République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MOHAMED KHIDER DE BISKRA

Faculté des sciences exactes et sciences de la nature et de la vie

Département d'informatique

N° d'Ordre: ....
N° de Série: ....

Thèse

En vue de l'obtention du diplôme de

DOCTORAT 3ème Cycle LMD en INFORMATIQUE
Option: Sciences et Techniques de l'Image

par

Mlle Sihem NITA

THEME

Conception de système de traitement de données sur les émotions d'un être humain dans un environnement mobile et incertain

Soutenue devant le jury

Prof. Laid Kahloul	Université Med Khider, Biskra	Président
Prof. Salim BITAM	Université Med Khider, Biskra	Rapporteur
Prof. Abdelhamid MELLOUK	Université Paris-Est Créteil, France	Co-rapporteur
Prof. Ghalem Belalem	Université Oran 1	Examinateur
Dr. Okba Tibermacine	Université Med Khider, Biskra	Examinateur

Année Universitaire : 2022/2023

People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

MOHAMED KHIDER UNIVERSITY OF BISKRA

Exact Sciences, Natural Sciences and Life Sciences Faculty

Department of Computer Sciences

Order Number: ....
Serial Number: ....

Thesis

Presented in Partial Fulfillment of the Requirement for the Degree of

DOCTORAT Third Cycle LMD in Computer Science
Option: Image Sciences and Techniques

by

Mlle Sihem NITA

THEME

Design of a data processing system on the emotions of a human being in a mobile and uncertain environment

In front of a jury composed of

Prof. Laid Kahloul	Med Khider University, Biskra	Chair
Prof. Salim BITAM	Med Khider University, Biskra	Supervisor
Prof. Abdelhamid MELLOUK	Paris Est Creteil University, France	Co-Supervisor
Prof. Ghalem Belalem	University of Oran 1	Examiner
Dr. Okba Tibermacine	Med Khider University, Biskra	Examiner

Academic year 2022/2023

Acknowledgements

Praise be to God first and foremost

This thesis would not have been accomplished without the support and guidance of several individuals to whom I would like to express my gratitude in this acknowledgment.

First and foremost, my deepest and sincerest gratitude goes to my supervisor Pr.Salim BITAM, who guided me through the entire journey. His patience, valued insight and great support underpinned this PhD. He has been an ideal teacher, mentor, and thesis supervisor, offering advice and encouragement with a perfect blend of insight and humor. I am proud of, and grateful for, my time working with Pr.Salim.

I would also like to thank deeply Pr.Abdelhamid MELLOUK from University of Paris-Est, LISSI Laboratory France. It was a great privilege and honor to work under his guidance. I am extremely grateful for his support, guidance and relevant advices.

The most eminent thanks go to my dearest parents who have always been there for me, for the sacrifices made towards me, for their support and all the efforts made for my education and training. To my brothers Dr.Lokmane and Dr.Mohamed and my sister Dr.Hadjer.

Many thanks to all the members of the jury Pr.Laid KAHLOUL, Pr.Ghalem BELALEM and Dr.Okba TIBERMACINE for accepting to debate this thesis and for the time they spend to review this work, I am grateful for the attention they paid to my work.

Finally, I cannot end without thanking my closest friends and all my family for supporting me and encouraging me at each step of this work.

Sihem NITA

Abstract

A human emotion is considered to be a state of mind of an individual that is complex and intense, onset abruptly and may last for a relatively brief period. Emotions generally affect both physiological and psychological state and can help improve human health and work efficiency if positive, while negative emotions can cause very serious health and behavioral problems.

The detection and monitoring of emotions is essential in many areas such as driving vehicles, in order to act in a timely manner in the event of the presence of a negative emotional state that can dangerously affect the life of the driver. In science, several methods of detecting emotions have been defined, which can be classified into two main categories; one uses human physical signals such as facial expression, speech, gesture, posture, etc., which have the advantage of being easily collected and studied, but which suffer from modest reliability due to from the possibility of not showing true physical signals to hide true emotions. The second category uses internal signals (physiological signals), which include electrocardiogram (ECG), electroencephalogram (EEG), temperature (T), electromyogram (EMG), etc. which are more reliable due to their internal nature and not directly controlled by the human being. In this thesis, we studied the problem of detecting human emotions in a vehicle driver based on the ECG signal. For this end, we have proposed three main contributions related to the detection of emotions. The first one is an approach for optimizing the classification parameters of the categories of ECG signals which among them an abnormal ECG class representing an unusual emotional state.

The second contribution is an improved version of the Random Forest approach for detecting a driver's state of stress. The third contribution is an emotion detection system suggesting a deep learning approach; it is a new convolutional and data-augmenting neural network that considers Heart Rate Variability (HRV-) as an essential criterion for detection. The proposed system has been well developed and proved by validation study and comparison with the similar reference works

proposed in the literature.

keywords: Human emotion detection, Electrocardiogram (ECG), Heart Rate Variability (HRV), Data augmentation, Convolutional Neural Network (CNN), Enhanced Random Forest (ERF), Simulated Annealing (SA).

Résumé

Une émotion humaine est considérée comme un état d'esprit d'un individu qui est complexe et intense, débutant de manière brutale et peut durer pendant une période relativement brève. Les émotions affectent généralement à la fois l'état physiologique et psychologique et peuvent aider à améliorer la santé humaine et l'efficacité au travail si elles sont positives, tandis que les émotions négatives peuvent causer des problèmes de santé et de comportements très graves.

La détection et la surveillance des émotions sont primordiales dans de nombreux domaines tels que la conduite de véhicules, afin d'agir dans le temps opportun en cas de présence d'un état émotionnel négatif qui peut affecter dangereusement la vie du conducteur. Dans la science, on a défini plusieurs méthodes de détection d'émotions qui peuvent être classifiées en deux grandes catégories ; l'une utilise les signaux physiques humains tels que l'expression faciale, la parole, le geste, la posture, etc., qui ont l'avantage d'être facilement collectés et étudiés, mais qui souffrent d'une fiabilité modeste en raison de la possibilité de ne pas montrer les signaux physiques vrais pour cacher de véritables émotions. La deuxième catégorie utilise les signaux internes (les signaux physiologiques), qui comprennent l'électrocardiogramme (ECG), l'électroencéphalogramme (EEG), la température (T), l'électromyogramme (EMG), etc. qui sont plus fiable due à leur nature interne et non contrôler directement par l'être humain.

Dans cette thèse, nous avons étudié le problème de la détection des émotions humaines chez un conducteur de véhicule en se basant sur le signal ECG. Pour cela, nous avons proposé trois contributions liées à la détection des émotions. La première est une approche d'optimisation des paramètres de classification des catégories des signaux ECG qui parmi elle une classe ECG anormale représentant un état émotionnel inhabituel.

La deuxième contribution est une version améliorée de l'approche Random Forest pour la détection de l'état du stress d'un conducteur. La troisième contribution est un système de détection d'émotions en suggérant une approche d'apprentissage profond ; il s'agit d'un nouveau réseau de neurones convolutif et d'augmentation de données qui considère la variabilité de la fréquence cardiaque (Heart Rate Variability -HRV-) comme critère essentiel de détection. Le système proposé a été bien développé et prouvé par une étude de validation et une comparaison avec les travaux de référence similaires proposés dans la littérature.

Mots clés: Détection des émotions humaines, électrocardiogramme (ECG), variabilité de la fréquence cardiaque (HRV), augmentation des données, réseau de neurones convolutifs (CNN), forêt aléatoire améliorée (ERF), Recuit Simulé (SA).

ملخص

تعتبر المشاعر الإنسانية حالة ذهنية للفرد معقدة ومكثفة ، وتبدأ فجأة وقد تستمر لفترة وجيزة نسبيًا. تؤثر العواطف بشكل عام على كل من الحالة الجسدية والنفسية ويمكن أن تساعد في تحسين صحة الإنسان وكفاءة العمل إذا كانت إيجابية، في حين أن المشاعر السلبية يمكن أن تسبب مشاكل صحية وسلوكية خطيرة للغاية.

يعد اكتشاف الانفعالات ومراقبتها أمرًا ضروريًا في العديد من المجالات مثل قيادة المركبات، وذلك من أجل التصرف في وقت المناسب في حالة وجود حالة عاطفية سلبية قد تؤثر بشكل خطير على حياة السائق. تم تحديد عدة طرق لاكتشاف المشاعر والتي يمكن تصنيفها إلى فئتين رئيسيتين؛ في الفئة الاولى، يستخدم المرء إشارات جسدية بشرية مثل تعبيرات الوجه، والكلام، والإيماءات، والموقف، وما إلى ذلك، والتي تتميز بسهولة جمعها ودراستها ، ولكنها تعاني من موثوقية متواضعة بسبب إمكانية عدم إظهار إشارات جسدية حقيقية لإخفاء المشاعر الحقيقية. في حين أن الفئة الثانية تستخدم الإشارات الداخلية (الإشارات الفسيولوجية)، والتي تشمل مخطط كهربية القلب (ECG) ، مخطط كهربية الدماغ (EEG) ، درجة الحرارة (T) ، مخطط كهربية العضل (EMG) ، إلخ. وهي أكثر موثوقية بسبب طبيعتها الداخلية ولا يتم التحكم فيها بشكل مباشر بواسطة كائن بشري.

في هذه الأطروحة درسنا مشكلة اكتشاف المشاعر البشرية لدى سائق السيارة بناءً على إشارة مخطط كهربية القلب. لهذا، اقترحنا ثلاث مساهمات تتعلق باكتشاف العواطف. الأول هو نهج لتحسين معلمات التصنيف لفئات إشارات تخطيط القلب والتي من بينها فئة ECG غير الطبيعية التي تمثل حالة عاطفية غير عادية.

بعد ذلك، و في المساهمة الثانية اقترحنا نسخة محسّنة من نهج *RandomForest* لاكتشاف حالة توتر السائق. بعد ذلك، اقترحنا حلا اخر لمشكلة نقص البيانت، لذلك في المساهمة الثالثة اقترحنا نظام الكشف عن المشاعر باستخدام نهج التعلم العميق، وهي عبارة عن شبكة عصبية تلافيفية جديدة ومضاعفة للبيانات، حيث يعتبر تقلب معدل ضربات القلب (HRV) معيارًا أساسيًا للكشف عن المشاعر. تم تطوير النظام المقترح جيدًا وإثباته من خلال دراسة التحقق من الصحة والمقارنة مع الأعمال المرجعية المماثلة المقترحة في الأدبيات.

الكلمات الرئيسية اكتشاف المشاعر البشرية ، مخطط كهربية القلب (ECG) ، تقلب معدل ضربات القلب (HRV) ، زيادة البيانات ، الشبكة العصبية التلافيفية (CNN) ، الغابة العشوائية المحسنة (ERF) ، التلدين المحاكي (SA) .

Contents

List of Figures **xi**

List of tables **xiii**

List of publications **xiii**

Introduction **1**

1 General context 1

2 Problem statement and motivations 2

3 Main contributions 3

4 Dissertation organization 3

1 Fundamental Concepts **5**

1.1 Introduction 6

1.2 Emotion Recognition 6

1.3 Emotion Recognition Using Physiological Signals: Challenges and Opportunities . . 8

1.4 Affective Computing and Emotion Recognition 10

1.4.1 Emotion Modelling 11

1.4.1.1 Discrete Emotional Models 12

1.4.1.2 Affective Dimensional Models 12

1.4.1.3 Discrete vs Dimensional Models 15

1.4.2 Emotion Recognition Modalities 16

1.4.2.1 Facial Emotion Recognition 16

1.4.2.2 Speech Emotion Recognition Speech 17

1.4.2.3 Body Gestures and Posture 19

1.4.2.4 Physiological Signals 20

1.5 Electrocardiographic Signal 22

1.5.1 Anatomy and Physiology 22

1.5.2 Variability 25

1.6 Conclusion 27

2 Human Emotion Detection: Related Work **28**

2.1 Introduction 29

2.2 Classification of Heart Diseases Based On ECG Signals 29

2.3 ECG and emotion recognition 31

2.4 Emotion detection methods 32

2.4.1 ECG-based Emotion detection methods 32

2.4.2 EEG-based Emotion detection methods 34

2.4.3 ECG-Based Driver emotion detection 35

2.4.4 Data augmentation methods for ECG 37

2.5 Conclusion 39

3 An enhanced random forest for cardiac diseases identification **40**

3.1 Introduction 41

3.2 Motivation: ECG for heart disease diagnosis 41

3.3 Basic concepts 42

3.3.1 Random forest (RF) classifier 43

3.3.2 Simulated annealing (SA) method 43

3.4 An enhanced random forest for ECG classification: our proposal 44

3.4.1 Phase 1: ECG Data Aggregation 44

3.4.2 Phase 2: Preprocessing phase 44

3.4.3 Phase 3: Features Extraction (BD of features) 45

3.4.4 Phase 4.1: Tree classifier building and decision aggregation 46

3.4.5 Phase 4.2: Discovering the optimal RF number of trees 46

3.5 Experimental results 48

3.5.1 Data set description and experiments environment 48

3.5.2 Experiments and discussion results 49

3.6 Conclusion 51

4 A body area network for ubiquitous driver stress monitoring **52**

4.1 Introduction 53

4.2 Motivation: ECG-Based Driver emotion recognition 53

4.3 An enhanced random forest for driver stress detection based on ECG: our proposal . 54

4.3.1 Phase 1: ECG Signal acquisition 55

4.3.2 Phase 2: Preprocessing phase 56

4.3.3 Phase 3: Features Extraction (BD of features) 57

4.3.4 Phase 4: Classification 57

4.4 Experimental results 58

4.4.1 Dataset description and experiments environment 59

4.4.2 Results and discussion 59

4.5 Conclusion 60

5 A New Data Augmentation Convolutional Neural Network for Human Emotion Recognition based on ECG Signals **61**

5.1 Introduction 63

5.2 Motivation: Data augmentation methods for ECG 63

5.3 The proposed ECG data augmentation for human emotion recognition using seven-layer CNN model 64

5.3.1 Data acquisition and preprocessing of ECG signal 65

5.3.2 Data augmentation strategy 65

5.3.2.1 Step 1: Detecting R-waves 65

5.3.2.2 Step 2: Periods calculation of R-R intervals 65

5.3.2.3 Step 3: Random selection of new R-R intervals 66

5.3.2.4 Step 4: R-R intervals concatenation 67

5.3.3 HRV features extraction 68

5.3.4 Architecture of seven-layer CNN model for ECG emotion recognition system 69

5.4 Experimental results 72

5.4.1 Experiments 72

5.4.2 Results obtained and discussion 73

5.4.2.1 Accuracy of valence detection 73

5.4.2.2 Accuracy of arousal detection 74

5.4.2.3 Accuracy of dominance detection 75
5.4.2.4 The confusion matrices: classification correctness 76
5.4.2.5 Precision, recall, and F1-Score 79
5.4.2.6 K-fold cross validation 79
5.4.2.7 PR and ROC curves 80
5.5 Conclusion 86

Conclusions and perspectives **87**
1 Summary of contributions 87
2 Perspectives and future work 88

Bibliography **89**

List of Figures

1.1 Plutchik's Wheel of Emotions 13

1.2 Two-Dimensional Space Model of Emotions 14

1.3 VAD Emotions representation in 3D space 15

1.4 Procedure used in conventional FER approaches 18

1.5 Conventional Architecture of Speech Emotion Recognition System 19

1.6 Two most common ways of modelling the human body: model based on ensemble body parts (left) and kinematic model (right) 20

1.7 General overview of an Emotion Recognition process using physiological signals under target emotion stimulation 22

1.8 The cardiac conduction system. (1) The sinoatrial (SA) node and the conduction system are at rest. (2) The SA node generates the impulse, which spreads across the atria. (3) After reaching the atrioventricular node (AV), there is a delay of approximately 0.1s, allowing the atria to complete pumping blood before the impulse is transmitted to the atrioventricular bundle. (4) The impulse then travels through the atrioventricular bundle and bundle branches to the Purkinje fibers. (5) The impulse spreads to the contractile fibers of the ventricle. (6) Ventricular contraction occurs and the cycle is ready to start again 24

2.1 ECG signal 32

3.1 An ECG one heartbeat normal pattern 42

3.2 Flowchart of the proposed ECG enhanced random forest classifier 45

3.3 The preprocessing and features extraction step 46

3.4 The pseudo-code of Enhanced Random forest 47

3.5 Comparison of Prediction Accuracy with Arrhythmia Data Set 50

3.6 Comparison of Prediction Accuracy with Heart Disease Data Set 50
3.7 Comparison of Prediction Accuracy with MIH-BIH Arrhythmia Database and European ST-T Database 51

4.1 Block diagram of the proposed method using Enhanced Random Forest classifier . . 55
4.2 General placement of physiological sensors 56

5.1 Overall block scheme of the proposed method for ECG-based emotion recognition system 64
5.2 R-wave detection process 66
5.3 Detecting R-waves 66
5.4 Periods calculation of R-R interval. 67
5.5 Random selection of new R-R interval. 67
5.6 Heart Rate Variability 68
5.7 Architecture of seven-layer CNN model for ECG emotion recognition system. 72
5.8 Accuracy comparison between CNN and other machine learning with/without data augmentation (emotion expressed in terms of valence). (a: 70% for training, and 30% for test, b: the 10-fold cross validation) 74
5.9 Accuracy comparison between CNN and other machine learning with/without data augmentation (emotion expressed in terms of arousal). (a: 70% for training, and 30% for test, b: the 10-fold cross validation) 75
5.10 Accuracy comparison between CNN and other machine learning with/without data augmentation (emotion expressed in terms of dominance). (a: 70% for training, and 30% for test, b: the 10-fold cross validation) 76
5.11 Valence dimension: (a: PR curve, b: ROC curve [without data augmentation]), (c: PR curve, d: ROC curve [with data augmentation]) 83
5.12 Arousal dimension: (a: PR curve, b: ROC curve [without data augmentation]), (c: PR curve, d: ROC curve [with data augmentation]) 84
5.13 Dominance dimension: (a: PR curve, b: ROC curve [without data augmentation]), (c: PR curve, d: ROC curve [with data augmentation]) 85

List of tables

2.1 Recent Emotion Recognition Studies based on ECG and EEG signals. 35

3.1 ECG databases used in the experiments 48

4.1 The time interval of each 07 segments of driving of usable bio-signal datasets. 57
4.2 Recognition rates for 02 classes (low/high) with 17 attributes. 60
4.3 Recognition rates for 03 classes (low/medium/high) with 17 attributes. 60

5.1 Notation of features extracted from ECG 70
5.2 Confusion matrix for CNN without data augmentation. 78
5.3 Confusion matrix for CNN with data augmentation. 79
5.4 Confusion matrix for SVM without data augmentation. 80
5.5 Confusion matrix for SVM with data augmentation. 81
5.6 Confusion matrix for NN without data augmentation. 81
5.7 Confusion matrix for NN with data augmentation. 82
5.8 Precision, recall and F1-score of CNN without data augmentation. 82
5.9 Precision, recall and F1-score of CNN with data augmentation. 82
5.10 The results of K-Fold Cross-Validation of CNN (With and without data augmentation). 82

List of publications

- **International journal**

  – NITA Sihem, BITAM Salim, HEIDET Matthieu, and MELLOUK Abdelhamid. A new data augmentation convolutional neural network for human emotion recognition based on ECG signals. Biomedical Signal Processing and Control, 2022, vol. 75, p. 103580.

- **International conferences**

  – NITA Sihem, BITAM Salim, and MELLOUK Abdelhamid. A body area network for ubiquitous driver stress monitoring based on ECG signal. In : 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). IEEE, 2019. p. 1-6.

  – NITA Sihem, BITAM Salim, and MELLOUK Abdelhamid. An enhanced random forest for cardiac diseases identification based on ECG signal. In : 2018 14th International Wireless Communications Mobile Computing Conference (IWCMC). IEEE, 2018. p. 1339-1344.

Introduction

1 General context

Emotions, which affect both human physiological and psychological status, play a very important role in human life. Positive emotions help improve human health and work efficiency, while negative emotions may cause health problems. Long term accumulations of negative emotions are predisposing factors for depression, which might lead to suicide in the worst cases. Compared to the mood which is a conscious state of mind or predominant emotion in a time, the emotion often refers to a mental state that arises spontaneously rather than through conscious effort and is often accompanied by physical and physiological changes that are relevant to the human organs and tissues such as brain, heart, skin, blood flow, muscle, facial expressions, voice, etc. Due to the complexity of mutual interaction of physiology and psychology in emotions, recognizing human emotions precisely and timely is still limited to our knowledge and remains the target of relevant scientific research and industry, although a large number of efforts have been made by researchers in different interdisciplinary fields.

Emotion recognition has been applied in many areas such as safe driving [1], health care especially mental health monitoring [2], social security [3], and so on. In general, emotion recognition methods could be classified into two major categories. One is using human physical signals such as facial expression [4], speech [5], gesture, posture, etc., which has the advantage of easy collection and have been studied for years. However, the reliability can't be guaranteed, as it's relatively easy for people to control the physical signals like facial expression or speech to hide their real emotions especially during social communications. For example, people might smile in a formal social occasion even if he is in a negative emotion state. The other category is using the internal signals—the physiological signals, which include the electroencephalogram (EEG), temperature (T), electrocardiogram (ECG), electromyogram (EMG), galvanic skin response (GSR), respiration (RSP), etc. The nervous system is divided into two parts: the central and the peripheral nervous systems (CNS and PNS). The PNS consists of the autonomic and the somatic nervous systems (ANS and SNS). The ANS is composed of sensory and motor neurons, which operate between the CNS and various internal organs, such as the heart, the lungs, the viscera, and the glands. EEG, ECG, RSP, GSR, and EMG change in a certain way when people face some specific situations. The physiological signals are in response to the CNS and the ANS of human body, in which emotion changes according to Connon's theory [6]. One of the major benefits of the latter method is that the CNS and the ANS are largely involuntarily activated and therefore cannot be easily controlled. There have been a number of studies in the area of emotion recognition using physiological signals. Attempts have been made to establish a standard and a fixed relationship between emotion changes and physiological signals in terms of various types of signals, features, and classifiers.

2 Problem statement and motivations

Electrocardiogram (ECG) has been proven to be a reliable source of information for emotion recognition systems. Automated ECG analysis can identify the affective states of users such as happiness, sadness, and stress, among others.

In the state of the art of ECG machine learning, Random forest is considered as one of the most effective ECG classifier, which motivates the RF choice in many research activities tackling ECG classification like in our study. Nevertheless, applying random forest requests introducing the number of RF trees in manual fashion as a parameter user, considered as a major drawback of this method.

Success of ECG analysis based on machine learning or deep learning relies on rich annotated dataset. Although large amount of ECG recordings is available nowadays thanks to fast developing digital medical devices, constructing a high-quality annotated ECG dataset for deep learning model remains challenging. On one hand, it is demanding for the experienced cardiologists to annotate the recordings.

Therefore, the annotated data are insufficient in quantity and diversity. On the other hand, the incidence of some special emotions like angry or sad is much lower than that of happy or calm emotions, resulting in a highly imbalanced dataset. If not properly treated, the imbalance will lead to biased classification results using deep learning models. Data augmentation may shed a light to the problem. It refers to a procedure that enriches the dataset by introducing unobserved samples.

The main motivations of this work are the following:

- Proposing a new classification system based on random forest classifier which is enhanced by a simulated annealing (SA) algorithm.
- Proposing a new monitoring system to detect driver stress using a body area network that senses the driver ECG
- Proposing a novel data augmentation strategy for ECG signals, aiming to increase the diversity of the samples and balance the number of samples among the classes.

3 Main contributions

In this section, we outline the main contributions of this work. More specifically, we propose novel approaches for emotion detection using ECG signals.

1. An Enhanced Random Forest method proposed for the classification of ECG signals in order to distinguish the normal class from the classes of medical disorders and from an abnormal class which may represent an unusual emotional state (see Chapter 3).
2. A new monitoring system for driver stress detection by an improved version of the Random Forest approach using ECG signals (see Chapter 4).
3. A new data augmentation convolutional neural network for human emotion recognition based on ECG signals (see Chapter 5).

4 Dissertation organization

The rest of this dissertation is organized as follows:

The general introduction presents the problem of the thesis which is the detection of emotions by an ECG signal based on the variability of the heart rate. The motivations, contributions and organization of the thesis were also initiated in this part.

Chapter 1 This chapter illustrates the basic and fundamental concepts concerning the electrocardiogram signal and the detection of emotions. In addition, the main methods of acquiring emotions have been analyzed and discussed in this chapter as well as the different models of emotions and the different modalities used for the recognition of emotions.

Chapter 2 This chapter is devoted to the different approaches proposed in the literature for the detection of emotions. In this chapter, we have exposed a significant number of related works that have been used in emotion recognition using ECG and EEG which are described and discussed in detail. Specifically, we focused on similar work to detect driver emotions using ECG signal. Furthermore, we presented the existing methods of data augmentation for the ECG signal, these methods are used to solve the problem of the limitation of the ECG datasets that are required for initiated learning in the proposed system.

Chapter 3 In the third chapter, we presented our first contribution which is the Enhanced Random Forest method proposed for the classification of ECG signals in order to distinguish the normal class from the classes of medical disorders and from an abnormal class which can represent an unusual emotional state. Improved Random Forest method combines the traditional Random Forest method and the Simulated Annealing algorithm to find the best parameters to introduce to the classification system for better detection

Chapter 4 A new monitoring system for driver stress detection by an improved version of the Random Forst approach has been proposed. This proposal analyzes and monitors the electrocardiogram (ECG) signal of the driver while driving in order to discover his state of stress belonging to one of the following three levels; low, medium or high.

Chapter 5 Presents a new emotional detection system based on a deep learning approach that is a network of convolution neurons (CNN) and increase of data. On the basis of an ECG signal, we proposed to enrich the dataset (dataset) with a significant number of representative ECG samples, which are generated according to a process of randomization, concatenation and rehabilitation of episodes Realistic ECG. Therefore, a new seven-layer CNN classifier is obtained aimed at detecting human emotions in terms of levels of valence, excitement and dominance.

Conclusions and perspectives The general conclusion concludes this dissertation and gives an insight into our ongoing and future work and perspectives in the area of emotion detection using ECG signals.

