelectKBest (S) and RFECV (R).

6.8. Negative Affect Classification Results

	Accuracy	Roc-Auc	F1	Precision	Recall	
	0.880 (0.058)	0.913 (0.046)	0.779 (0.101)	0.758 (0.133)	0.844 (0.060)	-
Over	0.876 (0.060)	0.911 (0.051)	0.811 (0.085)	0.724 (0.116)	1.000 (0.000)	
SMO	0.876 (0.064)	0.897 (0.056)	0.818 (0.090)	0.740 (0.125)	1.000 (0.000)	
ADA	0.871 (0.063)	0.909 (0.050)	0.808 (0.087)	0.720 (0.118)	1.000 (0.000)	
Border	0.880 (0.066)	0.936 (0.047)	0.819 (0.094)	0.761 (0.131)	1.000 (0.000)	
	0.889 (0.051)	0.923 (0.039)	0.797 (0.084)	0.778 (0.123)	0.889 (0.063)	S
Over	0.871 (0.067)	0.887 (0.063)	0.815 (0.092)	0.737 (0.127)	1.000 (0.000)	
SMO	0.867 (0.066)	0.891 (0.063)	0.800 (0.089)	0.757 (0.133)	0.956 (0.040)	
ADA	0.862 (0.068)	0.900 (0.059)	0.797 (0.091)	0.753 (0.135)	0.956 (0.040)	
Border	0.858 (0.070)	0.939 (0.055)	0.818 (0.100)	0.750 (0.137)	1.000 (0.000)	
	0.889 (0.067)	0.915 (0.069)	0.798 (0.092)	0.867 (0.119)	0.844 (0.097)	R
Over	0.867 (0.079)	0.917 (0.068)	0.815 (0.090)	0.733 (0.117)	1.000 (0.000)	
SMO	0.862 (0.077)	0.917 (0.067)	0.804 (0.087)	0.730 (0.117)	0.978 (0.020)	
ADA	0.871 (0.080)	0.926 (0.066)	0.826 (0.094)	0.752 (0.124)	1.000 (0.000)	
Border	0.889 (0.063)	0.925 (0.061)	0.819 (0.080)	0.798 (0.118)	0.933 (0.060)	

Table 6-16 LR negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R).

	Accuracy	Roc-Auc	F1	Precision	Recall	
	0.760 (0.070)	0.814 (0.091)	0.670 (0.084)	0.539 (0.112)	1.000 (0.000)	-
Over	0.680 (0.079)	0.790 (0.079)	0.479 (0.148)	0.395 (0.149)	0.689 (0.182)	
SMO	0.720 (0.065)	0.850 (0.081)	0.526 (0.126)	0.415 (0.114)	0.778 (0.175)	
ADA	0.738 (0.060)	0.880 (0.067)	0.680 (0.075)	0.504 (0.111)	1.000 (0.000)	
Border	0.782 (0.055)	0.933 (0.047)	0.680 (0.075)	0.545 (0.105)	1.000 (0.000)	
	0.849 (0.070)	0.920 (0.057)	0.761 (0.079)	0.715 (0.118)	0.911 (0.049)	S
Over	0.849 (0.071)	0.914 (0.070)	0.773 (0.090)	0.699 (0.122)	0.956 (0.040)	
SMO	0.844 (0.074)	0.925 (0.059)	0.771 (0.092)	0.697 (0.124)	0.956 (0.040)	
ADA	0.849 (0.075)	0.930 (0.059)	0.783 (0.092)	0.702 (0.124)	0.978 (0.020)	
Border	0.849 (0.055)	0.931 (0.058)	0.748 (0.075)	0.692 (0.119)	0.911 (0.049)	

	0.800 (0.050)	0.912 (0.056)	0.686 (0.057)	0.582 (0.103)	0.956 (0.040)
Over	0.804 (0.048)	0.917 (0.058)	0.690 (0.052)	0.553 (0.070)	0.978 (0.020)
SMO	0.813 (0.053)	0.920 (0.058)	0.710 (0.066)	0.594 (0.101)	0.978 (0.020)
ADA	0.809 (0.052)	0.920 (0.058)	0.704 (0.064)	0.586 (0.101)	0.978 (0.020)
Border	0.764 (0.049)	0.916 (0.057)	0.690 (0.028)	0.553 (0.037)	0.978 (0.020)

Table 6-17 LDA negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R).