Chapter 1

Fundamental Concepts

Contents

1.1 Introduction **6**

1.2 Emotion Recognition **6**

1.3 Emotion Recognition Using Physiological Signals: Challenges and Opportunities **8**

1.4 Affective Computing and Emotion Recognition **10**

1.4.1 Emotion Modelling 11

1.4.1.1 Discrete Emotional Models 12

1.4.1.2 Affective Dimensional Models 12

1.4.1.3 Discrete vs Dimensional Models 15

1.4.2 Emotion Recognition Modalities 16

1.4.2.1 Facial Emotion Recognition 16

1.4.2.2 Speech Emotion Recognition Speech 17

1.4.2.3 Body Gestures and Posture 19

1.4.2.4 Physiological Signals 20

1.5 Electrocardiographic Signal **22**

1.5.1 Anatomy and Physiology 22

1.5.2 Variability 25

1.6 Conclusion **27**

1.1 Introduction

Affective computing has become an increasing hotspot and principal direction in the area of human-computer interface. Emotion recognition is particularly significant in foregoing kingdom, which mainly includes the ways of face expression, speech, gesture, physiological signal and so on. Additionally, Emotion is considered a physiological and psychological expression associated with moods and personalities of individuals. However, physiological signal can truly reflect our emotions, which is more objective and robust than any other ways. As a result, sensing technologies integrated into wearable devices coupled with machine learning and deep learning techniques have recently been used to analyze physiological signals in order to classify or quantify human emotions. In order to better understand the context of this work, some knowledge needs to be acquired first. Therefore, in this chapter, we are presenting an overview of the fundamentals concerning both the electrocardiogram signal(ECG) and the emotion recognition theory and principles and the electrocardiographic signal, its variability and main acquisition methodologies. Then, the existing emotion models are presented and detailed, as well as the different modalities used for emotion recognition. Then we discuss the data augmentation strategy especially for ECG signal.

1.2 Emotion Recognition

Affective computing is a growing field, involving disciplines like engineering, psychology, cognitive science and even sociology that explores how the use of technology enables the recognition and interpretation of human emotions or affects. Emotions are considered the way to communicate beyond words. In this way, the automatic recognition of emotions may be an exciting opportunity for human-computer interaction or even the gaming industry. The exponential increase of smart technologies in society brings the desire to make them even more custom-made, by assessing the needs of their owner and choosing the most appropriate solution. In this way, machines with the capability of assessing human emotions may revolutionize the human-robot interaction [7].

However, other areas besides robotics [8; 9] may be interested in this automated emotion evaluation such as marketing, education or even the entertainment industry, as already mentioned. In marketing, advertisement can become more effective by taking into account the emotional state of customers [10]. As far as education is concerned, learning processes may be improved by analyzing emotional responses from students [11; 12]. Finally, regarding the entertainment industries, like gaming, more appropriate entertainment can be proposed by assessing the audience's response to the ones already available in the market [13].

Another example is call centres that could largely benefit from their customers' emotion recognition, allowing them to detect problems and maximise customer satisfaction [14]. However, the design of such robust algorithms for automated recognition of human emotions is still a major challenge [15].

Emotions are complex processes, including feelings, body language, cognitive reactions and behavior or thoughts [16]. Different models have been proposed for automatically recognise emotions, taking into account the way all of these processes may interact with each other. However, there's still no universally accepted formulation to model emotions. Nevertheless, to use engineering principles to the recognition of such a personal and non-exact parameter as the human emotion, affects need to be conceptualized in a clear and strict way. There are mainly two categories regarding the way emotional models are conceived: The first one considers Discrete Emotional Models (DEM), focusing on the six most basic emotions (happiness, sadness, anger, fear, surprise and disgust). This method considers that, regardless the situation or culture, humans perceive the environment and react to it in a similar way and with a distinguishable emotion. In this way, the main goal is to identify and label standard emotional states. On the other hand, Affective Dimensional Models (ADM) characterize an emotion as a set of parameters, forming an n-dimensional emotional space, and with arousal, valence, and dominance being the most commonly used dimensions [15] [17].

Arousal can be described as the measure of emotional stimulation (activation level) and it can vary from low/calmness to high/excitement. Valence is a measure of pleasure, defined by a polarity of positive or negative feelings. Finally, dominance is related to the subject's feeling of control, indicating if the human feels without control or empowered (dominant vs. submissive) [18]. For example, fear has a high level of arousal, negative valence and is a submissive emotion [19]. Having this said, pattern recognition approaches must be applied for affect recognition, relying on the acquisition of data with different affective states from subjects experiencing a given situation. Different modalities have been used to collect this data, from facial expressions, peripherical physiological signals or even speech intonation. Regarding the emotional stimuli, it may include media, like video, audio or even music clips, as well as the creation of different kinds of environments (relaxed, stressful, among others) [20–24].

Nowadays, most experiments developed in order to recognize emotion focus on behavioural (visible or audible) modalities [25]. Human communication is composed of verbal and non-verbal components, that are able to carry emotional information. Daily, humans rely on their own interpretation of facial and speech tone to infer the emotional states of other people. Thus, this recognition relies on the techniques used by humans to understand each other [14; 26–28]. Although emotion recognition from facial expressions or voice tone has been improved, there's still some uncertainty in the use of those kinds of evaluation parameters, since they might be purposely faked and altered. Individuals can consciously regulate their own emotions or naturally suppress them, without even acknowledging there are doing it [29; 30].

On the other hand, physiological signals are acquired in a more unconscious mode, allowing for a more trustworthy data collection. Physiological signals should provide relevant insights on emotion, since they're associated with the autonomic nervous system's responses [31; 32]. Emotion is both a psychological and a physiological expression, associated with mood, personality and all the cognitive processes involved. Besides, these signals are recorded in a continuous way, which enables detecting emotion variations through time [33]. Different physiological signals have been used in order to efficiently detect emotions, like electroencephalogram (EEG) [34], electrocardiogram (ECG) [35; 36], galvanic skin response (GSR) [32], electromyogram (EMG) [37], among others. The electrocardiogram (ECG) is a powerful signal, being considered one of the most used diagnostic tools in medicine. Recent researches have proven to be a prospective technique for emotion recognition, allowing to measure signals that can be affected by changing emotional states

1.3 Emotion Recognition Using Physiological Signals: Challenges and Opportunities

As already mentioned, one of the main benefits of detecting emotions using physiological signals is that they are involuntary and uncontrollable reactions of the body, and thus, difficult to hide or mask [38].

The human heart is known to be affected by what people feel. In a non-scientific way, people refer to the heart as what aches when something bad happens, or even what they give to someone they love. Expressions like "heartbreak", "big heart", "heart of stone" or "heavy heart" are just some examples of different ways of expressing a variety of emotions. However, in a more logical view, the heart is also associated with what humans feel. When people are nervous, the heart beats faster, when they're relaxed, the pulsation goes down. So it becomes reasonable to think about the possibility of detecting variations in emotions and affect through the heart's changing behaviour [29].

The electrocardiogram is a physiological signal obtained by the cyclic contraction and recovery of the heart [39]. In this way, the electrocardiogram can be seen as a code, conveying all the changes correlated to the humans' emotional states, especially by variations in the heart rate and other indicators. Various studies show that the use of ECG signal provides relevant information that can be correlated with emotion [40].

However, there are still several drawbacks regarding the use of physiological signals and specifically the electrocardiogram. Although it is accepted that physiological signals are affected by emotions, the effects on the waveform patterns are still unknown and not well defined. For this reason, many researchers are still trying to find the most emotion-related features, that can provide clearer information.

Besides, the experimental protocols are far more complex than those from behavioural emotion research. This problem is related to the need of obtaining high-quality physiological data, which is difficult, and the requirement of obtaining genuine emotions, which depends largely on the emotion elicitation material. The signal acquisition is also a challenge, since it is a more invasive process, with the sensors in contact with the human body during the recording session. Thus, proper methods and equipment should be chosen, having into account all the challenges mentioned, and trying to diminish these problems. Nonetheless, all the limitations mentioned above lead to commonly small and restrained databases, with a reduced amount of data, which can be highly problematic for deep learning approaches to be developed.

Another drawback is related to the labelling process, since physiological signals are subjective and it is difficult to establish their ground truth [29; 39]. Moreover, most databases and experimental protocols don't have a large number of subjects nor a relevant number of signals per subject. Due to this limited amount of data, the performance of the classifiers applied is compromised. To solve this problem, more subjects can be included in the experimental protocols, which should also become longer so that the number of data increases, or more samples can be used (for example, by considering multiple segments of the same physiological signal). The use of segments i highly recurrent in order to obtain a "fake" data enlargement, however, this multiple segmentation of the same signal implies that the same label is considered to all segments, which can sometimes be misleading. Furthermore, since random splits are usually applied, segments from the same signal are found in train and test sets, which can result in misleading performances presented in the literature, that must be evaluated under more controlled and realistic settings.

Having this said, signal and subject-independent settings can be one of the main problems regarding physiological signals, mainly due to the lack of large amount of data. Furthermore, concerning the ECG signal that, in normal conditions, presents the same behaviour and deflections for all subjects, it is considered to have a high degree of variability. This variability can be intra-subject, consisting of variations within the consecutive heartbeats of the same subject, or inter-subject, corresponding to variation between heartbeats of different subjects [39].

Especially the inter-subject variations can become a problem when the major goal is to detect specific and equal patterns to equal states of emotion. Having this said, the major problem nowadays is associated with this pattern detection when data shows a wide range of behaviour to the same affect felt by different people. When in biometric research, these inter-subject features can be highly positive, allowing for a more efficient recognition of each individual. However, when working to correlate an emotion to a given pattern, it can be challenging.

1.4 Affective Computing and Emotion Recognition

Emotion can be described as a “strong feeling deriving from one’s circumstances, mood, or relationships with others” or even as an “instinctive or intuitive feeling as distinguished from reasoning or knowledge” [41]. It is common sense that emotions affect both human physiological and psychological status, playing an important role in a person’s daily life and how someone may deal with his own happiness, victories, or even losses or disappointments. However, it is also true that emotions are still believed to be inherently non-scientific sensations, far from rational thought or common logic, being marginalized from science.

However, this public perception is not completely correct. Although emotions arise from a human’s ability to feel and process what happens around him, they also have a large impact as far as essential cognitive processes are concerned. Lisetti [42] highlights results from neurological literature indicating how emotions are not only associated with human creativity and intelligence, but also with basic rational thinking and decision making.

Furthermore, Lin Shu et al. [38] defines emotion as a “mental state, that arises spontaneously rather than through conscious effort and is often accompanied by physical and physiological changes that are relevant to the human organs and tissues such as the brain, heart, blood flow, muscle, facial expressions, voice, etc", pointing out the close link between emotions and rather scientific parameters like physiological signals and basic biology functioning.

With the exponential growth of human-computer interactions (HCI), it becomes reasonable to consider that the interpretation of emotions is an important and fundamental asset to enable accurate and efficient intercommunication [40; 42]. From this need, Affective Computing (AC) emerges as a new and exciting field that can enable smarter and more humane technology.

AC tries to assign to computers the ability to perceive, observe and even generate affect responses, increasing human-computer communication quality. This field is currently one of the most emerging and active topics in research, due to its large and promising spectrum of applications, in different areas. Affective Computing involves multidisciplinary knowledge like psychology, physiology and computer sciences [43]. Its great potential translates into a wide range of applications in several environments, like robotics, marketing or even education. The development of recommendation systems also becomes possible since Affective Computing will allow understanding customers preferences and opinions.

Within Affective Computing, two distinct research areas can be identified: sentimental analysis and emotion recognition [44]. While Sentiment Analysis consists of a simple classification task between a positive/neutral/negative state, emotion detection involves a more detailed and thorough methodology that aims to distinguish between a set of emotions. In this way, product reviewers can use Sentiment Analysis to assess how the product was perceived by its customers (liked or not). On the other hand, Emotion Recognition is able to differentiate if someone is angry, sad or bored. As it can be understood, these two areas can overlap, since a happy customer translates into a positive reaction. However, Emotion Recognition is a finer discrimination and assessment of a general reaction to a given stimulus. This scientific analysis of emotions and the capability of recognizing them depends on a correct emotion modelling, followed by proper data collection and the application of specific methodologies, which will be further explained in the following sections.

1.4.1 Emotion Modelling

Emotions have been studied for centuries and philosophical studies around them can date back to the ancient Greeks and Romans. The desire of defining emotions as quantitative parameters, in order to evaluate and assess them in a more scientific and logical way is as old as it can be, and started with Cicero, when he tried to organized emotions into four basic categories: metus (fear), aegritudo (pain), libido (lust) and laetitia (pleasure). In the late 19th century even Darwin focused some of his attention on emotions, developing a theory that stated that emotions evolved via natural selection [45]. However, since emotions are complex processes, involving body language, cognitive reactions, feelings and thoughts, modelling them is a very challenging task. Besides these first attempts, psychologists have been trying to develop a logical and universally accepted emotion model for decades, however, this goal hasn't yet been completely achieved. Nevertheless, the main scientific models developed and used in Emotion Recognition are the Discrete Emotional Models (DEM) and the Affective Dimensional Models (ADM).

1.4.1.1 Discrete Emotional Models

In the early 1970s, Ekman [46] conceived emotions as discrete and measurable categories, being psychological states that fit specific criteria. Six basic emotions were defined: happiness, sadness, anger, fear, surprise and disgust. All other emotions were considered to be combinations or reactions to these. Discrete models suggest that affective states are universal among people, regardless of their differences in gender, origin, or even moral beliefs. Also, when in a similar situation, people would react in the same way, showing analogous physiological expressions and patterns [17] However, in 1980, Plutchnik [47] extended this basic emotions list, by considering eight emotions of joy, trust, fear, surprise, sadness, disgust, anger and anticipation. Figure 1.1 presents the wheel model proposed by Plutchik, which describes emotions ranked by intensity, with the stronger ones in the centre and the weaker positioned at the flower blooms. The colour use represents that basic emotions can be combined to form complex emotion.

Finally, Izard [49]defined ten basic emotions: joy, surprise, sadness, fear, shyness, guilt, anger, disgust, shame and contempt. He hypothesized that these emotions were formed during human evolution, being a simple brain circuit. Although these simplistic models would allow an easier view of emotions, they become unsatisfactory and limited when analysing complex emotions, which cannot be described with a discrete label and need a more quantitative evaluation [38].

1.4.1.2 Affective Dimensional Models

In 1989, the first dimensional model was proposed by Russel et al. [50], showing that there would have to exist some evaluation parameters that could describe emotions, like intensity or positiveness. These models suggest that emotions are not discrete but continuous, and, according to Lang et al. [51], they can be projected into a two-dimensional space model, by valence and

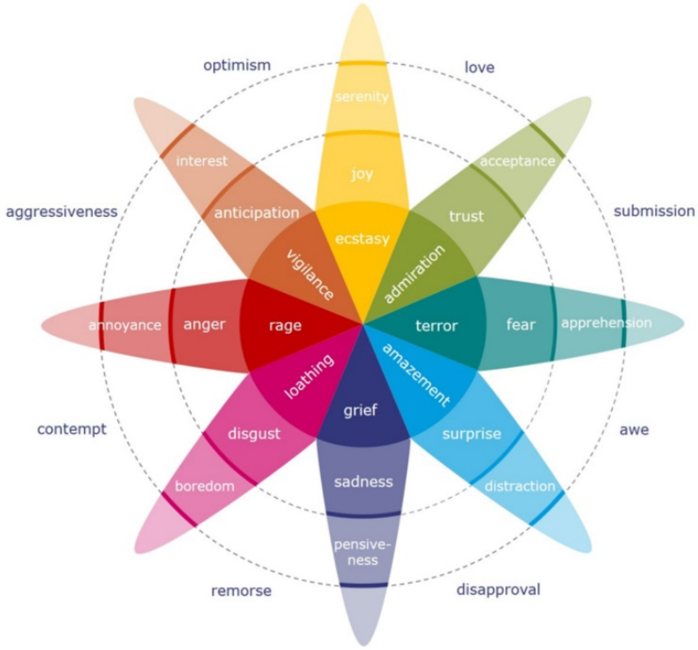


Figure 1.1: Plutchik's Wheel of Emotions, from [48]

arousal. Valence is a measure of pleasure, ranging from negative/unpleasant to positive/pleasant. On the other hand, arousal describes the level of activation and intensity concerning emotional stimulation, varying from low/calmness to high/excitement, or even passive and active. In this way, as shown in Figure 1.2, different emotions can be plotted, having a correspondent value of arousal and valence. For example, sad has a negative valence, indicating that it is not a pleasant emotion, and a passive arousal, proving a low activation level. On the other hand, angry is also a negative valence emotion but it presents a high level of activation and excitement, thus having an active arousal.

This dimensional grid became highly attractive due to its higher flexibility in describing an emotion besides using just words, and the possibility of considering emotional variations over time. Other metrics emerged through time like dominance, as it can be seen in Figure 1.3 [52]. Dominance corresponds to someone's level of control, varying from submissive to dominant. Having into account other measurements to characterize emotions like this one, it becomes possible to obtain a more complete discrimination between emotions. For example, while fear and anger both have negative valence and active arousal, dominance allows to divide them into the submissive and dominant axis, respectively [38]. Recently, another study presented by Warriner et al [54], adding the third dimension dominance to form the Valence-Arousal-Dominance space (VAD). So, Three main affective qualities were considered by the psychologists to describe the human emo-

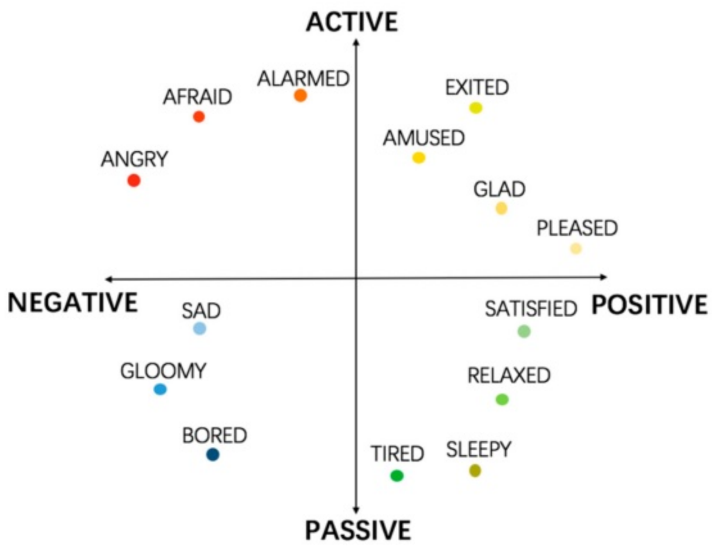


Figure 1.2: Two-Dimensional Space Model of Emotions, from [38]

tions which are: valence, arousal, and dominance [55]. Figure 1.3 illustrates a 3 dimensional emotion model, showing Valence, Arousal and Dominance (VAD in short). These emotions are explained below.

- **Valence:** represents the fear or happiness; it is the positivity or negativity of an emotion.
- **Arousal:** represents the intensity of emotion elicited by a stimulus. Precisely, a high arousal is an anger status and a low arousal represent a sadness status.
- **Dominance:** constitutes the state and level of control do by the stimulus, we distinguish two states such as dominant (i.e. with control) or submissive (i.e. without control).

As shown in figure 1.3, being Emotions are categorized on a 3D plan, the x-axis represents the valence, y-axis is for arousal and zaxis illustrates the dominance. There are 15 emotions grouped into 5 clusters: C1 represents the happy group with the emotions of Happy, Joy, Fun, Exciting, then, C2 represents is for the love group comprising: Love, Cheerful, Lovely, C3 shows the Sentimental Group which includes Depressing, Sentimental, Mellows, the C4 group is the sad group, with the emotions: Sad, Melancholy, and Terrible. C5 is the hate group with the emotion of Shock and Hate. Therefore, a human emotion is formed a combination of the three dimensions valence, arousal and dominance. In the happy group (Joy, Exciting, Happy and Fun), the valence and arousal are relatively high. However, in the sad group composed of Sad, Depressing and Melancholy the valence in low. By this way, positive and negative emotions could be expressed by numerical values

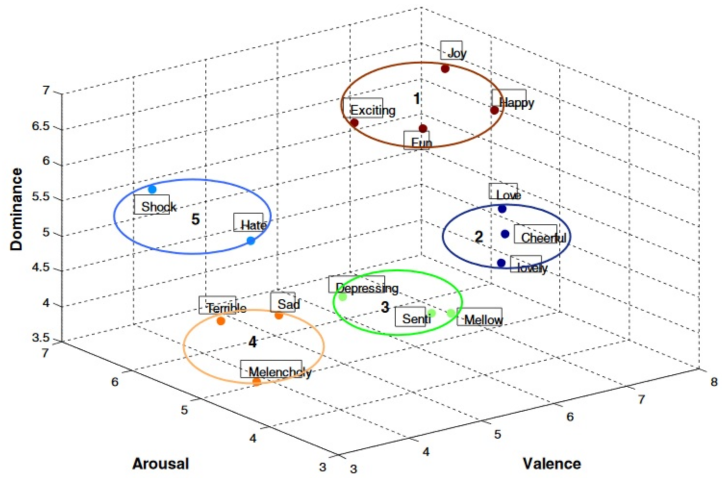


Figure 1.3: VAD Emotions representation in 3D space, from [53]

[53]. For example, when we have a perpendicular plan at valence 6.5, all the emotions on the right side are called positive emotions, however, on the other side, the emotions are considered as a negative emotions. We can also see the values of an emotional state of fun, which are 6.85, 5.85, and 6 for valence, arousal and dominance, respectively. According to [53], the range of the valence, arousal and dominance are fixed as following:

- **Valence value range:** (Low: 1 - 4.5), (Medium: 4.5 – 5.5), and (High: 5.5 – 9)
- **Arousal value range:** (Low: 1 - 4.5), (Medium: 4.5 – 5.5), and (High: (5.5–9)
- **Dominance value range:** (Low: 1 - 4.5), (Medium: 4.5 – 5.5), and (High: 5.5 – 9)

Human emotions can be the reason of many physiological changes through various involuntary neurological responses such as: respiration [56; 57], skin electricity [32; 58], temperature [59], muscular activity, [60] and cardiologic activity. [61; 62]. Various studies proved that ECG can be efficiently used for detecting human emotion [63; 64].In the next subsection, we explain how ECG can help to detect human emotions.

1.4.1.3 Discrete vs Dimensional Models

Concerning Discrete and Dimensional models, it is important to understand that both have advantages and disadvantages in their attempt to distinguish between emotions. Regarding the Discrete Emotional Models, as it can be easily understood, there are a variety of factors that can play an important role on someone's behavior or emotional response, which means that it is wrong to consider that all people would act in the same way when exposed to the same environment or situation. Furthermore, emotions are also complex processes, related to both physiological and neural activities, and, for this reason, it is limiting to consider a small list of basic emotions. On the other hand, even the Affective Dimensional Models are not enough to translate the emotional complexity already described. Besides assuming independence between axis, eliminating all possible relations between dimensions, which can also be a little uncertain and incorrect, these models do not consider the possibility of two or more emotions being felt at the same time. However, some new dimensional models are emerging, such as the Hourglass Model, trying to reproduce the entire range of emotions and considering situations where more than one emotion is being felt at the same time.

1.4.2 Emotion Recognition Modalities

Concerning Emotion Recognition and Affective states differentiation, it is essential to find a way of measuring the subject's emotional variations over time. As already mentioned, emotions are expressed by people in their daily life, through their bodily reactions, facial alterations or even the changes that can be detected in someone's voice or intonation. Yet, emotions are also closely related to neural activity and the body's physical and physiological states. In this way, different modalities can work as emotion indicators and be used to detect affective variations. Having this said, emotion recognition methods can be classified into two main categories. One is by using physical and more visible signals such as facial expressions, speech alterations, gestures or posture. The other has into account physiological signals, including EEG [34; 35; 65], ECG [30; 35; 36; 66], GSR [32; 67; 68] [26; 53; 54], EMG [37; 69], HRV [43; 70], among others. The present section focuses on the current modalities used for emotion recognition, analyzing the advantages and disadvantages of each.

1.4.2.1 Facial Emotion Recognition

In everyday life, humans rely on their own perception of emotions through the analysis of people's facial expressions and body movement. Especially the latter plays an important role as the main non-verbal communication channel. The face conveys diverse information regarding someone's age, sex, identity, background or even what they are feeling. In this way, it is not surprising that various behavioral scientists showed interest in how facial expressions could indicate a large amount of information from someone [71]. At the beginning of 1970, computer scientists started to use the face as a biometric modality, since it is the main natural method to recognize a given person. Later, the idea was to analyze and synthesize facial expressions. Ekman and Friesen developed one of the leading methods in facial expression recognition, the Facial Action Coding System (FACS) [72]. This system describes facial expressions by considering individual muscle movements, denominated Action Units (AUs) - the smallest movements that can be distinguished - that allow detecting micro-expressions.

Ekman's work inspired many researchers to study facial expressions using both images and videos. With time, image processing techniques started to be used and applied in facial emotion detection. Nowadays, Facial Emotion Recognition (FER) techniques are normally composed of three major steps: preprocessing, feature extraction, and classification.

Preprocessing consists of a preliminary process that can be used to improve FER performance. It is normally done before the feature extraction process, facilitating both extraction and further classification. Regarding images, most common preprocessing methods include image clarity, contract adjustments, scaling and other possible enhancements [73]. After that, some features are extracted and fed into a classification system. Finally, the output consists of one of the pre-selected emotion categories or labels.

However, it is possible to divide FER techniques into two groups, considering if features are handmade or automatically produced through a deep neural network. As far as handcrafted features are concerned, they can be further discriminated in "feature-based" and "region-based". "Feature-based" consists of detecting specific facial structures like the eyes or the corners of the mouth. On the other hand, "region-based" focuses on specific regions of the face to detect expression variations. For example, this analysis can be done by taking into account the eye/eyebrow and mouth regions, discarding the rest of the face. Figure 1.4 presents the three main steps in conventional FER approaches: facial structures detection, feature extraction and expression classification [74].

1.4.2.2 Speech Emotion Recognition Speech

Speech is the most natural way of communication and interaction between humans. In addition, there is a lot of information regarding how people are feeling through their way of talking, not only by what someone says but also having into account the speech intonation, trembling or firmness. When people are nervous, their voice can become more shaking, however, when in an excited state of happiness, people tend to talk louder and wider. These facts motivated different re-

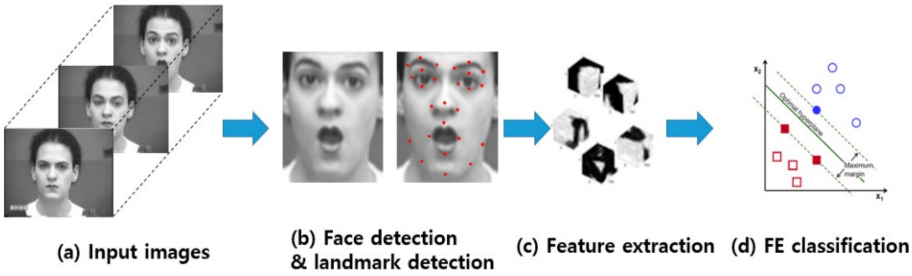


Figure 1.4: Procedure used in conventional FER approaches, from [74]

searchers to study speech alterations and relate them with possible emotion variations. Since the 1950s, different research has been developed on speech recognition, by converting human speech into words. However, when it comes to speech emotion recognition, research and achievements are still few [75]. Speech emotion recognition consists of extracting the emotional state of a given person by analyzing his or her speech. This goal can be achieved by using audio or transcriptions into text processing. However, it is not an easy task due to the variability introduced by different speakers, speaking styles and speaking rates [76].

Similarly to Facial Emotion Recognition and other classification problems, the initial step to be done, after possible preprocessing, is feature extraction, by identifying suitable features that may efficiently characterize emotions. These features can be local, by dividing the signal into small segments and analysing them in a stationary way, or global. Features like pitch and energy can then be extracted from each frame and further used to detect emotion variations. Global features are calculated as statistics of all speech features. Although there is some disagreement on which type of feature is better, most researchers consider that global features are more effective in terms of classification time and accuracy [14]. Speech features can also be grouped into four categories: continuous, qualitative, spectral and TEO-based features (Teager energy operator).

Figure 1.5 represents a possible method of speech recognition. Deep learning methods are also used for speech emotion recognition, avoid the need for manual feature extraction. Some of the most commons classifiers used are Support Machine Vectors (SVM) and Artificial Neural Networks (ANN). However, HMM (Hidden Markov Model) is the most used classifier in speech emotion classification and recognition [77; 78].

Although some good results have already been achieved, this field is still attractive and full of opportunities to be improved, by tackling problems like speaker-dependent classification. However, speech emotion recognition is not always completely trustworthy. Although it cannot be so

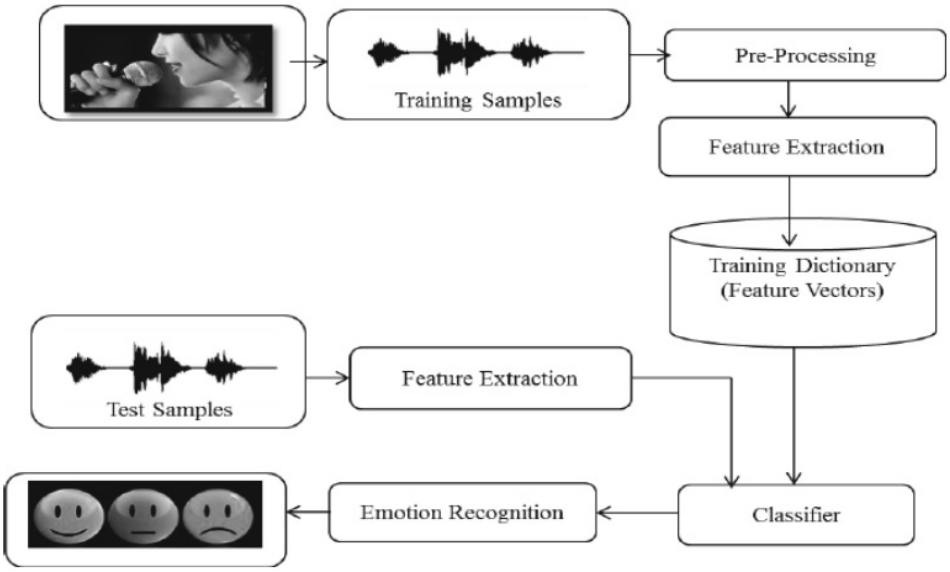


Figure 1.5: Conventional Architecture of Speech Emotion Recognition System, from [79]

easily manipulated and altered as facial expressions, it is possible to control parameters such as rhythm or intensity, in order to mask some emotional states.

1.4.2.3 Body Gestures and Posture

Comparing with affect analysis from speech and facial expression variations, body gestures and posture are regularly overlooked and diminished as strong emotion indicators. However, the power of body language has been gaining more and more attention, and it is already common sense that body gestures can provide information regarding if someone is, for example, comfortable or nervous. According to Mehrabian's 7-38-55 principle, the percentage distribution of a message is 7% verbal signals and words, 38% strength, height and rhythm and 55% facial expressions and body posture [80].

According to [81], body language includes a variety of non-verbal indicators, from facial expressions to body posture, eye movement, or the use of personal space. The hands are also a great source of body language information [82], and nowadays, it is starting to receive some attention, being used by politicians during their speeches or debates. In the same way, details like the head positioning or chin lifting angle can also be a source of information, conveying emotions and intents [83].

More importantly, body gestures and posture are generally natural and unintentional, unlike facial expression or speech intonation, which can be purposely masked or faked. Thus, increased

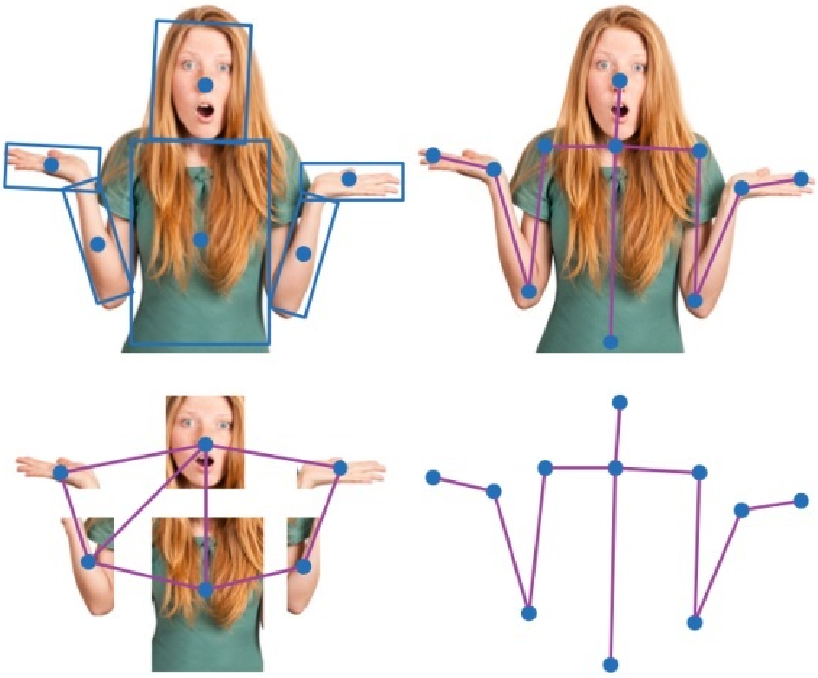


Figure 1.6: Two most common ways of modelling the human body: model based on ensemble body parts (left) and kinematic model (right), from [84]

interest in using body language to emotion recognition started to grow and be further researched.

As far as body analysis is concerned, human modelling is an important preparation step, that may influence all the subsequent phases. The most common ways of modelling the human body are either as an ensemble of body parts as a kinematic model (see Figure 1.6). In the first way, different body parts are independently detected and some restrictions can be applied to refine the detection. On the other hand, the kinematic model consists of a collection of interconnected joints with predefined degrees of freedom identical to the human skeleton [84].

1.4.2.4 Physiological Signals

The nervous system is composed of two distinct parts: the central and the peripherical nervous systems (CNS and PNS). Considering the PNS, it is divided into the autonomic and somatic nervous systems (ANS and SNS). The ANS is associated with the sensory and motor neurons, allowing the connection and communication between the CNS and all internal organs. In this way, physiological signals respond to both the CNS and the ANS, which implies their natural and unconscious nature. Nowadays, the autonomic nervous system (ANS) activity is viewed as a determinant component of the emotional response in a variety of recent theories concerning emotion [85]. Although there's still some disagreement regarding the degree of specificity of ANS activation in emotion, it can be inferred that emotions have an impact on ANS, which, in its turn, influences physiological signals.

Psychophysiology is the branch of psychology that correlates psychological states with physiological variations and measurements. Although someone can choose not to externally show his/her emotion, there's an inevitable change in physiological signals that cannot be hidden or avoided, since the ANS sympathetic nerves are activated whenever a person is positively or negatively stimulated [86]. This sympathetic activation leads to heart, respiratory and blood pressure rate variations, considered some of the most common reactions of the human body to a given emotion [87]. In 1983, Izard and Fridlund were responsible for one of the earliest works considering emotion recognition and physiological signals. They used Linear Discriminant Analysis on facial EMG activity, which became a landmark research since it was able to prove the existence of a correlation between physiological data and emotional states [88]. After that, a great number of studies regarding emotion recognition using physiological signals have been conducted, which can be found in the literature. However, it is still uncertain how emotion variations translate into actual pattern alterations in each physiological signal.

It is highly relevant to have an efficient and successful data collection when developing an emotion recognition system based on physiological data. Unlike facial or speech variations, physiological signals are harder to collect since they are highly prone to noise and ask for more complex setups. On the other hand, image and video can be easily acquired even by non-specialists [89]. Since the ANS controls physiological signals, there's the need to naturally induce a given emotion. For this, different emotion elicitation techniques can be used, like pictures [90], videos [18] or even music [33].

Regarding the specific methodology of emotion recognition using physiological data, it can be divided into two major categories, being the first the usage of more traditional machine learning methods, and the other using deep learning methods.

Like it was already mentioned in other modalities, automated classification tasks normally rely on these two options (see Figure 1.7). The first one implies handcrafted feature extraction and optimization, while deep learning methods learn from data that was practically unaltered. Especially when considering end-to-end approaches, preprocessing is almost non-existent since the network should be able to deal with the raw signals and extract all important patterns to recognize emotions. In this way, using raw signals makes the network more robust in its detection [25; 91].

In model-specific methods, in which feature extraction is done, there's the need to explore

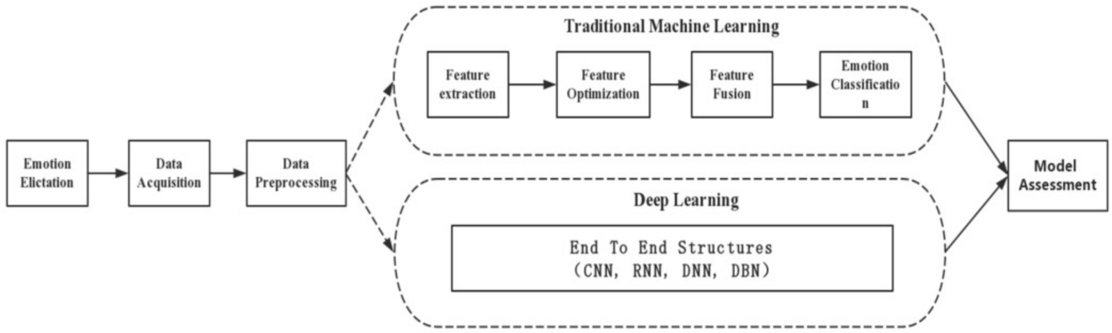


Figure 1.7: General overview of an Emotion Recognition process using physiological signals under target emotion stimulation, from [38]

emotion-specific characteristics concerning each different physiological signal, in order to focus on relevant features that may improve the model performance. On the other hand, deep learning approaches don't ask for such specific knowledge or challenging feature selection.

1.5 Electrocardiographic Signal

1.5.1 Anatomy and Physiology

The human heart is located in the mediastinum, a place within the thoracic cavity and between the lungs, slightly to the left of the sternum. This muscular organ is essential to human life since it works as a pump, by contracting and forcing the blood through the blood vessels of the body. The heart has three major functions: (1) generating blood pressure, (2) routing blood, which allows for the separation between the pulmonary and the systemic circulation, and (3) regulating blood supply, in which the heart rate and its force contraction change in accordance to the human metabolic needs [92; 93].

Regarding its anatomy, the heart is enclosed in the pericardial membranes, composed of three layers. The more external is the fibrous pericardium, made of a dense fibrous connective tissue that protects the heart and anchors it to the surrounding structures. In the middle, the serous pericardium, a thin serous membrane, is folded into two layers: parietal and visceral. The latter is the one in contact with the surface of the heart, also known as epicardium, which means "upon the heart". The epicardium is considered to be the most inner layer of the heart wall, followed by the myocardium (middle layer), and the endocardium (outer layer). The myocardium is mainly composed of cardiac muscle, forming the bulk of the heart and the walls of the four chambers. The endocardium covers both the valves and chambers, preventing abnormal clotting [94].

The heart is formed by four chambers. The two upper ones are the right and left atria, which are separated by the interatrial septum. The lower chambers are the right and left ventricles, presenting thicker and stronger walls, and being separated by the interventricular septum. Caval veins carry blood from the body to the right atrium. From this atrium, the blood is then led to the right ventricle upon atrial contraction, passing through the right atrioventricular (AV) valve, also called tricuspid. In the ventricle, the blood is pumped to the lungs, carried by the pulmonary artery. On the other hand, the left atrium receives blood from four pulmonary veins, that flows to the left ventricle. The valve separating these two chambers is the mitral or bicuspid, and after arriving at the ventricle, the blood is finally pumped to the body through the aorta, the largest human artery [95].

This sequence of events occurs in a cyclic way with each heartbeat, and it is dominated as the cardiac cycle. The electrical activity of the myocardium regulates this cycle, which is stimulated to contract without any external stimulation [96], and the electrical impulse follows a specific order and route, throughout the myocardium [39; 97] (see Figure 1.8):

1. **Atrial depolarisation** - The sinoatrial node (SA), considered the natural pacemaker of the heart, generates the impulse and is the fastest structure in the myocardium to depolarize. The electrical impulse follows its path, reaching both atria and the AV node;

2. **Atrial depolarisation complete** - After its depolarization, the atria contraction begins. Conduction through the AV node is ten times slower than through the surrounding heart tissue, delaying the impulse for 0.1s. This allows the atria to finish its contraction before the impulse gets into the ventricles and it also protects the ventricles from high atrial rates during possible atrial arrhythmias. The electrical conduction occurs through the AV node to the bundle of His, that divides into left and right bundle branches, carrying the electrical impulse into the left and right ventricles, respectively;

3. **Ventricular depolarisation** – Upon reaching the apex, the impulse spreads onto the myocardium cells through the Purkinje fibers, beginning the ventricular depolarization. Simultaneously, the atria begin to repolarise;

4. Ventricular depolarisation complete – the ventricles depolarise completely, and their contraction start immediately after the depolarisation;

5. **Ventricular repolarisation** – The ventricles start to repolarise;

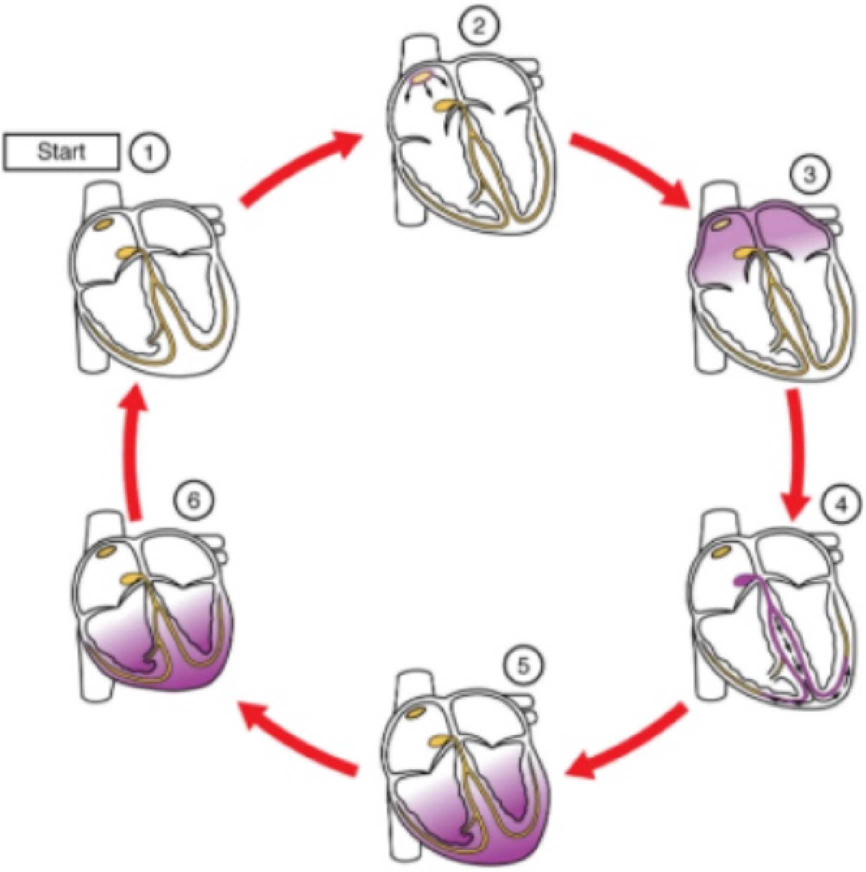


Figure 1.8: The cardiac conduction system. (1) The sinoatrial (SA) node and the conduction system are at rest. (2) The SA node generates the impulse, which spreads across the atria. (3) After reaching the atrioventricular node (AV), there is a delay of approximately 0.1s, allowing the atria to complete pumping blood before the impulse is transmitted to the atrioventricular bundle. (4) The impulse then travels through the atrioventricular bundle and bundle branches to the Purkinje fibers. (5) The impulse spreads to the contractile fibers of the ventricle. (6) Ventricular contraction occurs and the cycle is ready to start again, from [98]

6. **Ventricular repolarisation complete** – the ventricles repolarise completely and the cardiac cycle starts again.

The cardiac cycle is a quick process, taking a total time of approximately 0.22 s (220 ms) and, as it can be perceived, it consists of an electrical current that is generated and conducted through the heart. This electrical activity may be detected and recorded by an electrocardiograph, obtaining a graphic record of the heart activity: the electrocardiogram (ECG) [94]. A typical ECG is composed of three distinct waves: the P wave, the QRS complex and the T wave. The P wave results from the depolarisation of the atria, whereas the QRS is associated with the depolarisation of the ventricles and the simultaneous repolarisation of the atria. Finally, T-wave translates the ventricular repolarisation. The PQ or PR interval is the time between the beginning of the P wave and the QRS complex, during which the atria contracts and begins to relax. After the PQ interval, ventricles begin to depolarize.

The QT interval, which goes from the beginning of the QRS complex until the end of the T wave, represents the time needed for ventricular depolarization and repolarization. Within this interval, the S-T segment represents the time between the moment in which ventricles are completely depolarised until their repolarisation [93; 94]. All these waveforms are registered in a repeating rhythm, the sinus rhythm, originated by the SA node. Furthermore, in some people, generally with slower heart rates, a fourth waveform known as U-wave may appear. Some researchers believe that it represents late stages of ventricular repolarisation or even a post-repolarisation phenomenon [97].

1.5.2 Variability

Although in normal conditions the ECG presents the same major deflections and waves through each heartbeat for all subjects, there's a high degree of variability considering this physiological signal. There are two different kinds of variability in the ECG: intrasubject, which occurs when variations are detected between heartbeats of the same person, and intersubject, consisting of variations between heartbeats of different people [92]. According to J. R. Pinto et al. [36], these types of variability can be associated with different aspects being the most relevant:

1. **Heart Geometry** - The geometry of the heart depends on its size, shape and positioning. All these features highly influence the electrical current path during the heartbeat, the heart and muscle cells to be recruited and depolarised, and the average amount of time of a single heartbeat [99; 100];

2. **Physical Exercise and Meditation** - Athletes with high levels of physical training generally have larger hearts and thicker myocardium, indicating how physical exercise can influence the general heart and ECG characteristics. In addition, both exercise and meditation affect the heart rate, which translates into variations especially visible on the interval between the QRS complex and the T wave [101];

3. **Individual Characteristics** - Overall features like age, weight, height and pregnancy can affect the heart orientation and position. These shifts change the orientation of the electrical current conduction vectors, which alters the ECG waveforms [102];

4. **Cardiac Conditions** - Heart diseases or other medical conditions are also visible in the ECG waveform, since they may interfere in the general electrical conduction dynamics [103];

5. **Position and Shape of the Organ** - The body positioning like standing or laying down lead to significant variations in the position and shape pf the organ. Altering the position of the heart in the thorax will also change its reference position with the electrode placement. In this way, different body positions result on different ECG signals collected [102];

6. **Emotions and Fatigue** - Physiological and psychological states highly influence the autonomous nervous system, which translates into variations in heart rate considering different states of fatigue or emotion, from excitement to calmness [29];

7. **Electrode characteristics and placement** - The ECG acquisition methods and general specifics like the type, size and number of electrodes, as well as their positions on the chest and limbs, have an influence on the quality of the signal collected. The mispositioning of electrodes is also a source of variability, since it can change the general overview and perspective of the electrocardiographic signal [100; 102].

All the factors presented result in visible differences in the ECG data collected. Regarding the first three mentioned (Heart Geometry, Physical Exercise and Meditation and Individual Characteristics), they highly contribute to inter-subject variability, which is potentially positive to the development of biometric systems based on the ECG. However, concerning emotion recognition, subject-dependent settings are one of the most difficult problems that affect the already developed systems, since this can result in different ECG signals and patterns corresponding to the same emotion felt by different subjects. On the other hand, as mentioned, emotions are one of the factors that can also produce variability. However, this variability is mainly intra-subject and depends on the emotional state, which allows for the detection of different emotions through time. The nerve-endings of the Autonomic Nervous System (ANS) within the cardiac muscle have a major effect on the cardiac response, namely on the heart rate. The sympathetic system has fibers that run along the atria and the ventricles. When activated, they stimulate the myocardium to increase the heart rate. On the other hand, this rate and general cardiac workload are reduced by the parasympathetic system [103]. Concerning a stressful state, the sympathetic system overtakes the parasympathetic and different effects occur like the increase of the conduction rate, the dilation of coronary blood vessels, and the increase in perceptiveness to internal and external stimuli [104].

All these effects result in specific patterns and variations that can be detected in the ECG, thus, identifying emotional changes. Nevertheless, the physiology and anatomy behind these biological processes also differ considering the subject, translating into distinct ECG patterns for the same emotion depending on the subject, which can be one reason for the subject-dependency problems found in a large scale of emotion recognition algorithms [103]. Furthermore, other factors may negatively influence emotional detection. For example, body posture during data acquisition may influence HR measurements, consisting of one of the most common features to detect arousal. Thus, all these variability aspects have to be considered since they can either ease or hamper the task at hands.

1.6 Conclusion

The electrocardiogram is a cyclic signal that represents the electrical activity of the heart. As mentioned earlier, because the heart is directly connected to the ANS (parasympathetic and sympathetic nervous systems), it is possible to correlate emotional states with changes in the ECG signal. Although the electrocardiogram is prone to high variability, which may have its origin in various factors and lead to false emotional interpretations, techniques should be applied to reduce these problems, especially regarding subject dependence. In terms of acquisition methods, traditional emotion recognition experiments use simpler and less complex ECG measurement techniques than those used in the field of medical diagnostics. Wireless sensors are widely used, as well as a reduced number of electrodes, allowing a more comfortable and relaxed setup for the subjects, who can move more freely. After presenting overview of the fundamentals concerning both the electrocardiographic signal and the emotion recognition theory, some of the most commonly used approaches and techniques in the literature for Emotion Recognition using ECG and EEG are presented in chapter 2.

Chapter 2

Human Emotion Detection: Related Work

Contents

2.1 Introduction **29**

2.2 Classification of Heart Diseases Based On ECG Signals **29**

2.3 ECG and emotion recognition **31**

2.4 Emotion detection methods **32**

2.4.1 ECG-based Emotion detection methods 32

2.4.2 EEG-based Emotion detection methods 34

2.4.3 ECG-Based Driver emotion detection 35

2.4.4 Data augmentation methods for ECG 37

2.5 Conclusion **39**

2.1 Introduction

For more than a century, since the appearance of the first electrocardiogram, this electrical signal has been developed and constantly improved to obtain more reliable results and measurements. Nowadays, the main application of ECG remains in the field of medical diagnosis. This chapter addresses the a large number of related works (Section 2.2).

On the other hand, Emotion Recognition to human-computer interaction first gained attention in the 1980s. Although most research in this area uses facial expressions [105], emotions are known to strongly influence the Autonomous Nervous System (ANS) activity, which is responsible for regulating a variety of body parameters. In this way, physiological signals began to be considered as possible indicators of emotional fluctuations. Thus, the use of ECG for emotion detection is still quite recent compared to other modalities such as face or voice. However, as a physiological signal, ECG patterns can translate changes in emotion, mainly through heart rate and heart rate variability [38].

At present, emotion recognition methods can be divided into two categories: emotion recognition based on human behavior pattern and emotion recognition based on human physiological signal ECG ,EEG,... In this chapter, we are going to present the existing similar works for Emotion recognition based on ECG (Section 2.4.1) and EEG (Section 2.4.2). In section 2.4.3 an overview of related work in Driver's emotion recognition based on ECG signals is presented. However, the success of ECG analysis based on machine learnin and deep learning relies on rich annotated dataset. Besides, this chapter explores the existing methods of ECG data augmentation to tackle the challenge of limited amount of data (Section 2.4.4).

2.2 Classification of Heart Diseases Based On ECG Signals

The literature of ECG machine learning presented many works. In [106], Naseer et al., proposed a system of automatically differentiating between normal and abnormal heartbeats of patients using electrocardiography (ECG). They studied the components of the ECG signals in order to extract the following features: different time intervals of P-wave, QRS complex and T-wave for classification. The accuracy obtained was above of 80% on average. The results of this proposed system will help cardiologist to efficiently and automatically detect heart diseases based on machine learning. They concluded that linear discriminant analysis will be preferred for real-time application, however, more improvement of accuracy can further be obtained.

Gautam et al., introduced in [107] the concept of pattern recognition of ECG, this study is based on the classification of data patterns and characterizing them into classes of predefined sets. All important information about the activity of the heart offered in the ECG signal generated waveform. In this study, the authors used the ECG signal feature extraction parameters which are spectral entropy, Poincare plot and Lyapunov. To identify the abnormalities of heart disease, the classifier is an artificial neural network, performing data collected from MIT-BIH database. The result obtained using back propagation neural system accomplishes the separation in the normal and abnormal heart rate with a classification accuracy of 93.3%.

Udhaya et al., introduced in [108] the automatic beat classification, which is a significant method to support clinical specialists to categorize arrhythmia signals in ECG recording. The authors constructed a novel automatic classification system to analyze the ECG signal as a decision-making purpose. This method is composed of three essential steps: the first one is a denoising, in which a discrete wavelet transforms (DWT) is applied to delete the noise from the detected signal. After that, the feature extraction is executed followed finally by the classification approach, which is the neighborhood rough set (NRSC). NRSC classifies the ECG signal into normal and four abnormal heartbeats. The proposed algorithm classifies ECG with an overall accuracy of 99.32%

In [109], Rao et al. proposed an improved neural network classification system to identify effectively the heart disease. To obtain discriminative features of the ECG signals (correspond to various cardiac minor and major conditions), the authors presented a signal processing method and feature extraction techniques. The extracted features are classified using neural network classification techniques. This work proposed two efficient approaches for ECG classification, the first one uses transform techniques for feature extraction, where the second one uses a multiwavelet technique to find the smallest set of features to maximize the classification accuracy. The effectiveness of the proposed algorithm proved in terms of error percentage, which is lower than the compared algorithms. Authors of [110] applied the support vector machine (SVM) to detect cardiac diseases, in this proposal the disease is modeled by the time domain features of ECG signal, which is extracted using a software called BIOPAC AcqKnowledge. The extracted features used in this work are heart rate, QRS complex, PR inter-val, ST segment elevation, ST interval of ECG signal. These parameters extracted from the ECG help to detect different heart disease such as atrial fibrillation, sinus tachycardia, myocardial infarction and apnea. The results produced a classification rate of overall accuracy of 84.6%.