	Accuracy	Roc-Auc	F1	Precision	Recall
	0.866 (0.062)	0.807 (0.053)	0.667 (0.085)	0.796 (0.126)	0.634 (0.058)
Over	0.884 (0.057)	0.903 (0.038)	0.808 (0.079)	0.781 (0.122)	0.933 (0.060)
SMO	0.884 (0.057)	0.903 (0.038)	0.808 (0.079)	0.781 (0.122)	0.933 (0.060)
ADA	0.822 (0.066)	0.831 (0.046)	0.692 (0.078)	0.643 (0.103)	0.844 (0.087)
Border	0.858 (0.051)	0.886 (0.024)	0.763 (0.065)	0.706 (0.116)	0.933 (0.040)
	0.871 (0.050)	0.809 (0.051)	0.692 (0.083)	0.875 (0.112)	0.689 (0.123)
Over	0.933 (0.029)	0.892 (0.057)	0.817 (0.090)	0.870 (0.074)	0.822 (0.116)
SMO	0.876 (0.052)	0.822 (0.052)	0.719 (0.089)	0.825 (0.110)	0.733 (0.116)
ADA	0.876 (0.057)	0.856 (0.058)	0.753 (0.099)	0.746 (0.117)	0.822 (0.102)
Border	0.867 (0.041)	0.858 (0.059)	0.723 (0.090)	0.645 (0.097)	0.844 (0.097)
	0.782 (0.073)	0.731 (0.096)	0.564 (0.125)	0.538 (0.129)	0.644 (0.152)
Over	0.849 (0.036)	0.831 (0.053)	0.671 (0.070)	0.676 (0.096)	0.800 (0.119)
SMO	0.889 (0.033)	0.849 (0.043)	0.721 (0.064)	0.811 (0.091)	0.756 (0.111)
ADA	0.827 (0.045)	0.808 (0.045)	0.658 (0.073)	0.628 (0.107)	0.778 (0.094)
Border	0.809 (0.037)	0.847 (0.039)	0.664 (0.058)	0.541 (0.065)	0.911 (0.080)

Table 6-18 DT negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R).

	Accuracy	Roc-Auc	F1	Precision	Recall
	0.831 (0.062)	0.858 (0.069)	0.577 (0.154)	0.633 (0.179)	0.556 (0.144)
Over	0.822 (0.066)	0.927 (0.044)	0.745 (0.093)	0.643 (0.131)	1.000 (0.000)
SMO	0.831 (0.065)	0.934 (0.037)	0.755 (0.091)	0.653 (0.128)	1.000 (0.000)
ADA	0.831 (0.065)	0.935 (0.036)	0.755 (0.091)	0.653 (0.128)	1.000 (0.000)
Border	0.871 (0.071)	0.922 (0.045)	0.822 (0.098)	0.753 (0.135)	1.000 (0.000)
	0.831 (0.080)	0.882 (0.058)	0.763 (0.097)	0.696 (0.132)	0.956 (0.040)
Over	0.884 (0.065)	0.883 (0.067)	0.823 (0.092)	0.755 (0.129)	1.000 (0.000)
SMO	0.871 (0.071)	0.915 (0.047)	0.822 (0.098)	0.753 (0.135)	1.000 (0.000)
ADA	0.867 (0.073)	0.928 (0.039)	0.818 (0.100)	0.750 (0.137)	0.750 (0.137)
Border	0.871 (0.054)	0.985 (0.009)	0.763 (0.087)	0.772 (0.125)	0.889 (0.099)
	0.840 (0.068)	0.885 (0.069)	0.706 (0.109)	0.745 (0.140)	0.822 (0.116)
Over	0.876 (0.069)	0.885 (0.068)	0.826 (0.096)	0.758 (0.133)	1.000 (0.000)
SMO	0.876 (0.069)	0.888 (0.066)	0.826 (0.096)	0.758 (0.133)	1.000 (0.000)
ADA	0.871 (0.071)	0.888 (0.068)	0.822 (0.098)	0.754 (0.135)	1.000 (0.000)

Border	0.862 (0.068)	0.894 (0.070)	0.797 (0.091)	0.753 (0.135)	1.000 (0.000)	
---------------	---------------	---------------	---------------	---------------	---------------	--

Table 6-19 SVM negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R).