Masetic et al., proposed in [111], an automated heartbeat classification system to detect a congestive heart failure. This study consists of two phases namely feature ex-traction using autoregressive (AR) Burg method and classification phase. In the classification phase, the authors applied five different classifiers such as C4.5 Decision Tree, Support Vector Machine, k-Nearest Neighbor, Artificial Neural Net-works and Random Forest classifier. The experimental results showed that among the classifiers used the Random Forest method gives a good classification accuracy.

2.3 ECG and emotion recognition

In cardiology, heart rate (HR) is defined as the number of beats (or systolic contractions) per minute, the ECG records this cardiac electric activity responsible for the myocardial contraction, by the number of ventricular electric QRS waves see figure 2.1). HR is measured by counting the number of R waves registered in a minute. The time interval between two electrical R waves is called the R-R interval, and relates to the clinical interbeat interval. The physiological R-R interval is not constant, and varies depending on several factors, such as breathing, hormonal stimulation, or emotion (e.g. stress).

More specifically, the researchers mentioned that there are innervations of the ANS within the heart four chambers (i.e. two atria and two ventricles). These innervations play a major role on the cardiac output. Indeed, they have an effect on both the physiological pacemaker (sinus node), which controls the HR (chronotropic effect), and on the conduction of the electric signal running from this node through the rest of the heart (dromotropic effect). The sympathetic nervous system (SNS) has positive chronotropic and dromotropic effects (increases HR and speed of electric signal conduction), whereas the parasympathetic nervous system (PNS) has negative effect on both functions (slows pace and delays electric conduction) [112]. According to the intensity of a specific emotion, the sympathetic system is stimulated to prepare the body against a strong activity (fight-or-flight response). The parasympathetic system dominates in calm, resting activities (rest-and-digest response). Moreover, the HRV is very linked to the ANS and it is responsible for keeping the balance between the two systems: parasympathetic branches which are defined as the rest and digest response and sympathetic branches which are defined as the “fight and flight” responses [113].

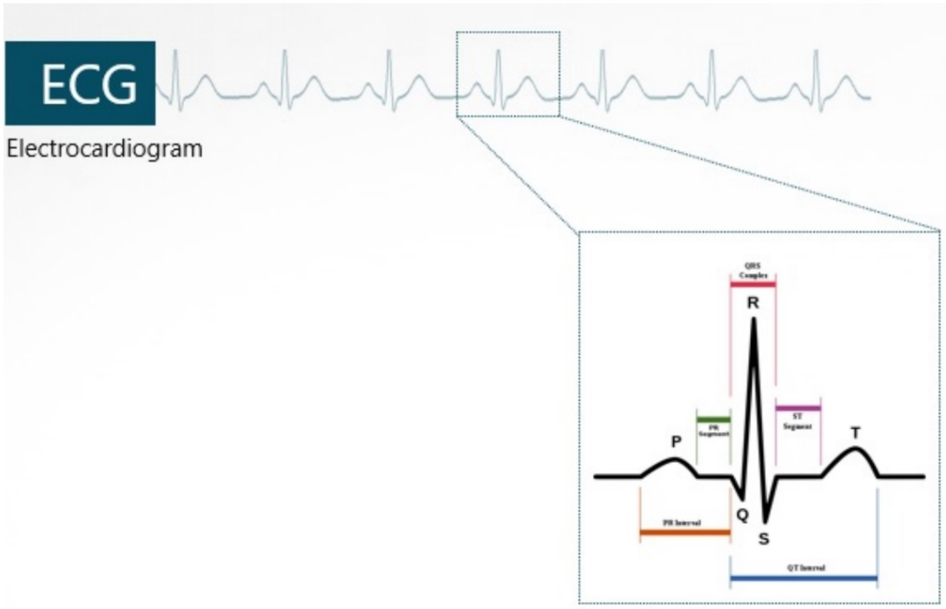


Figure 2.1: ECG signal

2.4 Emotion detection methods

Up to now, several techniques have been conceived to recognize emotion in the basis of ECG signal and EEG signal. This section presents and discusses these works. As well as ECG data augmentation and ECG data classification using machine learning.

2.4.1 ECG-based Emotion detection methods

In the past decades, ECG-based emotion recognition was considered as one of the most important branches of emotion recognition [114], [115], [33], In order to identify emotions in response to music, Kim et al. in [116] used different physiological signals namely, ECG, EMG, respiration, and skin conductivity. 110 features are calculated from various analysis domains, including HRV/breathing rate variability (BRV), geometric analysis, entropy, multiscale entropy, time/frequency, subband spectra, etc. These calculated features have been used to detect the best emotion-related features then to relate them to emotional states using the backward feature selection method. Furthermore, a novel scheme of emotion-specific multilevel dichotomous classification (EMDC) was developed to ameliorate the accuracy of four musical emotions in terms of arousal dimension such as positive/low arousal, positive/high arousal, negative/low arousal and negative/high arousal. As a classifier, the authors used an extended linear discriminant analysis (pLDA). After a set of experiments, the authors achieved a recognition accuracy of 70% for subject-independent classification against 95% for subject-dependent classification.

Nardelli et al. proposed a novel approach to identify emotions automatically, these emotions being evoked by emotional sounds. The HRV features extracted from ECG signals were used as input for the automatic emotion detection system. In this study, the emotions are expressed on valence with two classes and on arousal with four classes. The results obtained using the quadratic discriminant classifier for arousal and valence achieved a recognition accuracy rates of 84.26% and 84.72%, respectively [117].

Katsigiannis and Ramzan in [15] presented a multi-modal database called DREAMER, this latter consists of various ECG signals to recognize emotions that are evoked by audio-visual stimuli. The authors considered HR and HRV as features to detect the emotion in terms of valence, arousal and dominance using support vector machine (SVM) classifier, where the classification accuracy achieved are 62.37%, 62.37% and 61.57%, respectively. In [20], Correa et al. presented a new dataset named A dataset for Multimodal research of affect, personality traits and mood on Individuals and GrOupS (AMIGOS). AMIGOS is conceived to detect human emotions using neurophysiological signals. This dataset contains different multimodal records of the participants as well as their reaction emotional videos. The first set of videos is classified into one to four quadrants of the valence-arousal (VA) space such as: LVLA, LVHA, HVLA and HVHA (where: L/H: Low/high and V/A: Valence/Arousal). However, the second set of videos contains eight video extracts from movies according to their score in Internet Movie Database (IMDb) Top Rated Movies 3 list. Those selected videos from movies are affective multimedia content and do not demand a prior knowledge of the participants to be understood. In this case, the authors used short and long videos to generate different types of emotions, this is in two social situations: one for individual viewing and the other for viewing with groups of viewers. Using the wearable sensors, they collected participants' physiological signals such as: EEG, ECG and Galvanic Skin Response (GSR). The experimental evaluation showed that there are important correlations between the internal and external effect of valence and arousal which is good to predict the affective state of participants.

Sarkar and Etemad in [114], [118] applied an existing self-supervised deep multi-task learning framework on ECG recordings for emotion detection. In this research activity, the authors used four public datasets namely, SWELL, WESAD, DREAMER and AMIGOS. Compared to a fully-supervised method, the results obtained showed that the proposed method is able to improve classification performance. Specifically, arousal and valence detection is performed on DREAMER dataset, achieving an accuracy of 77.1% and 74.9%, respectively, where the accuracy obtained by WESAD is 95.0% for 4 affective states: amused, meditated, stressed and neutral. Also, AMIGOS has given an accuracy of 78.3% and 79.6% for valence and arousal, respectively. Finally, with SWELL, the achieved accuracy is of 92.6%, 93.8% and 90.2% for arousal, valence, and stress.

L.Granados et al. in [119] used a dataset of physiological signals (ECG, GSR and galvanic skin response) from AMIGOS dataset. They suggested a Convolutional Neural Network (CNN) as an automatic feature extractor of GSR and ECG to detect valence and arousal. The experimental results of this research activity showed a better precision to classify different emotional states, compared with the results obtained by [20], when using classic algorithms of machine learning: Gaussian Naıve Bayes and SVM.

Despite physiological signal analysis being considered as an effective method to recognize the emotion of humans using HRV of ECG signal, the proposed techniques applied a short dataset like AMIGOS, DREAMER, WESAD and SWELL in terms of size of ECG samples and their diversity. This limitation is due to two main reasons; the first one concerns the limited number of participants for measurement due to some difficult conditions (e.g. persons in the cars), and the second reason is that the only person authorized to interpret and annotate each sample is the cardiologist; it is a very hard task.

2.4.2 EEG-based Emotion detection methods

As a physiological signal, the EEG can provide important and complex information about a person's emotional state. In the past decades, EEG-based emotion recognition has received great interest from researchers. In order to detect autism spectrum disorder (ASD) in children, Aslam et al [120] developed an emotion recognition processor, based on an eight-channel EEG signal. This research activity combines a patient special SVM classifier with a hardware-efficient feature extraction engine realized to discriminate the emotions in real-time. The accuracy results in valence and arousal were 63% and 60%, respectively.

Aslam et al [121], proposed a processor for Chronic neurological disorders (CND's) to detects human emotions using eight EEG channels. Using linear SVM (LSVM) as a classifier, they achieved an accuracy of 70.71% on SEED dataset for valence, in addition, the classification accuracy achieved with DEAP dataset is about 72.96% and 73.14% for valence and arousal, respectively. To solve the problem of limited EEG data in order to use deep learning methods to identify emotions from EEG signals, the authors of [122] applied a simple data augmentation method on MAH-

NOBHCI dataset, aiming at generating more EEG training samples. The obtained results showed the effectiveness of the data augmentation method to improve the performance of deep models.

Table 2.1 summaries recent ECG and EEG-based emotion recognition studies proposed in the literature as well as various comparisons classification criteria.

Table 2.1: Recent Emotion Recognition Studies based on ECG and EEG signals.

Ref	Recorded signals	Method	Emotions	Detection precision rate	Detection confusion	Amount of used data
ECG signal						
[114]	SWELL AMIGOS	Self-supervised approach	Arousal Valence	High	Medium	Low
[15]	ECG signals from 23 subjects DREAMER	SVM	Valence Arousal Dominance	62.37% 62.37% 61.57%	Low	Low
[33]	ECG signals from 27 subjects AMIGOS	Least squares SVM LS-SVM	Positive/negative valence high/low arousal 4 types of emotions	82.78% 72.91% 61.52%	Medium	Low
[118]	ECG signals from 40 subjects AMIGOS	Gaussian Naıve Bayes SVM	Valence Arousal	54.5% 55.1%	Medium	Low
EEG signal						
[68]	DEAP database	Spectral and time features, multiple-fusion-layer based ensemble classifier of stacked auto-encoder (MESAE)	Arousal Valence	77.19% 76.17%	Medium	Low
[123]	EEG signals from 57 subjects	K-nearest Neighbour (KNN), Probabilistic Neural Network (PNN)	Sad, disgust, fear, anger, happy and surprise	82.32% only sad emotion is highly perceived	Medium	High
[124]	EEG signals from 12 subjects	Linear Discrimination Analysis (LDA)	Positive and negative	64.73 %	Medium	Low
[125]	EEG signals from 15 subjects SEED dataset	Regularized graph neural network	Neutral, sad, fear, and happy	73.84 %	Medium	High

2.4.3 ECG-Based Driver emotion detection

With the increasing use of many types of electronics such as on-board electronics as well as in-vehicle information systems, the task of evaluating the driver has become an urgent requirement for both industry and government. In a driver-vehicle-road system, the driver has a crucial role of this system and the drivers state is very important to achieve the so-called safe driving. Three seconds of inattention to the driver is likely sufficient for a fatal traffic accident, according to the National Highway Traffic Safety Administration (NHTSA) [126], for this reason, the monitoring of a drivers states during driving is important and beneficial to improve their safety and the security of the road users. We can define the state of a driver on two sides: the psychological state and the physiological state.

The physiological state can be diagnosed using medical knowledge (Clinical medical examination), however, the emotions represent the psychological state of the driver, this state appears subjectively, it is an neural state that appears unexpectedly and subjectively [127]. These physiological signals can be collected periodically without interrupting or obstructing the driver's task performance while driving on the road. The information obtained from physiological signals is very useful metric about both driver's health state and psychological state. This information can help the driver to deal properly with the stress to avoid the risk of fatal traffic accidents. This information can also help road authorities to take the necessary measurements to maintain good traffic.

In the literature, Dhananjay et al. treated in [128] the fatigue analysis by acquisition their physiological signals which is ECG signal motorists. They used Pan Tompkins algorithm to detect the QRS wave complex of ECG signal, specifically the ECG features such as width and amplitude. This research activity is done in time domain, in order to detect the fatigue of drivers to lessening the figure of accidents to a major extent in roads. For the classification, the authors used Neural network using nntoolbox of MATLAB. The obtained results showed that the 03 levels of stress could be detected and the accuracy could achieve 97.4%. This is done by using ten minutes intervals of used data, also the ECG and heart rate showed the highest global correlations with continuous driver stress levels.

Keshan and al. in [129] focused on ECG monitoring which can be obtained from a wearable patches and sensors easily. The authors developed an efficient and robust approach to identify the stress using ECG signal accurately. In this research, the analysis of the stress of an individual were done at three stress levels: low, medium and high considered as unique aspect of this presented work. The accuracy of classification achieved was about 88.24% from the ECG signals alone. The authors proved in this work that the high stress could be easily recognized for a person compared to his or her rest period; the accuracy in this case can be reached at 100% with NaiveBayes.

Healey et al. in [130] suggested to collect and analyze physiological data such as Electromyogram, Electrocardiogram, respiration and skin conductance, which are saved in a continuous way while driving. For the analysis, the authors collected data from 24 drivers with a minimum duration of 50 minutes. The analyzed data is about two driving conditions: in highway and in city, in order to identify 3 stress levels of driver: low, medium and high. They obtained 97% of accuracy, from many drivers with several days of driving. However, this experiment was tackled only two metrics: skin conductivity and heart rate, which could be completed by other metrics like ECG signal. In [131], Bichindaritz et al, focused on electrocardiography (ECG) to detect stress, the ECG could be obtained with minimally invasive wearable patches and with different type of sensors. In their research, the authors have analyzed the personalized individual stress specific for every human being, including 03 levels. To do that, different machine learning algorithms were applied using only one physiological ECG signal, where the accuracy achieved up to 100%. Regardless of the ECG classification obtained by the aforementioned studies using different machine learning algorithms, these proposed approaches require that the classifier parameters are introduced by the user manually and in a random manner. This manual choice of parameters influences negatively on the classification rate.

In [62], Nita et al. used an ERF classification approach to detect and diagnose driver's stress level while driving on the road. This approach was tested on MIT-BIH physioNet dataset [132]. The results proved that the ERF is more efficient than SVM in terms of recognition accuracy.

2.4.4 Data augmentation methods for ECG

Despite the importance of class balancing through data augmentation to the quality of classification output, data augmentation in ECG has been quite limited, as was also pointed out by [133]. Pan et al. [133] applied time series alteration methods to augment ECG data and compared the efficacy of four such methods—window slicing (Cui et al. [134]), permutation (Um et al. [135]), concatenation and resampling (Cao et al. [136]), and window warping (Le Guennec et al. [137]). Pan et al. [133] evaluated these methods using LSTM RNN and observed good results, but the data augmentation and classification was done on the entire ECG time series, and thus their methods are not applicable to our work, which needs to augment data at the level of individual ECG beats. There are a few known works on ECG data augmentation and classification at the beat level. Jun et al. [138] rendered 1D ECG signals to 2D images and used image cropping and masking for use with CNN. This method is not applicable in our work, which alters ECG as a 1D signal. He et al. [139] divided each class into five subsets and duplicated classes to match the number of samples in the most dominant class. The classification accuracy measured on nine classes in the 12-lead China Physiological Signal Challenge (CPSC) dataset was 80.6%. While this method may be able to balance the dataset sizes across classes, the trained classifier may not be robust to classes containing data duplicated from the ECG of a small number of patients.

Yao et al. [140] compressed or stretched the ECG signal along the timeline and removed parts of the segments (or beats). The classification accuracy measured on nine classes in the 12-lead

CPSC dataset was 81.2%. This method is problematic because the ECG signal is a time series, and, therefore, its diagnostic classification is sensitive to alteration along the timeline. Note that our work uses an amplitude scaling method, which limits the signal alterations only to the amplitude, not the timeline. Acharya et al. [141] altered samples by amplitude scaling, too. They varied the standard deviation and the mean of Z-scores from the original normalized ECG segments (of length 260 samples). The specifcs of the implementing the scaling, however, is not stated; while it appears randomly selected scale factor was applied to each segment. The classification accuracy measured on five classes in one lead ECG was 94.03% with noise removal. While these results are quite good, the setup is limited compared with than our work where nine classes in 12 lead ECG are evaluated.

For instance, Cao et al. in [136] developed a new data augmentation method to improve deep neural networks (DNN), conceived to detect atrial fibrillation (AF) from ECG recordings. The principle of this method is to concatenate the original ECG episode to the duplicated one, this is based on some characteristic points. After that, this concatenated signal is resampled at random. This algorithm makes a balance in the number of samples between the different classes of the dataset and also increases the variety of the dataset. The results of this work ameliorate the performance of DNN for AF detection.

Based on generative adversarial networks (GANs), Haradal et al. [142] proposed the use of a synthetic generation method for time series as well as a related application to increase data for biosignal classification. In order to generate data of time-series, the authors developed every neural network in the GANs using Long short Short-term Term Memories (LSTM) units this is done for its hidden layers based on a Recurrent Neural Network (RNN). The experiment results showed the capability of this method to generate synthetic biosignals using the EEG and ECG datasets and to analyze with better precision the studied system.

The authors of [135] used the permutation method combined with Window Slicing (WS) for the first time as a data augmentation method to monitor Parkinson's disease using wearable sensor data. This proposed method and CNN are applied in order to determine the motor state of Parkinson's Disease patients. The results obtained reached an accuracy of 86.88%. Also, Nonaka al.[143] applied a suitable method of data augmentation in a DNN model in order to classify atrial fibrillation. Using ECG augmentation with a single lead ECG data, the results showed that the proposed method improve classification of atrial fibrillation with an accuracy of 84.27%.

2.5 Conclusion

In this chapter, we presented relation between ECG and emotion recognition then we presented ECG-based Emotion detection methods as similar as for EEG signal. After that, we presented a number of related works to detect emotion for driver. Lastly, we presented different existing methods of data augmentation for ECG signal.

In the next chapters, we present our contributions, there are novel emotion recognition algorithms based on ECG signals, focusing on two major manageable challenges in which, albeit correlated could be treated as separate research problems namely, a good classifier problem (1) and the problem of limited size and imbalanced number of ECG samples in datasets (2).

The first problem investigates to enhance the random forest method by suggesting a new simulated annealing (SA) algorithm to find the optimal number of trees where the accuracy of classifying the ECG signal is tackled as an objective function (see chapter 3), this method was applied to detect and diagnostic stress level of driver(see chapter 4). The second problem addresses a novel data augmentation strategy aiming to increase the diversity of the samples and balance the number of samples among the classes (see chapter 5).

Chapter 3

An enhanced random forest for cardiac diseases identification

Contents

3.1 Introduction **41**

3.2 Motivation: ECG for heart disease diagnosis **41**

3.3 Basic concepts **42**

3.3.1 Random forest (RF) classifier 43

3.3.2 Simulated annealing (SA) method 43

3.4 An enhanced random forest for ECG classification: our proposal **44**

3.4.1 Phase 1: ECG Data Aggregation 44

3.4.2 Phase 2: Preprocessing phase 44

3.4.3 Phase 3: Features Extraction (BD of features) 45

3.4.4 Phase 4.1: Tree classifier building and decision aggregation 46

3.4.5 Phase 4.2: Discovering the optimal RF number of trees 46

3.5 Experimental results **48**

3.5.1 Data set description and experiments environment 48

3.5.2 Experiments and discussion results 49

3.6 Conclusion **51**

3.1 Introduction

Cardiac diseases are one of the foremost reasons of mortality in the worldwide. To cope with this issue, cardiology doctors insist on the early detection of cardiac diseases often with the use of an electrocardiogram (ECG) signal, providing timely and appropriate treatment for heart patients. In the literature, there are many efficient classification approaches like random forest method, conceived for ECG signal analysis to detect cardiac diseases. However, the execution of random forest requests introducing manually the number of trees as a parameter user, which is considered as a major drawback of this method, since often the user did not find the optimal tree value. In this chapter, we propose to enhance the random forest method by suggesting a new simulated annealing (SA) algorithm to find the optimal number of trees where the accuracy of classifying the ECG signal is tackled as an objective function.

3.2 Motivation: ECG for heart disease diagnosis

Recent statistics showed that heart disease is the number one cause of death for both men and women in the world, according to center for disease control and prevention (CDC), for example heart disease killed only on USA 610.000 people alone in 2014 [144]. Nowadays, the heart disease has become one of the most common disease, which has an important affect humans health worldwide. Due to the stress and some news habitudes of peoples (People of all age groups) in their daily life, heart attacks or myocardial infarction (MI) are widely known. The Electrocardiogram (ECG) signal can discover the disorders occurring in the heart, which present a heart disease or heart attack. So, to prevent unwanted heart attack, early detection and timely treatment of arrhythmia is necessary. ECG is a tool that is used to access the electrical recording and muscular function of the heart, also it is used in the diagnosis of heart related diseases. The introduction of computerized classification for ECG can help to reduce health care costs, moreover, ECG analysis is considered as a key element regarding the evaluation of human health status and its prevent. The use of the ECG allows verifying if the heart has a normal activity or not, the patient ECG representing one heart-beat, should be compared to a normal and representative ECG pattern (see figure 3.1). Depending on the type of pattern that the ECG trace may have, we can conclude the type of heart disease the patient may have. A normal ECG pattern has many part (i.e. spikes and dips) called waves and segments with known default measurements, this leads to detect any abnormal measurement, which may be an abnormal heart activity, referring later to a potential heart

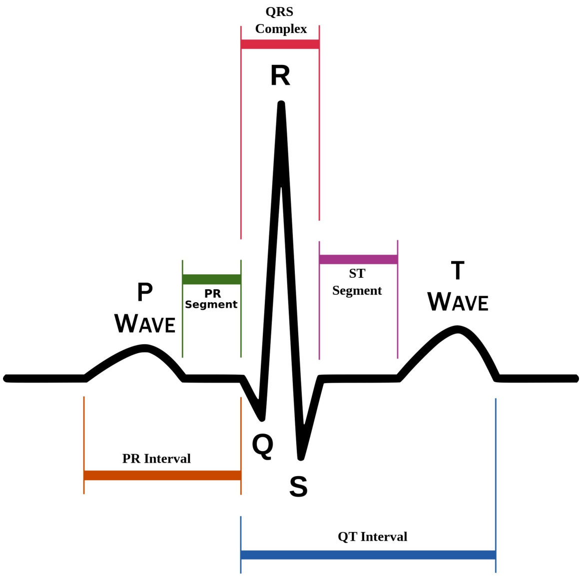


Figure 3.1: An ECG one heartbeat normal pattern

disease. The first electrical positive signal in the normal ECG is P wave, it has positive polarity and its duration is less than 120 milliseconds. Additionally, a QRS complex portion is the largest part of the ECG signal (the result of contraction of the ventricles), where:

- **Q wave** Q wave is the first negative or downward deflection,
- **R wave** is the first positive deflection,
- **S wave** is the negative deflection,
- **T wave** represents the relaxation (repolarization) of the ventricles [145], (see figure 3.1).

The rate of the normal sinus rhythm is varying from 60 to 100 beats/min [146], and when there are abnormalities occur in ECG, automatically the standard features change their values; for example ST depression, T wave changes and so on [147]. In the state of the art of ECG machine learning, several efficient classification approaches were proposed such as support vector machine (SVM) [110; 148], Neural Network [107; 109], Kmeans [149], and random forest method (RF) [150–152].

3.3 Basic concepts

This subsection presents random forest classifier and simulated annealing optimization method considered in our proposal.

3.3.1 Random forest (RF) classifier

Proposed by Breiman in 2001 [153], considered as a decision trees method, RF is generally more efficient than simple decision trees approach offering more accurate results [154]. RF is based on trees structure constructed following bagging and bootstrap methods. The data set is subdivided into several parts by the bootstrap and then a decision tree is learned from each part. A new example is tested by all trees built and its class is the majority class. Bootstrap method or replacement sampling is a powerful statistical method whose purpose is to estimate a quantity from a sample of data. The model trained on a set of N randomly selected examples of the set of examples, where examples can be chosen more than once or cannot be chosen at all. To obtain an improved precision of the model, this operation can be repeated several times [155]. Bagging or Bootstrap Aggregation is a simple and very powerful ensemble method. It is considered as an application of the Bootstrap procedure to a high-variance machine learning algorithm in order to reduce the variance (decision trees) [155]. The set D of examples is subdivided into a subset by Bagging. From each subset, an M_i model is learned using the Bootstrap method. This set of learned models forms a composed model M. In order to classify a new example and to obtain a class, this example is exposed to each model M_i . Each generated decision is considered as a vote. The most voted class is taken as a final decision.

3.3.2 Simulated annealing (SA) method

This optimization method is inspired by annealing in metallurgy, it is a technique based on involving heating and controlled cooling of a material, and this to increase the size of its crystals and reduce their defects [156]. Basically, SA is composed of two stochastic steps: the first one is to generate solutions and the second one is to accept best solutions [157]. SA starts with an initial solution chosen at random, then a set of iterations is run. For each iteration, a neighbor solution is generated and selected, if it is better than the current one, therefore the best found solution moves to this newer one. SA can also choose a worst neighbor solution but with a low probability aiming at moving to a better solution in next iterations. This iteration is run until satisfying a stopping criterion.

3.4 An enhanced random forest for ECG classification: our proposal

In this section, the novel automatic classification system for ECG beats analysis is illustrated. Our proposal is based on the introduction of a simulated annealing (SA) algorithm in the aim at finding the optimal number of random forest trees leading to a high accuracy of ECG classification. Considered as one of the most efficient metaheuristics, SA performs with a reasonable linear complexity [157]. The proposed system involves four main steps namely data collecting of ECG signal, pretreatment and denoising this data, feature extraction and classifying this signal using the enhanced random forest approach. This proposal uses different features contained of the ECG signal to classify the different beats within this signal, which will help on diagnosing for potential diseases. We use also a discrete wavelet transform to extract the morphological features namely P, T waves and QRS. The feature extraction and classification algorithm are implemented and evaluated based on ECG European ST-T data set and MIH-BIH from physionet, in addition to both US databases called Arrhythmia Data Set and Heart Disease Data Set. The complete methodology adopted in this work is presented in figure 3.2 composed of four phases.

3.4.1 Phase 1: ECG Data Aggregation

The first step of the proposed system is the ECG data gathering aiming at collecting the most representative information of the heart state and the widely used in the monitoring of the heart disease. To achieve this, two kinds of data aggregation are suggested; either an ECG data which is directly captured from patients or a representative data reached from universal ECG databases. For instance, we acquired in this study the ECG from two sources; the European ECG databases and MIH-BIH arrhythmia of PhysioNet [158] (available on [159]) and from both US Arrhythmia Data Set and Heart Disease Data Set from UCI ECG data set available on [160; 161] respectively.

3.4.2 Phase 2: Preprocessing phase

Initially, detected from the patient as suggested in phase 1, the ECG signal is exposed to several types of noises during the generation and aggregation steps. It is due to the nature of human body and the way of how ECG is measured. Several types of noise from various sources are enumerated like muscular activities, powerline, skin stretching and electrode motion, movement of heart due to respiration, etc. Consequently, ECG denoising is aimed to eliminate or at least to

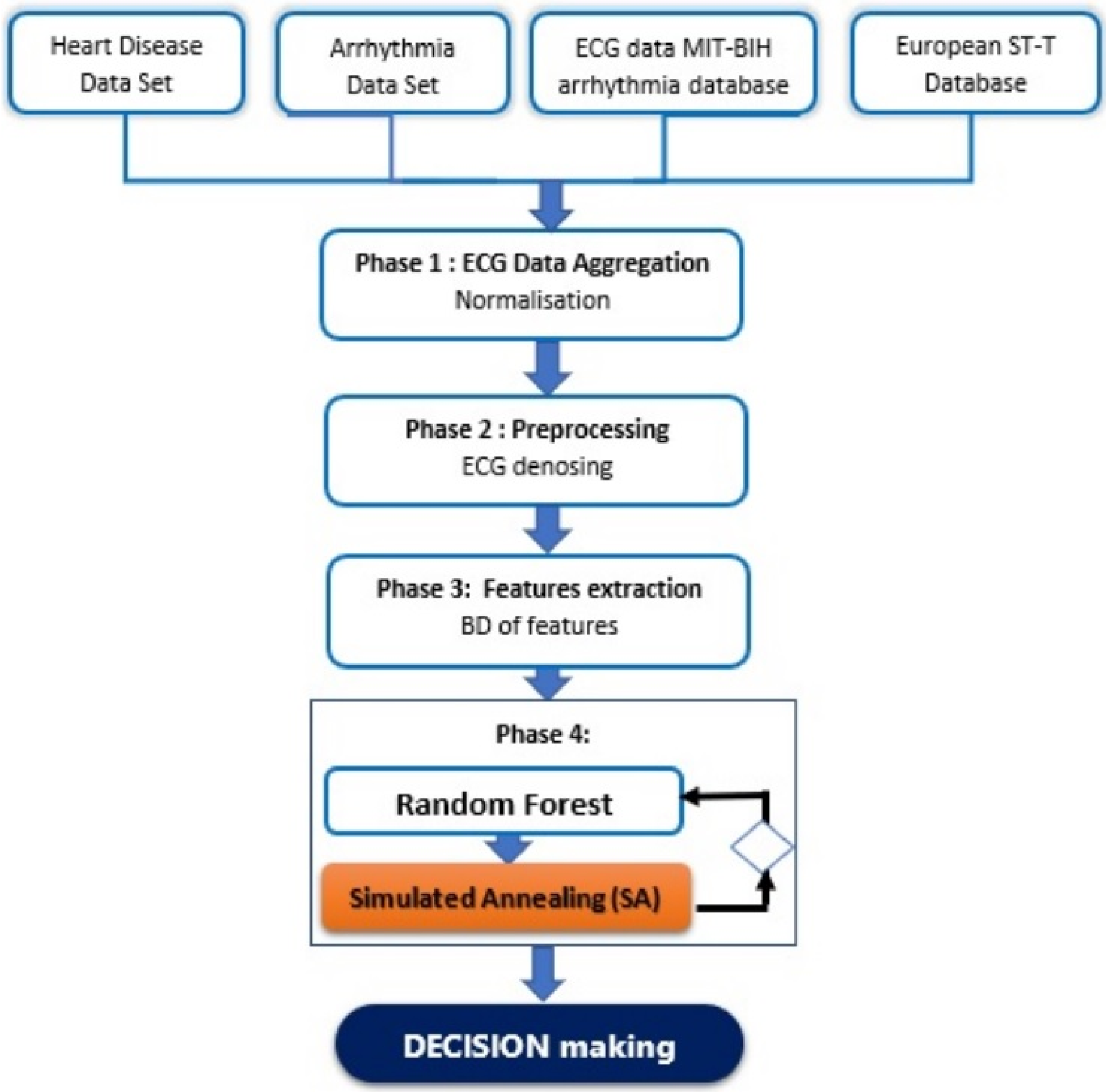


Figure 3.2: Flowchart of the proposed ECG enhanced random forest classifier

minimize the unwanted signals from an ECG record, without hampering the clinical information contained within the signal itself. In this study, we consider that this phase is executed as a black box by technical denoising approaches proposed in prior signal processing works like [162], as shown in figure 3.3

3.4.3 Phase 3: Features Extraction (BD of features)

R-peak and QRS complex are the most crucial feature in the cardiac cycle, and their location is used to obtain and understand the rhythm statement of the ECG. Many research works are done by scientist in the aim to improve the rate R-peak and other features (T wave, P wave, QRS complex). Those features are fundamental in the interpretation of the ECG, and they are the main

Figure 3.3: The preprocessing and features extraction step

characteristic used in the beat classification system. After reaching the clean data and its essential features, the feature selection phase is applied. In this phase, we select and combine features in order to get desired and significant features to our scheme. The data resulted will form the new features database, which is divided into two parts; the first one is used as a training data named D in a space X of dimension M, while the second one is considered to validate our classification. Here, we generate training data sampling where the bagging method is run to generate k subsets of training data D_1,D_2,...,D_k by sampling data D at random with replacement. The data used to train the model is a portion from the whole database; often 70% of data is used.

3.4.4 Phase 4.1: Tree classifier building and decision aggregation

For each training data set $D_i (1 \leq i \leq K)$, we apply a decision tree algorithm to grow a tree from a random selected sub samples (based on bootstrap method) [163; 164]. At each node, the algorithm randomly samples a subspace X_i of F features where (F << M), and computes all splits in subspace X_i, after that it selects the best split as the splitting feature to create a child node. This operation is repeated until the satisfaction of a stopping criterion, finally, a tree $h_i(D_i, X_i)$ built by a training data D_i under subspace X_i is then obtained; the next step is the decision aggregation: it is a collection of the k trees $h_i(D_1,X_1)$, $h_2(D_2,X_2)$,...,$h_k(D_k,X_k)$ generated to compose the random forest, then the majority vote of these trees is processed for making a classification decision set.

3.4.5 Phase 4.2: Discovering the optimal RF number of trees

Since the most important user parameter of random forests is k the number of classification trees [163], we suggest in this phase to reinforce this classifier by a simulated annealing (SA) metaheuristic technique aiming at finding the optimal number of trees, leading to a higher classification rate. SA discovers an optimal number of trees and checks the accuracy of RF classifier obtained with this number of trees, this process is repeated until the absolute optimal number of

Step 1: Initialize the system configuration:
 Initial number of trees at random $Nb.trees(0)$
 $Nb.trees = Nb.trees(0)$ *// Save the optimal tree number in* $Nb.trees$
 Evaluate function $RF(Nb.trees)$
 Calculate $Best\_Acc = Accuracy(Nb.trees)$ *// Calculate accuracy of ECG classification according to the initial solution*
 Initialize T with a large value *// T: Temperature*

Step 2: ***Repeat:***
 Generate a random neighboring solution $Nb.trees = Nb.trees + \Delta x$
 Evaluate function $RF(Nb.trees)$ *//Evaluate neighbor solution*
 Calculate $Current\_Acc = Accuracy(Nb.trees)$ *// Calculate accuracy of ECG classification according to the new neighbor solution*
 If (Current Acc > Best Acc) *// this a new Hight Accuracy*
 Keep the 'new neighbor' to 'best found'
 else
 accept the new worst solution with probability P= $e^{\frac{-\Delta E}{T}}$
 end if
 $T = T - \Delta T$
 until (a stopping criterion)
Step 3: **return** Nb_trees and Best_Acc

Figure 3.4: The pseudo-code of Enhanced Random forest

trees is found, as presented in figure 3.3.

The proposed SA starts with an initialization (step 1) in which an initial tree number called NB.$trees$(0) is chosen at random. After that, the random forest as an objective function is called to evaluate the ECG classification accuracy according to the chosen NB.$trees$(0). As SA algorithm suggests a temperature (T) parameter to conduct the SA iteration evolution, we propose to initialize T with a very high integer value. In next step 2, the general SA loop is performed. Here, a neighborhood generation is executed and the best neighbor solution (i.e. best new number of tree) is selected to be evaluated. This new number of trees is considered as the new best one if it is better than the old number of trees, otherwise it is accepted with a probability P as depicted in figure 3.4. Note that this worst tree number can lead in next SA iterations to a global optimum, then SA can surmount the optima local problem. After that, temperature value is slightly decreased. This second step is repeated until the best solution is reached (i.e. SA can not improve the number of trees; it is the stagnation stage).

3.5 Experimental results

We remind that our proposal tries to find the optimal number of trees considered as a user parameter of RF, affecting the performance of Random Forest classifier, where the classification accuracy of ECG signal is the objective function, which should be increased. In these experiments, the enhanced Random forest is implemented on R studio framework [165]. As the first step of the proposed enhanced random forest classifier is the ECG database collection, and in order to validate this scheme, we have chosen to use the online European ST-T Database and MIH-BIH Database from Physionet website as well as both US databases called Heart Disease Data Set, and Arrhythmia Data Set from UCI, respectively.

3.5.1 Data set description and experiments environment

In this subsection, we describe the used ECG databases to evaluate the proposed method as shown in Table 3.1.

Table 3.1: ECG databases used in the experiments .

Database	Number of records
Heart Disease Data Set	303 records
Arrhythmia Data Set	452 records
European ST-T	090 records
MIT-BIH Arrhythmia	048 records

Heart Disease Data Set [160]: this database contains 76 attributes, specifically, the Cleveland database. The objective is referring to the presence of heart disease in the patient. It is an integer value varying from 0 to 4, where 0 means that is no presence of heart disease, however the values 1 to 4 indicate the presence of different heart disease. Arrhythmia Data Set [161]: this database contains 279 attributes, where its objective is to distinguish between the presence and absence of cardiac arrhythmia and to classify these cases in one of the 16 groups or classes. Class 01 refers to 'normal' ECG, and for the classes 02 to 15 it corresponds to different classes of arrhythmia, however class 16 refers to the rest of unclassified ones.

European ST-T Database [158]: the designer of this database is the European Society of Cardiology, this database consists of 90 annotated excerpts of ambulatory ECG recordings from 79 subjects. MIT-BIH Arrhythmia Database [132]: this collection of 48 fully annotated half-hour two-lead ECGs, and the most used database in the seen literature, the considered beats refer to the following classes: normal sinus rhythm (N), ventricular premature beat (V), atrial premature beat

(A), right bundle branch block (RB), left bundle branch block (LB), and paced beat (/).

Our experiments are based on four databases from two sources: UCI machine learning repository and physionet, two database for each one. The first source uses two databases: Arrhythmia Data Set and Heart Disease Data Set. For the second source, we use MIH-BIH Arrhythmia Database and European ST-T Database. For these two databases, we have used the large library of software PhysioToolkit, which is introduced for physiologic signal processing and analysis [166]. In order to apply the suggested tools, we have used the Wave Form Data Base (WFDB) Software Packages [167], and we have performed the interoperable command-line tools for signal processing, which are originally Unix compatible. Electrocardiographic (ECG) signals may be corrupted and affected by various kinds of noise, such as artifacts from electrodes, skin noisiness, noise from power line interferences and muscles signals. An elimination of all types of noises from ECG signal is increasing the system efficiency. In this study, ECG signals denoising and feature extraction are performed using discrete wavelet transform. In fact, Wavelet transform (WT) offers essential information of signal in both time and frequency domain. WT is a well suitable tool for analyzing biological non-stationary signals such as ECG, EMG [162]. Moreover, we use the Annotation C++ library of Mr. Chesnokov Yuriy to preprocess the ECG signal and to extract necessary features. The used C++ library is based on waveletanalysis and console application for extraction of vital intervals and waves from ECG data (P, T, QRS, PQ, QT, RR, RRn), ectopic beats and noise detection [168]. We considered this step as a black box tool (figure 3.3), the entry is Physionet (.dat) compatible file, and the result is text file containing the features detected and the position according to the original file. It is worth noting that both Heart Disease Data Set and Arrhythmia Data Set are tested separately, since they contain many information about patients, where only features extracted from ECG signals are considered, however, the other two databases from physionet website are merged to form a single one with the same features of the ECG signals

3.5.2 Experiments and discussion results

The classification results of our proposal are depicted in figures 3.5, 3.6 and 3.7. We make a comparative study of the Enhanced RF (ERF) and RF where the number of trees generated automatically in ERF is for 500 iterations, however, RF trees are generated manually by the user for 10 iterations. We mention clearly the impact of the found number of ERF trees on the classification accuracy to predict beats of ECG signals.

The found results depicted in figures 3.5, 3.6 and 3.7, showed that the enhanced random

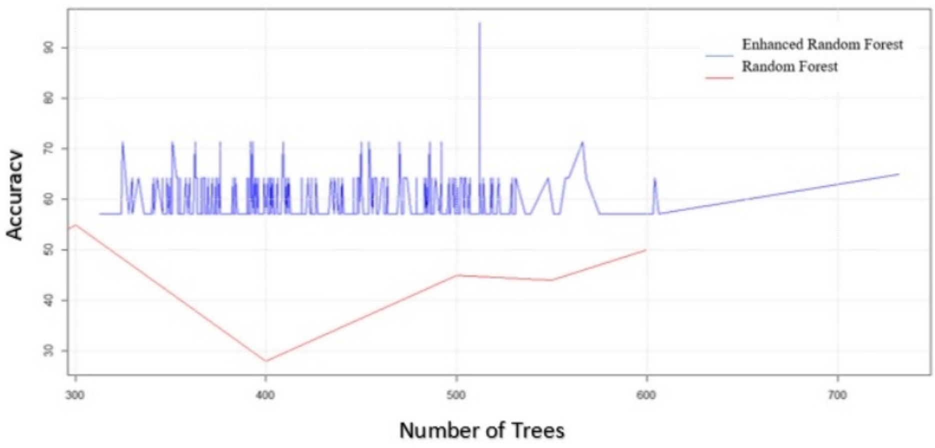


Figure 3.5: Comparison of Prediction Accuracy with Arrhythmia Data Set

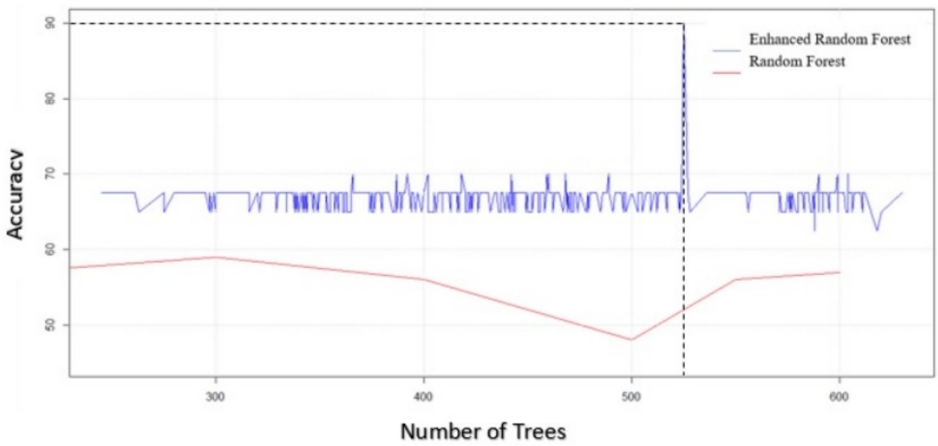


Figure 3.6: Comparison of Prediction Accuracy with Heart Disease Data Set

forest outperforms the original random forest in terms of the ECG classification accuracy of 99.62% after reaching the optimal number of trees, which is 584 trees (see figure 3.7). We note that the simulated annealing performs with a linear complexity specifying SA algorithm.

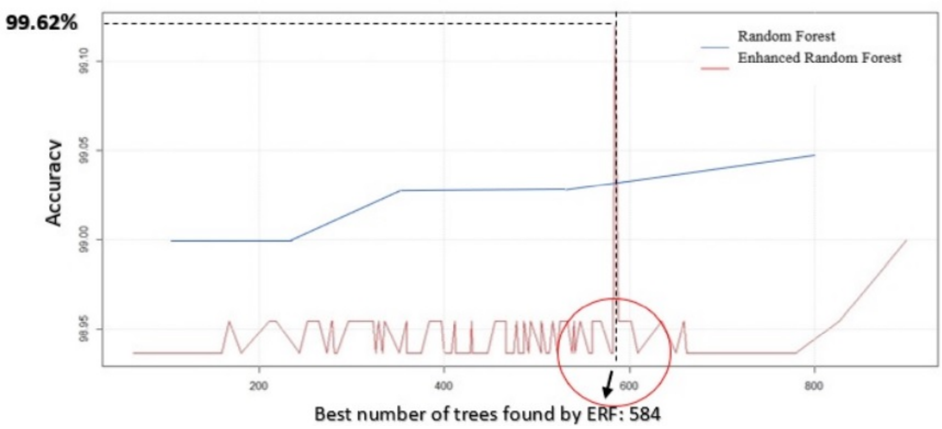


Figure 3.7: Comparison of Prediction Accuracy with MIH-BIH Arrhythmia Database and European ST-T Database

3.6 Conclusion

The ECG is vital physiological signal that is used exclusively to know the state of the cardiac patients. The operation of the features extraction of ECG plays a crucial role in the automatic analysis of ECG. In this chapter, we have proposed an enhanced Random forest algorithm for ECG beat classification combined with a simulated annealing algorithm to ensure a complete automatic and optimal process of the ECG classification. This proposal consists of four major steps, the first one preprocessing, then a feature extraction and the Enhanced Random Forest are performed. The experimental results on various data sets have shown that the classification accuracy is affected by the number of the trees using in random forest classifier and it is improved when we use our Enhanced Random Forest that applies a Simulated Annealing technique to find the optimal number of trees. These results are very hopeful in this health domain, the Enhanced Random Forest method could distinguish the normal class from the classes of medical disorders. Moreover, the abnormal class may represent an unusual emotional state.

Chapter 4

A body area network for ubiquitous driver stress monitoring

Contents

4.1 Introduction **53**

4.2 Motivation: ECG-Based Driver emotion recognition **53**

4.3 An enhanced random forest for driver stress detection based on ECG: our proposal **54**

4.3.1 Phase 1: ECG Signal acquisition 55

4.3.2 Phase 2: Preprocessing phase 56

4.3.3 Phase 3: Features Extraction (BD of features) 57

4.3.4 Phase 4: Classification 57

4.4 Experimental results **58**

4.4.1 Dataset description and experiments environment 59

4.4.2 Results and discussion 59

4.5 Conclusion **60**

4.1 Introduction

During recent years, a body area network is becoming an important tool to improve the healthcare by monitoring patient's health state at distance, at home, in work, or when traveling. This body sensor-based network is easy to use, and available with low cost into two types; the first one can be swallowed or implanted under the skin where the second type is wearable, these became available within the Internet of Things (IoT). Physiological sensor based systems have been recently designed to detect the emotional stress. By this way, we propose in this chapter, a new monitoring system for driver stress detection based on an enhanced random forest classification approach.

This proposal analyses and monitors driver electrocardiogram (ECG) signal when driving in order to discover its stress state belonging to one of the following three levels namely, low, medium or high. This proposal could help to detect and diagnostic stress level and alert the driver, its family and the other road users to avoid accidents caused by high stress state. The proposed system suggests the integration of a simulated annealing algorithm to enhance the random forest classification method in order to reach the highest classification accuracy. According to various drivers' ECG acquired from MIT-BIH physioNet dataset, the experimental study showed that the proposed random forest algorithm outperforms support vector machine (SVM) classification method to detect driver stress levels in terms of recognition accuracy.

4.2 Motivation: ECG-Based Driver emotion recognition

According to Virginia Tech Transportation Institute (VTTI) and US National Highway Safety Administration (NHTSA) organizations, 80% of all US car accidents, are caused by lack of attention while driving [131]. Moreover, according to Dr. Herbert Benson from Harvard Mind and Body Medical Institute, about 80% of medical consultations are stress-related in one way or another. In biology and psychology, the stress is the reaction of an organism to a stressor which could be an environmental condition, so-called physiological or biological stress.

The stress also represents the used method that the body's can react to a condition that can be a threaten his life, a challenge, an aggression or face a new situations or physical and psychological barrier [169]. The most stressors are intellectual, emotional, and perceptual for a human being [170]. Driving in city or highway is very stressful for drivers especially for people who take a lot of time in road like truck drivers and Bus drivers. Moreover, the prolonged stress can diminish the attention of drivers span due to the length of driving time in the road. Stress can be the main cause of most traffic accidents. The availability of portable sensors can be worn or implanted in the Internet of growing things (IoT), we are then able to improve the healthcare and diagnose certain diseases through the measurement of physiological functions, such as blood pressure, heartbeat. Many researches in this field have been done, whose purpose is to detect stress using multiple physiological sensors like skin conductance, body temperature, heart rate variability (HRV), respiration rate, blood pressure, Electroencephalogram (EEG), Electrocardiography (ECG) and Electromyogram (EMG). Many studies ignore the Electrocardiography (ECG), because of some limitations of measurements of ECG signal that often require 12 to 16 leads, in addition to the high quality of signal measurement required to detect some heartbeat. On the other side, ECG signals are very useful for the accuracy of detection of R-peak, allowing to measure the heart rate in an excellent way. With the use of non-invasive portable patches and other sensors, car driver ECG monitoring can be done simply. This makes stress detection, as a very interesting field of study and research [171].

For this end, we suggest in this study, a new monitoring system to detect driver stress using a body area network that senses the driver ECG and then classifies this signal to discover one of the following stress state: low, medium or high. To do that, we introduce a random forest (RF) classification approach combined with a simulated annealing algorithm (SA) responsible for finding the optimal number of trees considered as the main parameter required by RF [163], which is often set by the user randomly. This process could help to reach the highest classification accuracy compared to the native RF or other classification methods like SVM.

4.3 An enhanced random forest for driver stress detection based on ECG: our proposal

There are many methods of recognizing emotions [172] and each of them use different characteristics. Nowadays, the monitoring of driver's emotions is a big challenge due to two causes where the first one is to develop a method trying to achieve a very higher recognition rate with a result in real-time. For the second cause, the input signals must be processed with high quality for a better interpretation. The aim of our study is to develop an efficient and practical method seeking to accurately identify drivers stress while driving. Therefore, we propose a new Enhanced Random Forest, to detect stress driver from ECG signals. The proposed system is composed of

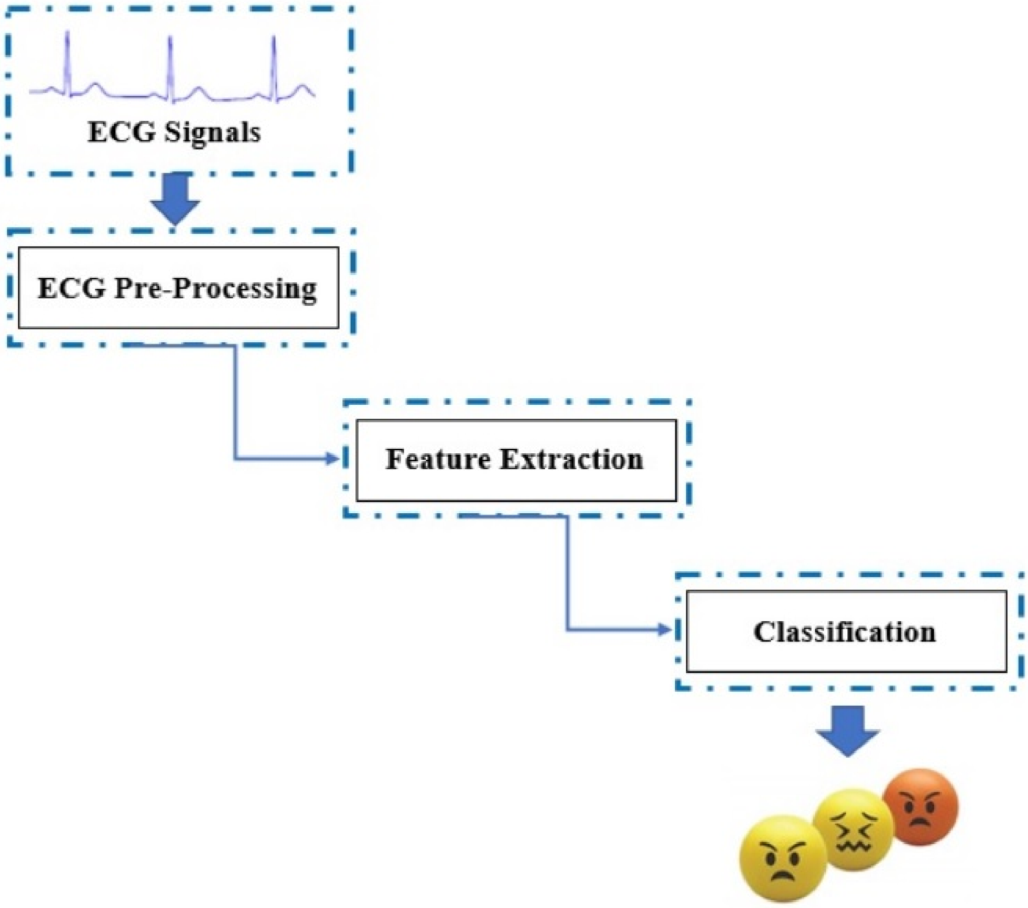


Figure 4.1: Block diagram of the proposed method using Enhanced Random Forest classifier

04 steps: the first step is to collect ECG signal data, then a second step is performed to treat and to denoise this data, after that, the extraction of features from this aggregated data is performed and finally, these features are determined from ECG signals using the enhanced random forest to recognize and classify the psychological state of drivers. The classification result helps to detect stress from ECG signals of drivers while driving in different types of environmental stress (in city, in highway, etc.). In order to extract the morphological features of ECG signal, such as QRS complex, P-wave and T-wave, we use a Discrete Wavelet Transform (DWT). The proposed system is evaluated on data from PhysioNet: MIT-BIH PhysioNet Multi-parameter Database. The global methodology followed in our work is presented in figure 4.1. Our proposal is detailed as follows:

4.3.1 Phase 1: ECG Signal acquisition

In this first step, we use ECG signals of stress from MIT-BIH PhysioNet Multi-parameter Database [158] available on [159] (PhysioNet, 2019). 17 participating drivers were participated in

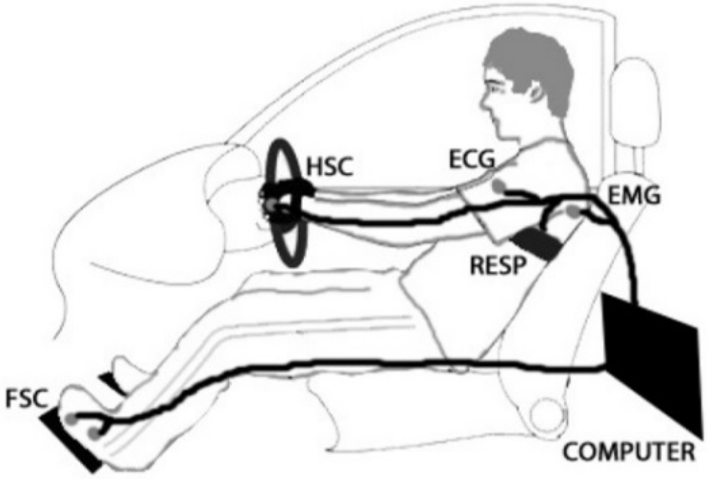


Figure 4.2: General placement of physiological sensors[173].

this experience and 08 types of raw data which are: ECG, Electromyogram (EMG), time stamp, foot galvanic skin response (GSR), hand GSR, Intermittent Heart Rate (IHR), respiration and marker, the general placement of the physiological sensors which are placed on the driver showed in figure 4.2. The drivers drove on a specific road in Greater Boston area, from MITs East Garage to River Street Bridge, this road passes through 03 cities and 02 highways. The datasets used in our study were segmented according to stress levels as follows: low stress (initial rest and final rest), moderate stress (while driving in highway), and high stress (while driving in cities), the stress gained is based on traffic conditions only, and not for another reason. However, the time durations taking by drivers period in rest, in highway and in city calculated by Akbas [173] (see table 4.1). According to Akbas, there are only 10 drivers datasets that can be used in our study, because in 07 of the datasets, the segmentation of different driving periods was not well defined [173]. We have used the methods available from Physionet [158] in order to extract the annotated files for every driving period of the ten drivers.