	Accuracy	Roc-Auc	F1	Precision	Recall	
	0.889 (0.057)	0.876 (0.068)	0.779 (0.079)	0.872 (0.114)	0.800 (0.091)	
Over	0.787 (0.058)	0.869 (0.048)	0.650 (0.061)	0.597 (0.116)	0.844 (0.060)	-
SMO	0.813 (0.045)	0.901 (0.039)	0.657 (0.049)	0.583 (0.075)	0.822 (0.067)	
ADA	0.871 (0.059)	0.888 (0.055)	0.786 (0.083)	0.728 (0.114)	0.933 (0.060)	
Border	0.889 (0.058)	0.958 (0.024)	0.823 (0.080)	0.781 (0.122)	0.956 (0.040)	
	0.822 (0.067)	0.877 (0.047)	0.689 (0.084)	0.742 (0.142)	0.800 (0.096)	
	0.862 (0.053)	0.875 (0.049)	0.719 (0.071)	0.771 (0.109)	0.778 (0.094)	S
SMO	0.876 (0.056)	0.888 (0.047)	0.699 (0.116)	0.779 (0.136)	0.667 (0.113)	
ADA	0.831 (0.075)	0.838 (0.060)	0.618 (0.131)	0.719 (0.158)	0.578 (0.119)	
Border	0.893 (0.057)	0.946 (0.040)	0.799 (0.073)	0.872 (0.114)	0.822 (0.074)	
	0.858 (0.061)	0.909 (0.051)	0.715 (0.101)	0.764 (0.130)	0.800 (0.119)	R
Over	0.880 (0.060)	0.905 (0.053)	0.804 (0.088)	0.778 (0.124)	0.933 (0.060)	
SMO	0.871 (0.059)	0.910 (0.043)	0.786 (0.083)	0.728 (0.114)	0.933 (0.060)	
ADA	0.836 (0.055)	0.915 (0.046)	0.744 (0.098)	0.778 (0.124)	0.867 (0.119)	
Border	0.867 (0.055)	0.920 (0.040)	0.748 (0.076)	0.784 (0.121)	0.844 (0.097)	

Table 6-20 ADB negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R).

	Accuracy	Roc-Auc	F1	Precision	Recall	
	0.907 (0.042)	0.967 (0.019)	0.774 (0.074)	0.886 (0.102)	0.778 (0.104)	
Over	0.924 (0.030)	0.942 (0.036)	0.828 (0.049)	0.900 (0.089)	0.822 (0.067)	-
SMO	0.898 (0.044)	0.942 (0.042)	0.763 (0.082)	0.827 (0.112)	0.733 (0.067)	
ADA	0.880 (0.046)	0.927 (0.046)	0.707 (0.109)	0.731 (0.090)	0.756 (0.123)	
Border	0.911 (0.033)	0.976 (0.011)	0.815 (0.053)	0.806 (0.087)	0.867 (0.049)	
	0.916 (0.028)	0.961 (0.025)	0.768 (0.077)	0.907 (0.083)	0.733 (0.102)	
	0.902 (0.028)	0.944 (0.025)	0.741 (0.069)	0.875 (0.087)	0.711 (0.092)	S
SMO	0.898 (0.039)	0.986 (0.008)	0.754 (0.085)	0.840 (0.097)	0.778 (0.113)	
ADA	0.867 (0.050)	0.962 (0.014)	0.684 (0.086)	0.825 (0.110)	0.689 (0.123)	
Border	0.933 (0.046)	0.955 (0.045)	0.880 (0.072)	0.886 (0.102)	0.933 (0.060)	
	0.893 (0.053)	0.993 (0.004)	0.816 (0.068)	0.804 (0.114)	0.911 (0.049)	R
Over	0.920 (0.057)	1.000 (0.000)	0.872 (0.076)	0.875 (0.112)	0.933 (0.040)	
SMO	0.911 (0.055)	0.997 (0.002)	0.847 (0.071)	0.875 (0.112)	0.889 (0.044)	
ADA	0.920 (0.036)	0.994 (0.006)	0.833 (0.061)	0.845 (0.094)	0.867 (0.058)	
Border	0.911 (0.055)	0.999 (0.001)	0.849 (0.071)	0.855 (0.109)	0.911 (0.049)	

Table 6-21 RF negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R).