4.3.2 Phase 2: Preprocessing phase

In this phase, the data must be cleaned from several types of noises because of the nature of the human body (various sources of noises such as: muscular activities, powerline, skin stretching, the electrodes motion, movement of the heart due to respiration, etc), the noises also can be generated from the method of measuring ECG. For this reason, we need an efficient method to remove it, in our study, we consider this step as a black box, using technical denoising approaches

Table 4.1: The time interval of each 07 segments of driving of usable bio-signal datasets.

Rec. Name	Driving period (min) Initial Rest	City1	Hw1	City2	Hw2	City3	Final Rest	Total Rec. Time(min)
Driver05	15.13	16.00	7.74	6.06	7.56	14.96	15.78	**83.23**
Driver06	15.05	14.49	7.32	6.53	7.64	12.29	15.05	**78.37**
Driver07	15.04	16.23	10.96	9.83	7.64	10.15	15.03	**84.88**
Driver08	15.00	12.31	7.23	9.51	7.64	13.43	15.07	**80.19**
Driver09	15.66	19.21	8.47	5.2	7.06	13.21	**NA**	**68.82**
Driver10	15.04	15.30	8.66	5.27	7.04	12.06	14.79	**78.16**
Driver11	15.02	15.81	7.43	7.15	6.96	11.72	14.99	**79.08**
Driver12	15.01	13.41	7.56	6.50	8.06	11.68	15.01	**77.23**
Driver15	15.00	12.54	7.24	5.99	6.82	12.12	15.00	**74.71**
Driver16	15.01	16.12	7.14	5.12	6.81	13.91	**NA**	**64.11**

conceived in signal processing module like in [162], this is illustrated in figure 3.3.

4.3.3 Phase 3: Features Extraction (BD of features)

In this important phase, we extract 17 important and fiducial points according to (P, Q, R, S) interval features, averaged over the time intervals from the annotated ECG signals. In our experiments, we include features related of T-wave, because there are studies proved that the QT interval is very useful and considerate as a distinctive biological marker to identify the cardiac anomaly [174]. It should be noted that in the cardiac cycle, the QRS complex and R-peak are seen as the most crucial features, that can be very helpful, to understand the ECG rhythm reporting, and consequently to deduce the psychological state of the driver. After that, we subdivide the new features database into two parts, where the first part designated for training and the second one is considered to validate the classification. The training data is generated from this features database, by the bagging method in order to form k subsets of training data D1,D2,...,Dk, by selecting randomly samples from data D with replacement. In our study, we use two cases: for the former, we consider 70% of data for training and 30% for test, and for the latter case, 90% of data is for training and 10% is for test.

4.3.4 Phase 4: Classification

Using the aggregated ECG signals in phase 3, predicting the different level of stress is considered as a typical and important classification task in data mining. To achieve this, we have designed an Enhanced Random Forest, and we have applied the traditional Random Forest and support vector machines (SVM) for comparisons. There are 3 different stress levels of drivers for classification purposes as following: level 1: explains the low stress (rest state), level 2: explains the moderate stress (highway driving) and level 3: explains the high stress (city driving).

- **Enhanced Random Forest** Random Forest method was applied in various domains like in [64]. However, to the best of our knowledge, RF-based study had been conducted in literature to detect driver stress. Therefore, we propose to apply a random forest classifier, which is combined with a Simulated Annealing (SA) algorithm [157], in order to find the optimal number of trees required by this classification method. Consequently, a driver stress level (among low, medium or high cases) could be detected with highest accuracy, helping to worn the driver, its family, and the other road drivers. In the first step, SA generates and saves a random tree number and performs a RF classification accuracy. After that, the main SA loop is run, in which a new number of trees is generated to check the new appropriate accuracy acquired and to make a comparison with the best found values in prior iterations, and so on until the best number of trees is reached with the highest accuracy, as shown in figure 3.4.

  SA is applied to find the optimal number of RF trees as follows: In the first step, the system initializes the number of trees at random. T value (Temperature) is also fixed as an integer with a high value. Thereafter, the system calculates the classification accuracy with the chosen number of trees, and saves its as an optimal trees number $nb.trees$. As a second step:

  1. The system generates a random neighboring solution $nb.trees = nb.trees + \Delta$ x and applies RF with this generated number of trees.
  2. It calculates the accuracy with the new neighbor solution Acc = Accuracy($nb.trees$).
  3. The new accuracy is compared with the best found one. If the new one is not the better, the system could accept this new solution with a probability P = ^{3x}, otherwise, the system keeps the new neighbor solution considered as the best one. The sub steps from 1 to 3 are repeated until reaching a stopping criterion.

4.4 Experimental results

In this work, our objective is to suggest an efficient classifier to detect stress of drivers while driving in roads, this section validates the proposed Enhanced Random Forest after a set of comparisons against the conversional classifiers namely, traditional Random Forest and SVM. For this validation, we consider the online MIT-BIH PhysioNet Multi-parameter Database from Physionet website explained below. Using data selected as source data, we train our proposed system on the R studio environment, the obtained model is submitted to the test and tuning step to validate this model.

4.4.1 Dataset description and experiments environment

The ECG signals of drivers stress used in this study, were obtained from MIT-BIH PhysioNet Multiparameter Database (PhysioNet, 2019) [158]. The proposed method was evaluated on the ECG databases from physionet, about 70 potential instances, the used databases contain 68 instances, for 10 participated drivers and for 7 intervals of driving for every driver. The experiments use 67 instances with 17 different extracted attributes (see figure 4.1). We have used the large library of software PhysioToolkit for these chosed databases, this library was presented for physiologic signal processing and analysis only [24]. Toward applying this suggested method, we need the Wave Form Data Base (WFDB) Software Packages [166] and for signal processing, we need to perform the interoperable command-line tools. This last is originally Unix compatible. The Electrocardiographic (ECG) signals are considerated as very sensitive signals which could be affected or corrupted from several different types of noise. The phase of elimination of all types of noise is essential for best results. For non-stationary signals like ECG and EMG, we need Wavelet transform (WT) tool [162].

4.4.2 Results and discussion

Tables 4.2 and 4.3 present the classification results of our proposal method, with a comparative study against RF and SVM classifiers.

- **Results on 2 Classes and 17 Features:** The results offered in the table 4.2 for 17 attributes and 67 instances, are the percentages of the classification accuracy on the dataset with 02 classes: low and high stress. From these results, we found that the most important attribute is the average heart rate. The obtained results for 2 classes: low and high stress and with 17 features proved the performance of our proposal classifier, with the highest accuracy of 99,6% with 87 trees. Additionally, for the small dataset (with a 90% split) the accuracy achieved is 99,7% for Enhanced Random Forest with a number of trees equal to 90. For traditional RF and SVM classifiers, the accuracies achieved are 99,68% and 30,12%, respectively.

- **Results on 3 Classes and 17 Features:** The results presented in the table 4.3 for 17 attributes and 67 instances, are the percentages of the classification accuracy on the dataset with 03 classes low, medium, and high stress during the rest, highway, and city driving periods respectively. The obtained results for this 03 classes proved the performance of our proposal classifier, with the highest accuracy: equal to 97,02% with a number of trees fixed automatically at 914. Moreover, for the small dataset (with a 90% split) the accuracy achieved is 96,33% for Enhanced Random Forest with a number of trees fixed at 932. For SVM classifier the accuracy achieved is 43,86% in 90% split, where traditional RF offers 95,03%.

Table 4.2: Recognition rates for 02 classes (low/high) with 17 attributes.

Classification methods	70% split	90% split
Random forest	97%	99.68%
Enhanced Random forest	99.6 with 87 trees %	99.7% with 90 trees
SVM	97.06%	30.12%

Table 4.3: Recognition rates for 03 classes (low/medium/high) with 17 attributes.

Classification methods	70% split	90% split
Random forest	94%	95.03%
Enhanced Random forest	97.02 with 914 trees %	96.33% with 932 trees
SVM	43.35%	43.86%

4.5 Conclusion

Monitoring and diagnosing personalized individual stress levels are useful and very important in order to save many human lives in roads and to avoid many accidents. In this paper, an approach to stress detection has been proposed, studied and evaluated. This approach suggests an Enhanced Random Forest method which combines traditional Random Forest with Simulated Annealing algorithm, whose goal is to ensure a complete automatic and optimal process, to detect three levels of stressed car drivers from ECG signals. The accuracy obtained with the Enhanced Random Forest achieved 97% in detecting the 03 levels of stress: low, medium and high for 17 features extracted from ECG signals of drivers such as heart rate, QRS complex, P-Wave. According to the obtained results, we could detect the high stress of driver successfully, and this in comparison to his/her rest period (the 02 levels of stress: low and high) with close to 100% accuracy. The classification model that gives a high classification rate is saved. Later this model is used to detect the stress level of the driver in real time, helping to avoid false detection.

Chapter 5

A New Data Augmentation Convolutional Neural Network for Human Emotion Recognition based on ECG Signals

Contents

5.1 Introduction 63

5.2 Motivation: Data augmentation methods for ECG 63

5.3 The proposed ECG data augmentation for human emotion recognition using seven-layer CNN model 64

5.3.1 Data acquisition and preprocessing of ECG signal 65

5.3.2 Data augmentation strategy 65

5.3.2.1 Step 1: Detecting R-waves 65

5.3.2.2 Step 2: Periods calculation of R-R intervals 65

5.3.2.3 Step 3: Random selection of new R-R intervals 66

5.3.2.4 Step 4: R-R intervals concatenation 67

5.3.3 HRV features extraction 68

5.3.4 Architecture of seven-layer CNN model for ECG emotion recognition system 69

5.4 Experimental results 72

5.4.1 Experiments 72

5.4.2 Results obtained and discussion 73

5.4.2.1 Accuracy of valence detection 73

5.4.2.2 Accuracy of arousal detection 74
5.4.2.3 Accuracy of dominance detection 75
5.4.2.4 The confusion matrices: classification correctness 76
5.4.2.5 Precision, recall, and F1-Score 79
5.4.2.6 K-fold cross validation 79
5.4.2.7 PR and ROC curves 80
5.5 Conclusion 86
1 Summary of contributions 87
2 Perspectives and future work 88

5.1 Introduction

In the literature, various machine learning algorithms were proposed for human emotion recognition based on electrocardiogram (ECG) signal. However, the recognition accuracy of these techniques is hampered by the hardness of acquiring huge and balanced number of ECG dataset samples, which is considered as a major challenge in this topic. Therefore, we propose in this chapter, a new data augmentation convolutional neural network (CNN) for human emotion recognition based on ECG signal. Specifically, we suggest to enrich the ECG dataset by a significant number of representative ECG samples, generated according to randomize, concatenate and resample realistic ECG episodes process. Hence, a new seven-layer CNN classifier is suggested, consisting of seven layers to detect human emotions in terms of valence, arousal, and dominance levels.

5.2 Motivation: Data augmentation methods for ECG

As mentioned above in chapter 2, ECG-based emotion recognition has received great interest from researchers especially by machine learning community. However, the success of the ECG analysis based on machine learning approaches depends mainly on a rich annotated dataset. Additionally, generating an annotated ECG dataset with high quality remains a major challenge. In fact, the model trained with small datasets does not generalize well data from the validation and test sets then, the study results will not be precise. Consequently, this type of models suffers from the problem of overfitting occurred when a good fit is achieved on the training data, while the model does not generalize well on new and unseen data. Data augmentation is one of the best solutions to reduce overfitting on models, it is based on increasing the amount of training data. In the literature, there are several data augmentation methods proposed to extend ECG dataset. Although the considerable success given by data augmentation methods cited above and applied for various research areas like medical image analysis tasks, this scheme was not applied yet to human emotion recognition, known as a critical domain.

Therefore, we propose in this work a new data augmentation convolutional neural network (CNN) for human emotion recognition based on ECG signal, where the data obtained is further used as an input for the new seven-layer CNN classifier to classify the different types of human emotions.

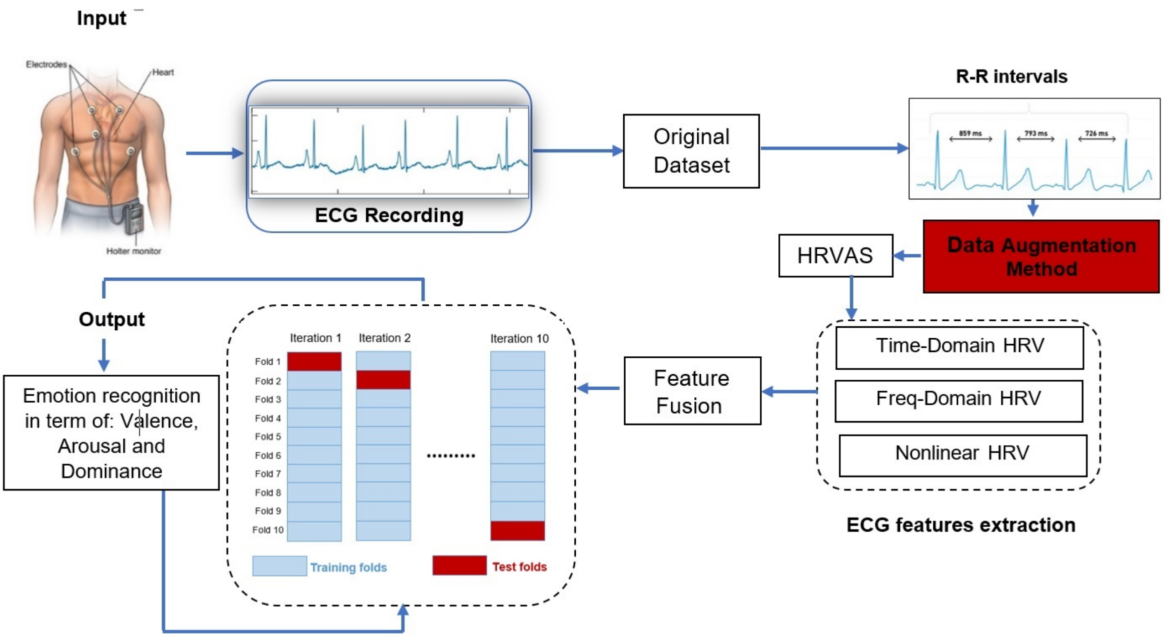


Figure 5.1: Overall block scheme of the proposed method for ECG-based emotion recognition system

5.3 The proposed ECG data augmentation for human emotion recognition using seven-layer CNN model

To recognize human emotion and to cope with the issue of limited size and imbalanced samples used to ML approaches, we detail in this section our proposal [175], presented in figure 5.1. We start with ECG signal acquisition and preprocessing, then the augmented data are generated to increase data samples using our novel approach. In order to extract heart rate variabiliy, we apply a HRV analysis tool named HRV Analysis Software (HRVAS). HRVAS extracts various HRV features such as time-domain features (e.g. RMSSD: Root Mean Square of the Successive Differences and SDNN: standard deviation of normal to normal R-R intervals and), frequency-domain features (e.g. LF: Low Frequency power, HF: High Frequency power and LF/HF) and nonlinear domain features (e.g. Sample entropy (SampEn)). For more details about how these features are calculated, please refer to [176]. According to these extracted features obtained from an ECG signal, the classification is performed by the suggested CNN to recognize the human emotion class such as valence, arousal, or dominance. In the next subsections, we present a detailed description of each step.

5.3.1 Data acquisition and preprocessing of ECG signal

In data analysis science, the data used for learning are often benchmark data, however, in this healthcare domain, this kind of datasets is limited in records number (i.e. samples quantity) and in diversity (i.e. different types of people like young or old people, men or females, etc.). The data augmentation may shed to this issue. In order to enrich the dataset by introducing unobserved and various samples, we propose to start by the use of a publicly available ECG dataset like DREAMER [15] expressing humans with valence, arousal, or dominance emotions. This dataset is considered as a small ECG datasets where the total size of the ECG for this dataset is 414 (for 23 persons exposed to 18 trials and each example of ECG contain 2 channels) [177]. More specifically, the ECG DREAMER dataset is a multi-modal database for analyzing emotions resulting from a set of audio-visual stimuli which are in the form of movie clips.

Furthermore, this dataset includes ECG records collected only from 23 participants, where each participant watches 18 videos chosen to evoke nine special emotions namely calmness happiness, excitement, fear, surprise, sadness, disgust and anger. Therefore, the participant's rated his emotional response in terms of valence, arousal and dominance on a scale from 1-5. We remind that emotions are recognized in the basis of HRV, which is in its turn detected using ECG R-peaks (i.e. R-R intervals) [178]. It is worth noting that R-R interval corresponds to the interval between two successive R peaks in an ECG. To do so, we used R-wave detection algorithm from the input ECG signals to extract R-R intervals series (see figure 5.2) [179].

5.3.2 Data augmentation strategy

In the aim generating many and balanced ECG samples, we suggest a data augmentation process over four steps, presented in figure 3 and illustrated as follows.

5.3.2.1 Step 1: Detecting R-waves

Based on the R-wave detection algorithm, the R-waves of the QRS complex of the normalized ECG signals are detected R1, R2,..., Rn, as shown in figure 5.3.

5.3.2.2 Step 2: Periods calculation of R-R intervals

Different periods between successive R-waves (R-R intervals) are then calculated, (R-R interval(1), R-R interval(2),..., R-Rinterval(n)) as shown in figure 5.4.

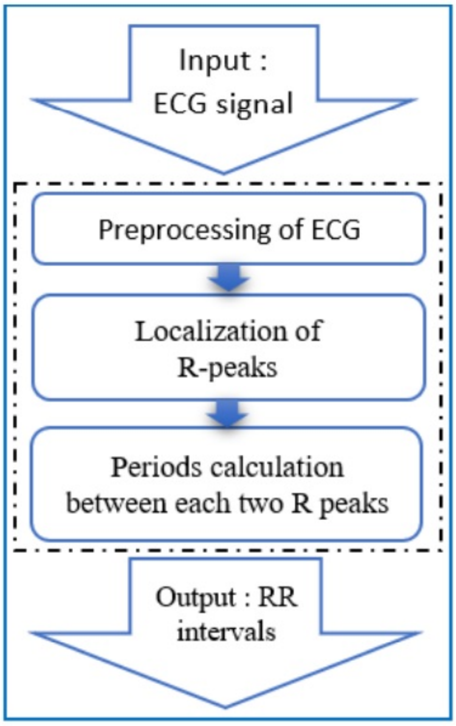


Figure 5.2: R-wave detection process

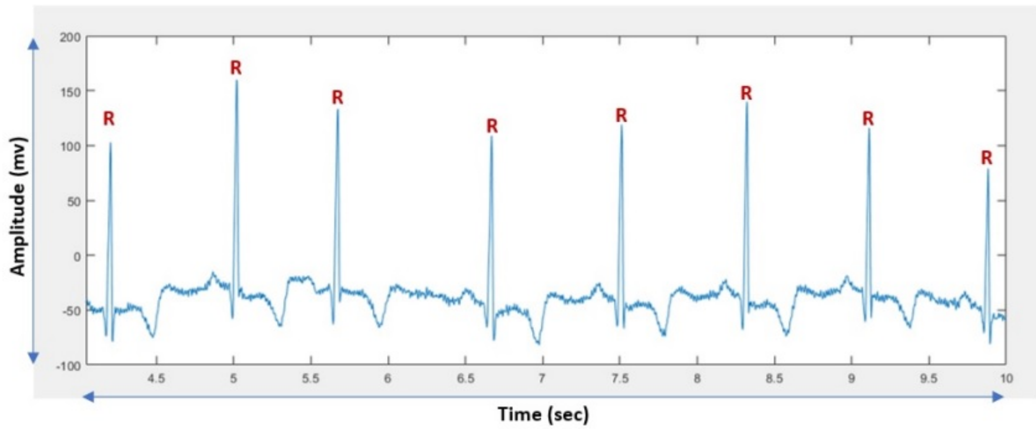


Figure 5.3: Detecting R-waves

5.3.2.3 Step 3: Random selection of new R-R intervals

To extend ECG samples and records, all the nn extracted R-R intervals from each signal is arranged and resampled in one signal with a new order. As shown in figure 5.5, RR1, RR2,..., RRn are selected to be arranged and resampled as following: RR7, RR6, RR1,..., RR4. These new R-R intervals are put in the selected ECG signal at random as shown in figure 5.5.

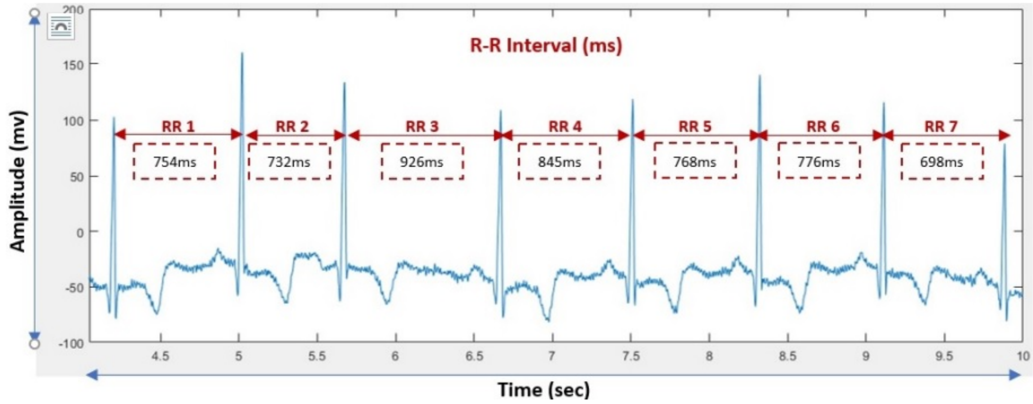


Figure 5.4: Periods calculation of R-R interval.

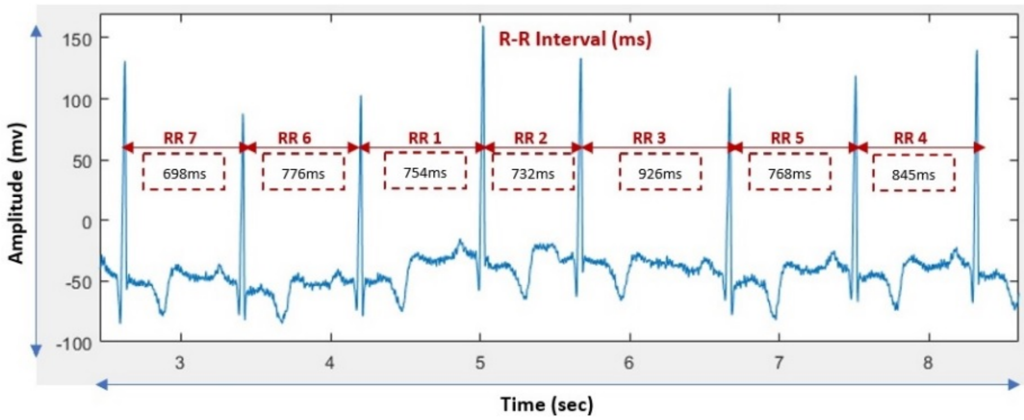


Figure 5.5: Random selection of new R-R interval.

5.3.2.4 Step 4: R-R intervals concatenation

In this step, the generated and selected R-R intervals are concatenated together and added to previous ECG records which giving rise to a new ECG signal like: RR7, RR6, RR1, RR2, RR3, RR5, RR4, as depicted in figure 5.5.

Our system starts with the use of an ECG dataset benchmark like DREAMER dataset to detect the human emotion, this dataset consists of only 414 samples (records) of ECG signal. The number of samples of the different ECG signal, could be balanced by adding the same number of samples in each signal of any of categories in order to enrich each category. To do so, we randomly extracted 24 samples from each ECG signal, the total number of samples is above 10000 samples. It should be noted that the resampled signals are the same length as the original's ECG recording. In this research activity, we try to examine the effects of various values of dataset size (like dataset containing 5000 or or 10000 samples, or more) on neural network training phase.