	Accuracy	Roc-Auc	F1	Precision	Recall	
	0.867 (0.052)	0.928 (0.035)	0.745 (0.076)	0.734 (0.111)	0.822 (0.067)	-
Over	0.880 (0.051)	0.951 (0.024)	0.771 (0.079)	0.759 (0.111)	0.844 (0.067)	
SMO	0.867 (0.056)	0.938 (0.027)	0.755 (0.090)	0.704 (0.113)	0.867 (0.073)	
ADA	0.871 (0.054)	0.907 (0.035)	0.777 (0.076)	0.731 (0.112)	0.911 (0.058)	
Border	0.884 (0.052)	0.981 (0.008)	0.789 (0.069)	0.763 (0.104)	0.889 (0.063)	
	0.871 (0.052)	0.956 (0.018)	0.749 (0.057)	0.784 (0.121)	0.800 (0.073)	S
Over	0.853 (0.063)	0.958 (0.025)	0.736 (0.081)	0.726 (0.116)	0.822 (0.067)	
SMO	0.871 (0.059)	0.901 (0.039)	0.767 (0.089)	0.731 (0.113)	0.867 (0.073)	
ADA	0.929 (0.022)	0.946 (0.031)	0.809 (0.064)	0.879 (0.068)	0.800 (0.096)	
Border	0.889 (0.052)	0.993 (0.007)	0.798 (0.067)	0.889 (0.063)	0.800 (0.119)	
	0.889 (0.052)	0.950 (0.033)	0.797 (0.070)	0.775 (0.102)	0.889 (0.063)	R
Over	0.876 (0.054)	0.974 (0.008)	0.764 (0.076)	0.784 (0.121)	0.822 (0.067)	
SMO	0.871 (0.059)	0.974 (0.011)	0.759 (0.081)	0.734 (0.111)	0.844 (0.067)	
ADA	0.898 (0.059)	0.929 (0.058)	0.822 (0.081)	0.822 (0.112)	0.889 (0.063)	
Border	0.898 (0.053)	0.979 (0.012)	0.823 (0.067)	0.813 (0.112)	0.911 (0.049)	

Table 6-22 XGB negative affect classification scores without (-) and with SelectKBest (S) and RFECV (R).

6.9. Positive Affect Classification Results

	Accuracy	Roc-Auc	F1	Precision	Recall
LR	0.991 (0.008)	1.000 (0.000)	0.994 (0.005)	1.000 (0.000)	0.989 (0.010)
S	0.996 (0.004)	1.000 (0.000)	0.997 (0.003)	1.000 (0.000)	0.994 (0.005)
R	0.996 (0.004)	1.000 (0.000)	0.997 (0.003)	1.000 (0.000)	0.994 (0.005)
LDA	0.920 (0.048)	0.994 (0.005)	0.945 (0.037)	1.000 (0.000)	0.906 (0.060)
S	0.956 (0.025)	0.991 (0.005)	0.971 (0.017)	0.994 (0.006)	0.950 (0.028)
R	0.907 (0.044)	0.993 (0.005)	0.929 (0.036)	1.000 (0.000)	0.878 (0.059)
DT	0.800 (0.055)	0.733 (0.070)	0.867 (0.038)	0.902 (0.034)	0.844 (0.059)
S	0.836 (0.043)	0.847 (0.027)	0.884 (0.034)	0.964 (0.013)	0.828 (0.058)
R	0.858 (0.039)	0.836 (0.027)	0.902 (0.032)	0.952 (0.019)	0.872 (0.061)
SVM	0.982 (0.012)	1.000 (0.000)	0.989 (0.007)	0.979 (0.013)	1.000 (0.000)
S	0.987 (0.012)	1.000 (0.000)	0.992 (0.007)	0.985 (0.014)	1.000 (0.000)
R	0.996 (0.004)	1.000 (0.000)	0.997 (0.002)	0.995 (0.005)	1.000 (0.000)
ADB	0.809 (0.047)	0.840 (0.055)	0.878 (0.032)	0.885 (0.028)	0.878 (0.050)
S	0.831 (0.056)	0.872 (0.057)	0.890 (0.037)	0.908 (0.032)	0.878 (0.050)
R	0.836 (0.056)	0.843 (0.056)	0.893 (0.044)	0.912 (0.029)	0.878 (0.050)
RF	0.960 (0.012)	0.991 (0.005)	0.976 (0.007)	0.959 (0.016)	0.994 (0.005)
S	0.973 (0.012)	0.988 (0.010)	0.975 (0.007)	0.979 (0.013)	1.000 (0.000)
R	0.956 (0.019)	0.996 (0.003)	0.961 (0.013)	0.979 (0.012)	0.972 (0.016)

XGB	0.916 (0.028)	0.979 (0.011)	0.949 (0.017)	0.934 (0.025)	0.967 (0.020)
S	0.924 (0.022)	0.954 (0.021)	0.955 (0.013)	0.931 (0.025)	0.983 (0.010)
R	0.929 (0.023)	0.910 (0.059)	0.951 (0.018)	0.945 (0.028)	0.961 (0.019)

Table 6-23 Positive affect classification scores without and with SelectKBest (S) and RFECV (R).