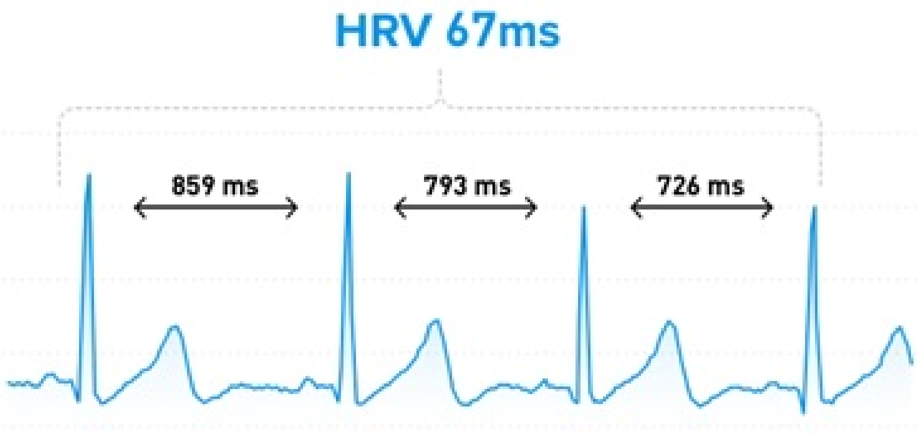


Figure 5.6: Heart Rate Variability.[186].

5.3.3 HRV features extraction

Standard HRV Measures: In the literature, the ECG in specific HRV signals play a crucial role in research on emotion assessment. In fact, the HRV is a measure that indicates variation in the HR over time as shown in figure 5.6. The ANS is responsible for control this variation through the balance between the SNS and PNS, in order to react to the daily stressors as well as to control the human body's most important systems such as respiration, digestion and heart rate. In summary, HRV is one of the most interesting and noninvasive way to recognize the ANS imbalances of the human because it has a direct impact on the activity of the heart [180; 181]. Standard HRV analysis is to extract a variety of parameters that are specified in the frequency domain and also in the time domain [182–184]. In the field of ECG-based emotion recognition, Zhao et al. in[185] studied the differences of HRV indices between six different emotions: happiness, sadness, anger, disgust, fear and neutral, the obtained results proved that there are important differences of HRV indices between these emotions. In this work, we relied on ECG signals to extract a series of features as well as its derived HRV features extracted from frequency-domain, time-domain, and nonlinear domains. Measuring these features is illustrated as follows:

- **Time-Domain Features:** in this domain, we measure the variation in heart rate over time (i.e. the intervals between successive normal cardiac cycles). This variation reflects to do some easy calculations such as: calculate the mean normal-to-normal (NN) intervals, the variance between NN intervals, the standard deviation of NN (SDNN) and the root mean square of differences between adjacent R-R intervals (RMSSD).

- **Frequency-Domain Features:** it is a complex analysis technique, the role of this technique is to show the amount of signal that lies one or more frequency bands (ranges). For the HRV, the technique uses the frequency bands that could tend to correlate with some physiological phenomenon (e.g. Parasympathetic nervous system activity) [187]. In addition, the influence of sympathetic and parasympathetic nerves on HRV can be distinguished by this method very well. From the power spectral density (PSD) analysis, several features are calculated in the frequency domain analysis. However, the Power Spectral Density (PSD) analysis is used to understand HRV, furthermore, there are three spectral bands of the PSD which are: Very Low Frequency (VLF) with spectral components less than 0.04 Hz; Low Frequency (LF) belongs to the interval [0.04, 0.15Hz] and High Frequency (HF) defined in the interval [0.15, 0.4Hz].

- **Nonlinear Features:** In the nonlinear analysis, several features are calculated, including ECG-derived respiration (EDR) related parameters, nonlinear dynamics related parameters, Poincaré plot related parameters, and self-correlation related parameters.

The extracted HRV features groups are listed in table 5.1. The use of HRV becomes an increasingly popular and important tool to identify the emotion of a human. In this phase, we suggest the use of the HRV Analysis Software (HRVAS) [188] which is a HRV analysis tool developed using Matlab software [189]. It is used to extract HRV features including the time-domain (e.g. SDNN and RMSSD, the frequency-domain (ex: Low-Frequency power (LF), High-Frequency power (HF) and LF/HF) and nonlinear domain (ex: SampEn), more details about the calculation of these parameters are presented in [176].

5.3.4 Architecture of seven-layer CNN model for ECG emotion recognition system

CNN is usually consisted of two main parts; the former is a feature extractor, which is responsible for an automatic features learning from raw input data, while the latter is considered as a fully connected multi-layer perceptron (MLP). This MLP is responsible for the classification according to the learned features, realized by the first part. The architecture of seven-layer CNN model for ECG emotion recognition system is presented in figure 5.7. The network consists of seven layers, including four convolutional layers (Conv1D), two max pooling layers, one fully connected layers, and one Softmax layer. The convolution operations are performed by the convolutional layers C1, C2, C3 and C4 according to equation 5.1, where each layer uses the output of the

Table 5.1: Notation of features extracted from ECG[190].

Features group	Variable	Unit	Description of the extracted features
Time-Domain features	SDNN	ms	standard deviation of all NN intervals
	SDANN	ms	Standard deviation of the averages of NN intervals in all 5-minute segments of the entire recording
	RMSSD	ms	The square root of the mean of the sum of the squares of differences between adjacent NN intervals
	SDNN index	ms	Mean of the standard deviations of all NN intervals for all 5-minute segments of the entire recording
	SDSD	ms	Standard deviation of differences between adjacent NN intervals
	NN50 count		Number of pairs of adjacent NN intervals differing by more than 50 ms in the entire recording; three variants are possible counting all such NN intervals pairs or only pairs in which the first or the second interval is longer
	pNN50	%	NN50 count divided by the total number of all NN intervals
	HRV		Total number of all NN intervals divided by the height of the histogram of all NN
	TINN	ms	Baseline width of the minimum square difference triangular interpolation of the highest peak of the histogram of all NN intervals
Frequency-Domain features	Total power	ms^2	Variance of all NN intervals < 0.4 Hz
	ULF	ms^2	Ultra low frequency < 0.003 Hz
	VLF	ms^2	Very low frequency < 0.003–0.04 H
	LF	ms^2	Low frequency power 0.04–0.15 Hz
	HF	ms^2	High frequency power 0.15–0.4 Hz
	LF/HF ratio	/	ratio of low-high frequency power
Nonlinear features	SD1	/	The standard deviation of the Poincare plot perpendicular to the line-of-identity
	SD2	/	The standard deviation of the Poincare plot along the line-of-identity
	SD1/SD2	/	Ratio of SD1/SD2
	ApEn	/	Measures the complexity or irregularity of the RR series
	SampleEn	/	Sample entropy : a tolerance (r) of 0.2 standard deviation of the R-R interval and an embedding dimension (m) of 2.

previous one using the current convolution kernel.

$$x_k^l = f(\sum_{i\in \mathrm{M}_k} x_i^{l-1} * w_{ik} + b_k) \tag{5.1}$$

Where, x_k^l is the output of the $k-th$ neuron in layer l, M_k represents the effective range of the convolution kernel, w_{ik} is the weight kernel between the i-th neurons in layer $l-1$, b_k represents the bias of the k-th neuron in layer l and the $k-th$ neuron in layer l, in addition, $f()$ is used as the activation function, we opt for a Rectified Linear Unit (ReLU) as an activation function [191].

First of all, our system has as an input the extracted features vector with a length of 104 features (e.g. SDNN,...). This vector is sent to the first layer C1, which applies a kernel size of 5 (i.e. filter) and 64 feature maps. So, the output of C1 layer is calculated as follows: $input\_size - (kernel\_size-1) = 104-(5-1) = 100 features$. Note that the relationship (feature maps-reached features) are written (feature maps@reached features) such as (64@100). The found 100 features are then presented to the next layer C2 with convolution kernel size of 5, which in its turn (i.e. C2) gives rise to 64@96 (as output). Next, max-pooling S1 is used to reduce the number of computations with a kernel size of 2, so the output is 64@48.

The subsampling layers (also called Pooling layers) (i.e. S1 and S2) are used to minimize the input size of the next layer and to extract more useful features. We suggest applying the Max pooling layers to keep only the effective features from the Conv1D layer. The equation 5.2 calculates the output of the $k-th$ neuron of the subsampling layers.

$$x_k^l = subsample(x_{k_{cluster}}^{l-1}) \tag{5.2}$$

Where, x_k^l is the output of the $k-th$ neuron of layer l, which is calculated by the down sampled operation performed on the output of $k-th$ cluster of layer l–1.

Similarly, C3 layer is performed after the max pooling process where the output is 128@46 (here, the filter size is 3). Next layer is C4, having as output 128@44. For the next step, S2 is applied and the output pooling features of S2 (128@22) are rearranged as a feature vector and input to the fully connected layer F1 for classification.

The vector resulting from ECG features extraction state is dispatched to the input neurons of the fully connected network layers (FCN) F1, this might help to perform the training and testing process of the model. The phase of prediction is performed by the final FCN. The equation 5.3 is used to calculate the output of the neuron in fully connected layer which is equal to 256.

$$x_k^l = f(\sum_{i=1}^{\mathrm{N}} x_i^{l-1} * w_{ik} + b_k) \tag{5.3}$$

Where, x_k^l is the output of the $k-th$ neuron in layer l, w_{ik} is the weight vector between the $k-th$ neurons in layer l and the $i-th$ neuron in layer $l-1$, b_k represents the bias of the $k-th$ neuron in layer l. However, the total number of neurons in layer $l-1$ is defined by N.

In the output layer, softmax activation function is used for the final classification to obtain the five levels of the arousal, valence, and dominance.

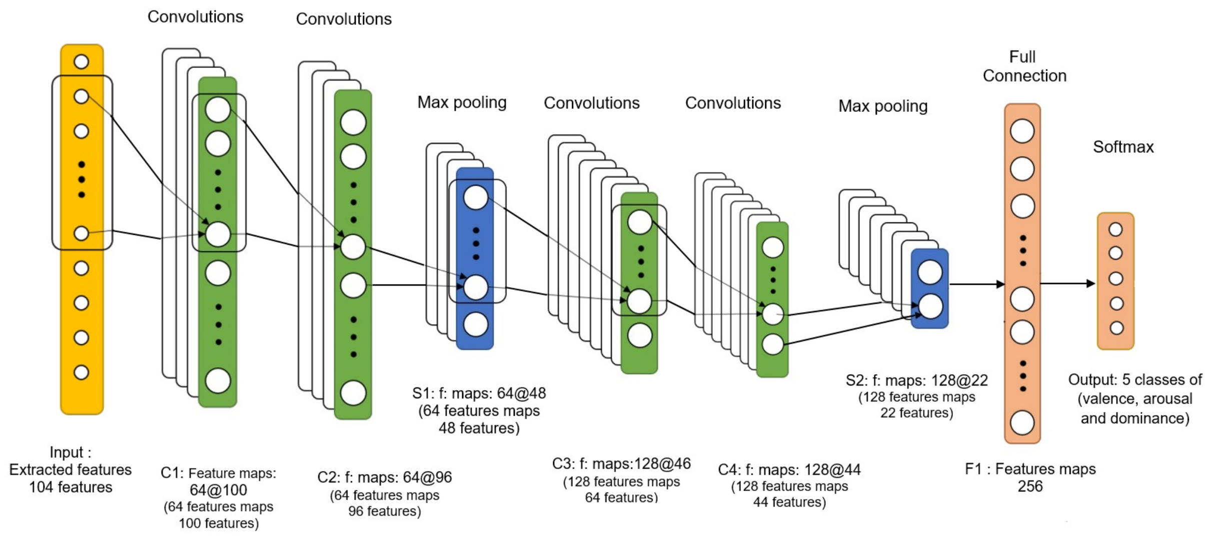


Figure 5.7: Architecture of seven-layer CNN model for ECG emotion recognition system.

5.4 Experimental results

5.4.1 Experiments

In this section, we describe the realized tests and obtained results of a series of experiments including the phase of features extraction and classification. In order to evaluate the effectiveness of the proposed method for ECG data augmentation approach to detect the emotion, comparative experiments are conducted in this work using different classifiers; this is done with and without data augmentation. The data generated from the proposed method will be used by the classifier in the learning phase. For our contribution, an seven-layer CNN classifier is adopted. The considered classifiers for comparisons are: NN, and SVM. Additionally, we also study the influence of every augmentation of samples on the classifiers by the use of 5000 and more than 10000 samples. We have used R studio platform to train the developed networks for emotion detection with and without data augmentation. We used in this study, the public database DREAMER containing 23 individuals' ECG data (14 males and 9 females), this data was elicited by audio-visual stimuli and using low-cost devices. Moreover, 18 film clips with a duration of 65 – 393s are proposed to tested persons in order to elicit different emotion. Note that the ECG signals were recorded at 256 Hz. Next, the proposed method is run to increase the number and the diversity of ECG dataset.

For the phase of extracting features, the HRV data from the ECG signal based on the extracted peaks are measured using HRVAS Software (HRV Analysis) to extract 104 features. After extracting and merging features, CNN is launched to classify and detect emotions.

5.4.2 Results obtained and discussion

In order to evaluate any classifier in machine learning field, it is common to divide the dataset into two separate sets: a training set and a testing set [192]. To this end, there are different partition schemes that could be applied on the dataset. In our study, we apply both 70-30 technique and 10-fold cross-validation technique. 70-30 technique is performed to illustrate the importance of training data and how vital training information is in the proposed classifier, in this technique, the dataset is split into 70% train data and 30% test data [193]. The 10-fold cross validation technique is used to estimate the skill of a machine learning model on unseen data [194]. Specifically, the 10-fold cross validation involves randomly dividing the dataset into 10 folds or subsets of approximately equal size. Of the 10 folds, a single fold is retained as the validation data for testing the model, and the remaining 9 folds are used as training data. This process is then repeated 10 times, with each of the 10 folds used for testing and the rest for training. This section describes the results obtained from the experiments that were conducted in this research. Results of the classifiers to recognize emotions in terms of valence, arousal, and dominance levels are listed in the tables below.

5.4.2.1 Accuracy of valence detection

Figure 5.8a presents the accuracy comparison of valence dimension between NN, SVM and CNN. For the CNN classifier, we used 63,76% of samples for training and 36,24% for testing, as user parameters experimentally selected given the best found results in this case.

Besides, the figure 5.8b compares these classifiers using 10-fold cross validation with and without data augmentation.

The results without data augmentation showed low classification accuracy with all classifiers with 24.36% for SVM classifier, 22.88% for NN and 22.00% for CNN. The reason for these poor results is the limited number of samples as well as the unbalanced number of each sample in the dataset. Also, figure 5.8a showed that with an augmentation of a number of 5000 samples, the CNN obtained the highest accuracy to detect valence dimension with 59.69% compared to the other classifiers such as SVM with 34.76% and NN with 27.53%. Moreover, the results with an augmentation of a number of more than 10000 samples showed that CNN (with 75,75% of samples for training and 24,25% for testing) performs better than the other classifiers, with an accuracy of 95.16% compared to SVM with 36.01% and NN with 31.02%.

From figure 5.8b and for the case of data augmentation, it can be clearly seen that the CNN

(a) (b)

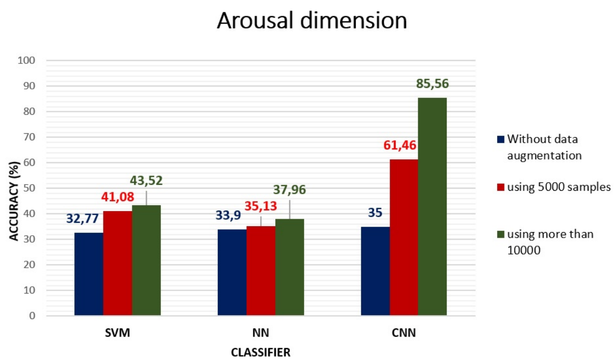
Figure 5.8: Accuracy comparison between CNN and other machine learning with/without data augmentation (emotion expressed in terms of valence). (a: 70% for training, and 30% for test, b: the 10-fold cross validation)

performs better than the other classifiers, with an accuracy of 75.1% compared to SVM with 33.6% and NN with 36.32%.

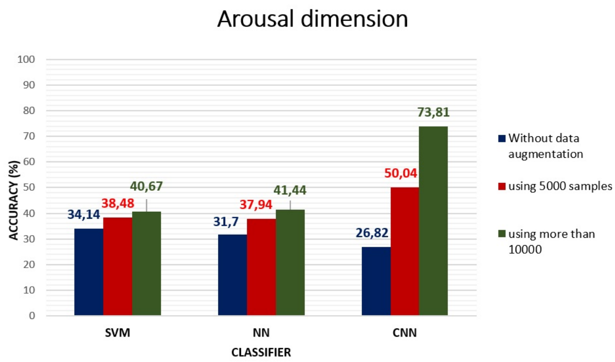
5.4.2.2 Accuracy of arousal detection

Figures 5.9a and 5.9b depicts the accuracy of arousal detected by various machine learning approaches (i.e. SVM, NN, and CNN) in both cases with and without data augmentation.

As shown in figure 5.9a and for CNN classifier, we have fixed 75.84% samples for training and 24.16% for testing, as user parameters selected in an experimental manner, leading to the best found results. Therefore, for the case when is no data augmentation, CNN records an accuracy of 35% to detect arousal, outperforming NN and SVM, which have given an accuracy of 33.90% and 32.77%, respectively. However, in the figure 5.9b, it was observed that using 10-fold cross validation, the accuracy without data augmentation is 32.16% for SVM, 31.7% for NN and 26.19% for CNN. It is clear that these values are very moderate, it is due to the limited number of recorders used to train each Machine Learning (ML) approaches in addition to restricted diversity of these processed samples. On the other side, when 5000 samples are considered to augment data to enhance machine learning performance, figure 5.9a showed that CNN obtained the best accuracy to detect the arousal with 61.46% compared to NN with 35.13%. Moreover, if we consider more than 10000 samples, we remark that CNN (with 75,84% of samples for training and 24,16% for testing) reached also the best accuracy with 85.56% against the other ML approaches that have given 37.96% for NN, and 43.52% for SVM. Even when we used 10-fold cross validation, the accuracy presented in figure 5.9b with data augmentation is increased with 40.31%, 41.44% and 73.81%, for



(a)



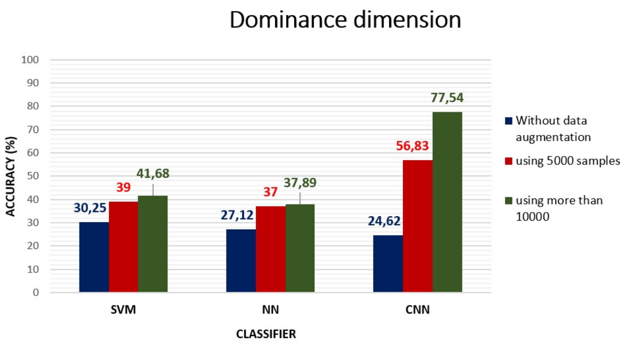
(b)

Figure 5.9: Accuracy comparison between CNN and other machine learning with/without data augmentation (emotion expressed in terms of arousal). (a: 70% for training, and 30% for test, b: the 10-fold cross validation)

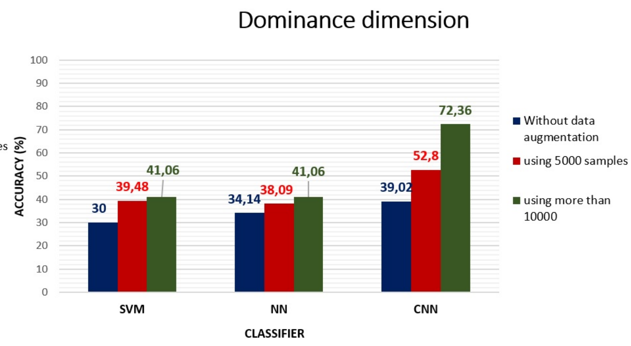
SVM, NN, and CNN, respectively.

5.4.2.3 Accuracy of dominance detection

Concerning the dominance dimension in both cases without and with data augmentation, we used 84,30% of samples for training and 15,70% for testing, as user parameters. Figure 5.10a depicts the accuracy obtained for each model (i.e. SVM, NN and CNN), where figure 5.10b presents the accuracy obtained using 10-fold cross validation. The results presented in this figure using the original dataset (i.e without data augmentation) showed that the accuracy of various classifiers is low due to the use of a small and unbalanced dataset, except for SVM, giving a classification accuracy equal to 30.25% to detect dominance, however, CNN has an accuracy of only 24.62% and NN achieved 27.12%. On the other hand, when the number of samples increases, as presented in figure 5.10a, the accuracy of the classifiers increases gradually. It can be seen that with an augmentation with 5000 samples, the CNN achieved the best accuracy with 56.83% compared with SVM with 39% and NN with 37%. Additionally, the accuracy of the network is also improved with an increase of more than 10000 samples, where the accuracy of the CNN using 83,10% of samples for training and 16,90% for testing is much higher than that of the other classifiers with 77.54% compared to SVM with 41.68% and NN with 37.89%. Moreover, the accuracy presented in figure 5.10b using 10-fold cross validation showed that with data augmentation, CNN achieved highest accuracy with 72.36%, compared to SVM with 40.77% and NN with 41.06%.



(a)



(b)

Figure 5.10: Accuracy comparison between CNN and other machine learning with/without data augmentation (emotion expressed in terms of dominance). (a: 70% for training, and 30% for test, b: the 10-fold cross validation)

5.4.2.4 The confusion matrices: classification correctness

To illustrate the quality of the proposed classifier against SVM and NN for valence, arousal and dominance, tables 5.2, 5.3, 5.4, 5.5, 5.6 and 5.7 present the confusion matrices for both cases without/with data augmentation. It is worth noting that the correct classifications (i.e. true positives) are shown in the diagonals of each dimension (i.e. valence, arousal and dominance). Each column of the each matrix represents the instances in a predicted class (i.e. valence, arousal, and dominance), while each row represents the instances in the actual class. In this study, we mention that the total number of the original dataset is 414 and we note that the choice of training or test samples percentages was done in an experimental fashion, with approximately to 70% / 30% splits between training and test data, leading to the best found results.

- **Without data augmentation case:**

  The table 5.2:a shows the confusion matrix of the training results on the test set for CNN. For this, we have used a roughly equal number of samples for each class of valence, arousal and dominance in training; this number is equal to 30 samples per class. So, in terms of valence, a total of 264 samples is used for training with 63.76% and 150 samples for the test with 36.24%. About the arousal, the total of 314 samples is selected for training with 75.84% and 100 samples for test with 24.16%, however, for dominance, we have opted for 349 samples for training with 84.30% and 65 samples for test with 15.70%.

  When we used the SVM classifier, we have applied the same percentages of dataset partition (training/test) for arousal, valance or dominance as follows: 71% of dataset is reserved for training and 29% for the test. Consequently, 295 samples are devoted for training and 119 samples for testing, as depicted in the Table 5.4:a.

Also for the NN classifier (see table 5.6:a), the same percentages of dataset partition is considered; 71.50% of data are used for training and 28.50% for the test, which means that 296 samples are devoted for training and 118 samples for testing.

When we analyze table 5.2:a, we can see that the CNN is confused when there are no data augmentation with an accuracy of actual/predicted valence equal to 22.00%, 35.00% for arousal, and 24.62% for dominance. We can also see in table 5.4:a that the SVM records only an accuracy of 24.36% for valence, 32.77% for arousal, and 30.25% for dominance. A close outcome is also seen in table 5.6:a, which is given by the NN with only 22.88%, 33.90% and 27.12% for valence, arousal and dominance, respectively.

In addition, we have performed 10 fold cross validation on a training data, the results reached are presented in tables 5.2:b, 5.4:b and 5.6:b, as confusion matrices of optimal model from the 10 generated models, according to CNN, SVM, and NN classifiers, respectively. We can clearly see that all these classifiers records low accuracy, we cite for instance CNN that gives an accuracy of 26.83% for valence, 26.82% for arousal and 39.02% for dominance.

- **With data augmentation case:**

  Considering the total number of original dataset which is 414, the proposed data augmentation increases the number to be 10350, which is calculated as follows: (Total number of original dataset * x) + (Total number of original dataset). Where, x is the number of copies generated from each ECG signal to be added to the original dataset. In our study, x is fixed at 24 (as an experimental parameter), then the new dataset (with data augmentation) will contain: (414*24) + (414) = 10350 samples.

  To evaluate the classification accuracy of the proposed CNN for the case of valence, we have chosen the total of 7840 samples for training with 75.75% and 2510 samples for test with 24.25%. About the arousal, a total of 7850 samples is selected for training with 75.84% and 2500 samples for test with 24.16%, however, for dominance, we have chosen the total of 8600 samples for training with 83.10% and 1750 samples for test with 16.90%. About the SVM, We have selected for all emotion levels, the same sampling as follows: 7301 samples (i.e 70.54%) devoted for training and 3049 samples (i.e 29.46%) for testing, as depicted in table 6:a.

  Concerning the NN and for valence, we have used 70.04% for training which represents 7249 samples and 29.96% of dataset for test with 3101 samples, however, for arousal and dominance, we consider 7460 samples (i.e. 72.06%) for training and 2890 samples (i.e. 27.94%) for test, see table 8:a.

When analyzing different confusion matrices (case of data augmentation), we can see that CNN classifier records the best classification precision with very low confusion. For instance, the results given by table 5.3:a showed that the CNN is able to label and classify correctly each class with accuracy of actual/predicted valence equal to 95.16%, 85.56% for arousal, and 77.54% for dominance.

Nevertheless, the accuracy achieved by SVM presented in table 5.5 is only 36.01% for valence, 43.52% for arousal, and 41.68% for dominance. Also, the table 5.7:a expressed low accuracy values of NN classifier equal to 31.02%, 37.96% and 37.89% for valence, arousal and dominance, respectively.

As depicted in tables 5.3:b, 5.5:b and 5.7:b that found from the 10 cross validation, it has been clearly demonstrated that the accuracy is obviously improved especially for CNN, which performs better than the other classifiers (SVM and NN), with an accuracy of 75.1% for valence, 73.81% for arousal and 72.36% for dominance.

Table 5.2: Confusion matrix for CNN without data augmentation.

a: 70% for training and 30% for test																
		Valence dimension					**Arousal dimension**					**Dominance dimension**				
		Predicted					**Predicted**					**Predicted**				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Actual	**1**	**7**	9	8	6	11	**0**	0	0	0	0	**0**	0	0	0	0
	2	7	**6**	6	5	3	6	**12**	4	4	4	4	**5**	1	1	1
	3	4	5	**7**	8	3	8	4	**12**	6	6	6	6	**5**	5	7
	4	9	6	6	**8**	8	4	3	3	**9**	8	2	2	5	**4**	3
	5	3	4	3	3	**5**	2	1	1	1	**2**	1	0	2	3	**2**
Accuracy		**22.00%**					**35.00%**					**24.62%**				
b: 10-fold cross validation																
		Valence dimension					**Arousal dimension**					**Dominance dimension**				
		Predicted					**Predicted**					**Predicted**				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Actual	**1**	**1**	1	3	3	2	**0**	1	1	2	0	**0**	0	1	0	0
	2	3	**4**	1	2	0	0	**2**	2	4	1	1	**3**	1	4	3
	3	1	2	**1**	3	0	0	4	**5**	4	2	0	0	**6**	3	1
	4	2	1	1	**3**	1	0	1	1	**3**	1	0	3	2	**5**	2
	5	1	1	1	1	**2**	1	2	1	2	**1**	0	2	2	0	**2**
Accuracy		**26.83%**					**26.82%**					**39.02%**				

Table 5.3: Confusion matrix for CNN with data augmentation.

a: 70% for training and 30% for test

		Valence dimension					Arousal dimension					Dominance dimension				
		Predicted					Predicted					Predicted				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
	1	**482**	5	3	1	1	**195**	0	0	0	1	**0**	0	0	0	0
	2	5	**480**	3	3	16	140	**487**	7	5	4	45	**336**	0	3	2
Actual	**3**	1	3	**482**	9	14	114	6	**490**	6	9	228	2	**341**	8	1
	4	10	11	11	**484**	28	50	5	3	**489**	8	52	11	9	**339**	6
	5	2	1	1	3	**451**	1	2	0	0	**478**	25	1	0	0	**341**
Accuracy		**95.16%**					**85.56%**					**77.54%**				

b: 10-fold cross validation

		Valence dimension					Arousal dimension					Dominance dimension				
		Predicted					Predicted					Predicted				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
	1	**160**	7	8	23	2	**39**	9	7	9	1	**22**	4	6	3	4
	2	13	**161**	14	16	6	2	**159**	34	44	11	1	**149**	24	34	3
Actual	**3**	11	19	**151**	22	11	4	23	**245**	41	4	8	28	**203**	27	24
	4	23	22	27	**208**	14	3	18	21	**235**	4	6	29	33	**260**	13
	5	4	4	7	5	**97**	0	10	15	11	**86**	2	12	16	9	**115**
Accuracy		**75.1%**					**73.81%**					**72.36%**				

5.4.2.5 Precision, recall, and F1-Score

As seen in 5.9, and for all the three dimensions valence, arousal and dominance the CNN reaches high precision with a high recall which means that each result restored by a search was relevant. Additionally, a high recall corresponds that the search retrieved all the true positives. Indeed, the F1-score is also calculated using both precision and recall, confirming the highest CNN precision and recall. In contrary, the results obtained without data augmentation in 5.8, the CNN has low precision with low recall (and low F1-score) which means that CNN returns a lot of false positives and returns so few results.

5.4.2.6 K-fold cross validation

In this subsection, we tested the proposed model to check the effectiveness of the CNN classifier on a dataset (with and without data augmentation) with the K-Cross-validation technique, using three statistical metrics which are:

- **RMSE:** Root means square error, that is, how far apart the predicted values are from the observed values in a dataset.
- **MAE:** The mean absolute error, it corresponds to the average absolute error between the model prediction and the actual observed data.
- **R-squared:** The measure of the correlation between the predictions made by the model and

Table 5.4: Confusion matrix for SVM without data augmentation.

a: 70% for training and 30% for test

		Valence dimension					Arousal dimension					Dominance dimension				
		Predicted					Predicted					Predicted				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Actual	**1**	**0**	1	0	0	0	**0**	0	0	0	0	**0**	0	0	0	0
	2	0	**0**	0	1	0	0	**1**	2	0	0	1	**1**	2	2	0
	3	7	6	**8**	7	4	1	5	**8**	7	3	2	8	**9**	13	5
	4	15	16	22	**21**	11	3	17	30	**30**	12	2	15	21	**26**	12
	5	0	0	0	0	**0**	0	0	0	0	**0**	0	0	0	0	**0**
Accuracy		**24.36%**					**32.77%**					**30.25%**				

b: 10-fold cross validation

		Valence dimension					Arousal dimension					Dominance dimension				
		Predicted					Predicted					Predicted				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Actual	**1**	**0**	0	0	0	0	**0**	0	0	0	0	**0**	0	0	0	0
	2	0	**0**	0	0	0	0	**1**	0	2	0	0	**2**	1	4	0
	3	2	5	**3**	1	3	1	1	**5**	4	3	1	3	**2**	5	3
	4	4	6	7	**5**	2	0	7	7	**6**	2	1	6	3	**8**	2
	5	1	0	1	0	**1**	0	0	0	0	**2**	0	0	0	0	**0**
Accuracy		**21.95%**					**34.14%**					**30.0%**				

the actual observations.

The results obtained using 10-fold cross validation 5.10, showed that the lower values of RMSE and MAE reflect the good ability of the model to accurately predict the data, conversely, the high value of R-squared means that the model can predict the actual observations and the model performance is good.

5.4.2.7 PR and ROC curves

We also evaluate Precision-Recall (PR) Curve and Receiver Operating Characteristic (ROC) curve. PR and ROC curves are considered as metric tools to evaluate the performance of the CNN classifier. The PR curve is a graph that represents Precision values against the Recall values, where ROC curve is a graph, representing true positives rate (TPR) values depending on false positive rate values (FPR). On the one hand, Figures 5.11a, 5.11b, 5.12a, 5.12b, 5.13a and 5.13b display PR and ROC curves, of valence, arousal, and dominance given by the CNN classifier without data augmentation, respectively. In these cases the CNN is not able to differentiate between the positive and negative classes. On the other hand, the PR and ROC curves of Figures 5.11c, 5.11d, 5.12c, 5.12d, 5.13c and 5.13d given by the CNN classifier with data augmentation, yield an excellent performance in order to distinguish between positive and negative classes.

Table 5.5: Confusion matrix for SVM with data augmentation.

a: 70% for training and 30% for test

		Valence dimension					Arousal dimension					Dominance dimension				
		Predicted					Predicted					Predicted				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
	1	**75**	28	9	17	26	**1**	0	0	0	0	**1**	0	0	0	0
	2	92	**191**	63	95	21	39	**188**	77	92	64	12	**29**	9	5	8
Actual	**3**	167	143	**250**	183	82	25	96	**339**	126	73	54	289	**536**	274	204
	4	239	225	339	**497**	163	65	414	479	**759**	170	42	281	330	**675**	269
	5	27	5	12	15	**85**	0	2	0	0	**40**	0	0	1	0	**30**
Accuracy		**36.01%**					**43.52%**					**41.68%**				

b: 10-fold cross validation

		Valence dimension					Arousal dimension					Dominance dimension				
		Predicted					Predicted					Predicted				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
	1	**42**	9	10	15	14	**1**	0	0	0	0	**1**	0	0	0	0
	2	15	**56**	11	34	11	24	**58**	52	31	19	7	**10**	2	1	13
Actual	**3**	49	50	**92**	55	39	9	47	**115**	45	38	13	90	**194**	88	68
	4	69	95	97	**147**	63	21	117	156	**233**	53	10	105	105	**213**	108
	5	10	6	7	9	**30**	0	1	1	0	**14**	0	0	0	0	**7**
Accuracy		**35.45%**					**40.67%**					**41.06%**				

Table 5.6: Confusion matrix for NN without data augmentation.

a: 70% for training and 30% for test

		Valence dimension					Arousal dimension					Dominance dimension				
		Predicted					Predicted					Predicted				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
	1	**3**	3	3	4	1	**0**	0	2	1	2	**0**	0	0	0	0
	2	0	**0**	3	1	1	0	**2**	0	0	1	1	**0**	2	5	4
Actual	**3**	9	17	**14**	20	8	2	7	**9**	15	0	4	7	**19**	21	6
	4	3	2	6	**3**	1	3	14	21	**27**	6	0	10	8	**9**	9
	5	5	3	0	4	**4**	0	2	1	1	**2**	0	2	3	4	**4**
Accuracy		**22.88%**					**33.90%**					**27.12%**				

b: 10-fold cross validation

		Valence dimension					Arousal dimension					Dominance dimension				
		Predicted					Predicted					Predicted				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
	1	**1**	0	1	4	0	**1**	0	0	0	0	**2**	0	0	0	0
	2	0	**1**	0	1	0	0	**0**	0	0	0	1	**3**	1	2	0
Actual	**3**	2	2	**2**	3	3	1	4	**1**	2	0	1	1	**2**	2	1
	4	4	2	6	**4**	1	3	5	8	**10**	5	1	3	10	**6**	4
	5	0	2	2	0	1	**1**	0	0	0	0	**1**	0	0	0	0	**1**
Accuracy		**21.96%**					**31.70%**					**34.14%**				

Table 5.7: Confusion matrix for NN with data augmentation.

a: 70% for training and 30% for test																
		Valence dimension					**Arousal dimension**					**Dominance dimension**				
		Predicted					**Predicted**					**Predicted**				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
	1	**161**	95	75	150	51	**1**	0	2	0	0	**5**	0	0	0	0
	2	47	**160**	109	122	53	28	**99**	58	39	20	4	**46**	30	43	8
Actual	**3**	149	149	**269**	201	83	21	102	**185**	120	26	52	172	**386**	243	165
	4	193	219	230	**319**	145	101	432	564	**778**	247	43	331	390	**602**	234
	5	25	15	12	16	**53**	0	9	13	11	**34**	5	23	28	24	**56**
Accuracy		**31.02%**					**37.96%**					**37.89%**				
b: 10-fold cross validation																
		Valence dimension					**Arousal dimension**					**Dominance dimension**				
		Predicted					**Predicted**					**Predicted**				
		1	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
	1	**51**	31	26	34	17	**5**	0	1	1	0	**16**	0	1	1	0
	2	21	**53**	30	36	16	7	**62**	25	25	10	5	**59**	28	27	10
Actual	**3**	39	61	**115**	69	39	11	46	**105**	71	37	14	50	**106**	71	30
	4	68	53	63	**113**	26	19	83	112	**213**	43	22	98	97	**205**	43
	5	10	7	5	8	**44**	12	21	48	34	**44**	13	21	49	30	**39**
Accuracy		**36.32%**					**41.44%**					**41.06%**				

Table 5.8: Precision, recall and F1-score of CNN without data augmentation.

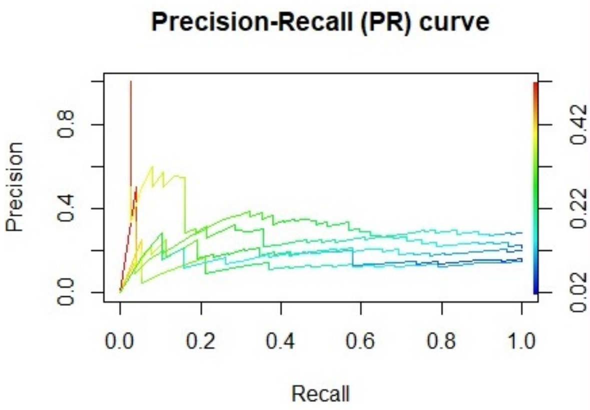
	Valence					**Arousal**					**Dominance**				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Precision	0.170	0.222	0.259	0.216	0.277	/	0.400	0.333	0.333	0.285	/	0.416	0.172	0.250	0.250
Recall	0.233	0.200	0.233	0.266	0.166	0	0.600	0.600	0.450	0.100	0	0.384	0.384	0.307	0.153
F1-score	0.197	0.210	0.245	0.238	0.208	/	0.480	0.428	0.383	0.148	/	0.400	0.238	0.275	0.190
Accuracy	**22.00%**					**35.00%**					**24.62%**				

Table 5.9: Precision, recall and F1-score of CNN with data augmentation.

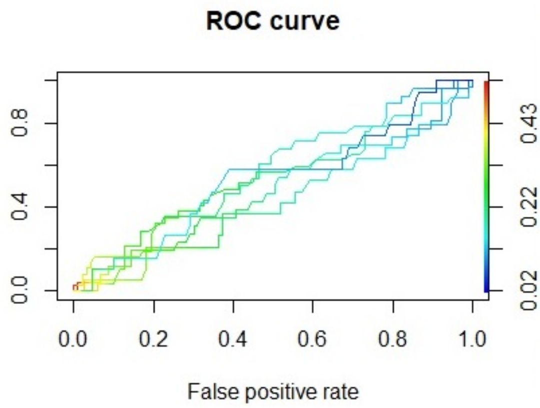
	Valence					**Arousal**					**Dominance**				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Precision	0.979	0.946	0.965	0.889	0.984	0.994	0.757	0.784	0.881	0.993	/	0.870	0.587	0.812	0.929
Recall	0.964	0.960	0.964	0.968	0.902	0.390	0.974	0.980	0.978	0.956	0	0.960	0.974	0.968	0.974
F1-score	0.971	0.953	0.965	0.927	0.941	0.560	0.852	0.871	0.927	0.974	/	0.913	0.733	0.884	0.951
Accuracy	**95.16%**					**85.56%**					**77.54%**				

Table 5.10: The results of K-Fold Cross-Validation of CNN (With and without data augmentation).

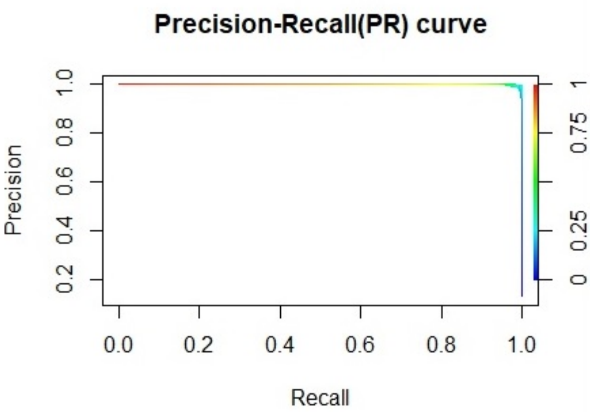
	Without data augmentation			**With data augmentation**		
	Valence	Arousal	Dominance	Valence	Arousal	Dominance
MAE	1.439	1.261	1.119	0.443	0.413	0.449
RMSE	1.893	1.640	1.535	0.985	0.885	0.934
R-squared	0.001	0.003	0.001	0.519	0.439	0.392



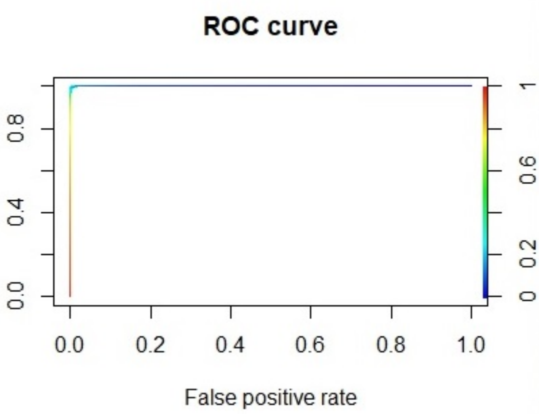
(a)



(b)

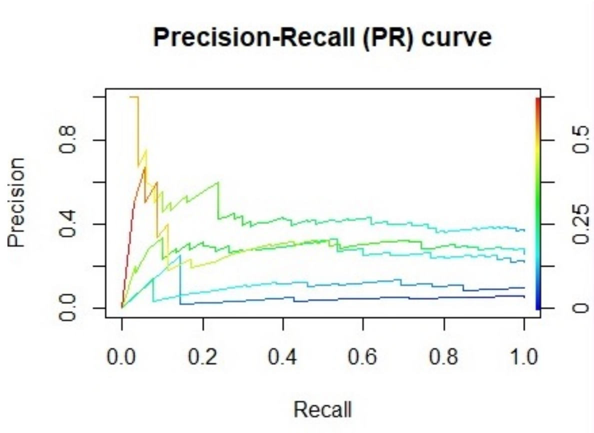


(c)

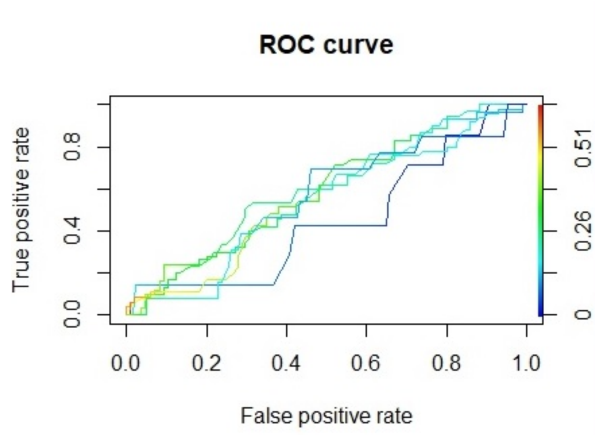


(d)

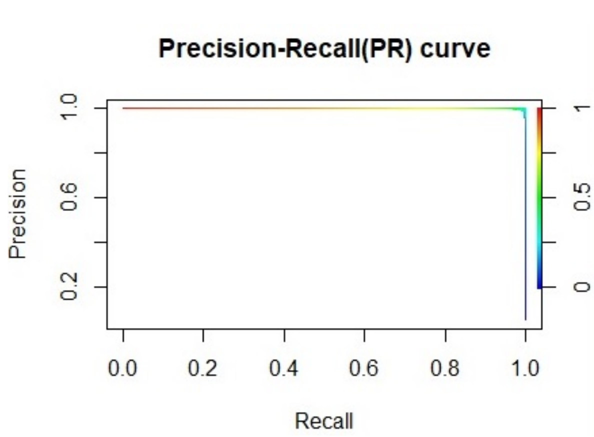
Figure 5.11: Valence dimension: (a: PR curve, b: ROC curve [without data augmentation]), (c: PR curve, d: ROC curve [with data augmentation])



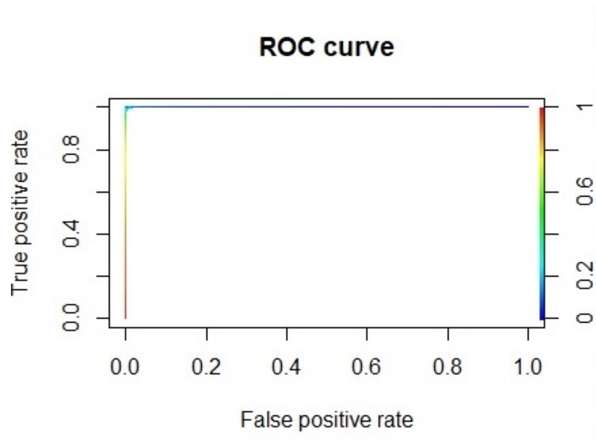
(a)



(b)

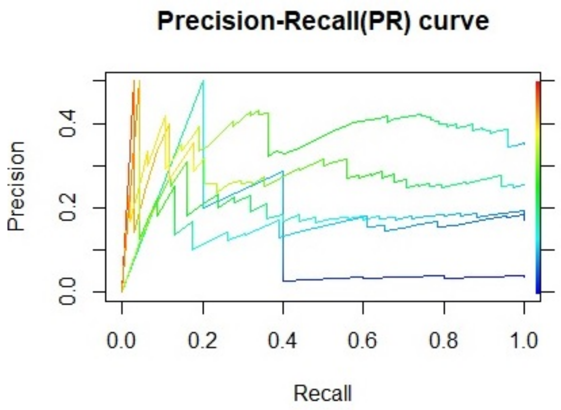


(c)

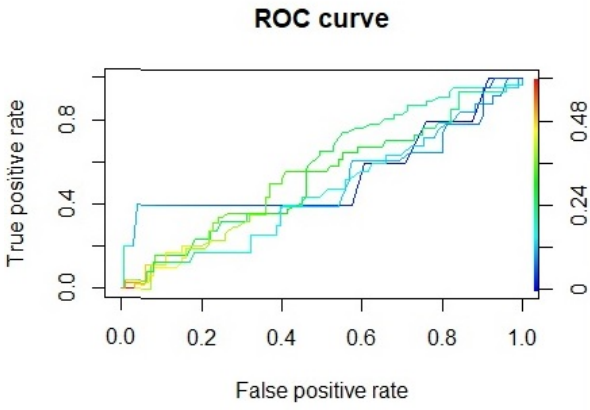


(d)

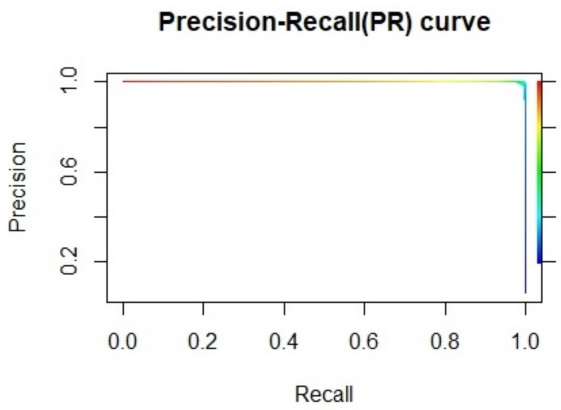
Figure 5.12: Arousal dimension: (a: PR curve, b: ROC curve [without data augmentation]), (c: PR curve, d: ROC curve [with data augmentation])



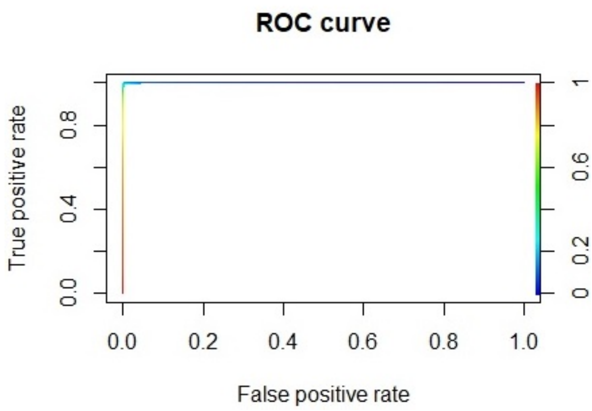
(a)



(b)



(c)



(d)

Figure 5.13: Dominance dimension: (a: PR curve, b: ROC curve [without data augmentation]), (c: PR curve, d: ROC curve [with data augmentation])

5.5 Conclusion

Detecting human emotion using ECG signal is an important research domain with promising application future. However, the existing methods to detect emotion using ECG are not able to identify accurately the emotion of a human. Consequently, it is very important to make this process automatic and to predict human emotion more precisely. Moreover, training CNN to be able to perform the task of detecting emotion using ECG signal needs a huge amount of ECG data. Furthermore, it is very difficult to provide this large number of data due to the sensitive nature of this medical data. To deal with this problem, we proposed, a novel ECG data augmentation strategy, this strategy is capable of generating artificial data in a way that resembles existing ones. Then, we have developed a predictive model using an CNN as a classifier compared to other machine learning models. Our proposal extracts RR intervals and randomly concatenates them from ECG episode intervals to form new samples of ECG signals. Using the DREAMER database, the emotions are expressed in terms of valence, arousal and dominance. Additionally, the HRV features were extracted from RR interval series time. As shown in the experimental results, we came to the conclusion that the performances of all the classifiers used are affected and improved differently from each other by the proposed method of ECG data augmentation.

The CNN classifier achieved the best performance in terms of valence, arousal and dominance with a recognition accuracy rates of 95.16%, 85.56% and 77.54%, respectively.

Conclusions and perspectives

1 Summary of contributions

The heart is one of the most critical organs in the human body, and electrocardiography (ECG) is considered to be one of the most powerful diagnostic tools in medicine that is routinely used for the assessment of the functionality of the heart. ECG being a physiological signal is used as the conventional method for noninvasive interpretation of the electrical activity of the heart in real time. But is usefulness not only in analyzing the heart's activity it can be also used for emotion recognition. Emotion recognition based on ECG signals has been a hot topic and applied in many areas such as safe driving, health care and social security. The main goal of this dissertation was to develop a new monitoring system for emotion recognition using ECG signals. To achieve this, there was the need to explore the fundamentals behind the electrocardiogram and emotions, as well as understanding their natural functioning and behaviour. After that, we presented a large number of related works of emotion recognition using ECG and EEG signals as well as the different methods of ECG data augmentation.

After that, we presented our first contribution, in which we have proposed an enhanced Random forest algorithm for ECG beat classification combined with a simulated annealing algorithm to ensure a complete automatic and optimal process of the ECG classification. ERF was used to distinguish the normal class from the classes of medical disorders and from an abnormal class which may represent an unusual emotional state. The obtained results are very hopeful in this health domain and inspire us to extend this study to other biological signals such as Electroencephalography EEG and other medical applications in order to save human lives.

Since the ECG is a reliable and effective source of information for human emotion recognition systems and has considerable potential for recognizing, and predicting human emotions such as anger, joy, trust, sadness, anticipation and surprise. More specifically, to detect these emotions, the Heart Rate Variability (HRV) values, extracted from ECG are required. In fact, HRV analysis is defined as a simple noninvasive and effective metric, reflecting the activity of sympathetic and parasympathetic components of the ANS on the sinoatrial node (known also as a sinus node -SA-), located in the wall of the right atrium of the heart. Therefore, HRV helps to differentiate among multiple emotions such as neutrality, happiness, disgust, fear, sadness and anger.

In this second proposition, we applied the Enhanced Random forest (ERF) to establish emotion recognition model for driver while driving in road. The proposal system could help to detect and diagnostic stress level and alert the driver, its family and the other road users to avoid accidents caused by high stress state. However, these techniques have been applied on a limited size of datasets. This limited size of dataset is due to the high cost of these sensitive data collection.

To deal with this issue, we proposed in this third contribution a new data augmentation convolutional neural network for human emotion recognition based on HRV, extracted from ECG signal. Notice that the proposed method increases and diversifies the considered samples and ensures a balanced number of samples in each ECG category or class. It worth noting that the existing studies regarding data augmentation for ECG signal were dedicated to detect heart disease like atrial fibrillation (AF) and not for emotion detection.

2 Perspectives and future work

Based on the promising results of this work, the study can be further extended with a variety of research perspectives.

- First, ERF proposed in the third chapter could be applied for other biological signals such as Electroencephalography EEG and other medical applications in order to save human lives.
- Our proposed system in the chapter four, could be enhanced to check the reliability of the returned results in order to avoid the false detection of driver stress state
- We suggest to conduct a new study aiming at conceiving an embedded system that could be integrated in different tiny devices and machines like smart watches, smartphones, on-board computer, etc. This could help to make concrete this proposal as a real-world application. To do that, some real-time and operating concerns could be tackled like respecting time constrain and task scheduling to ensure an integrated functioning of this system within a complicate computational device.

Bibliography

[1] Silvia DE NADAI, Massimo D'INCÀ, and Francesco PARODI. Enhancing safety of transport by road by on-line monitoring of driver emotions. In *In : 2016 11th System of Systems Engineering Conference (SoSE)*, pages 1–4. IEEE, 2016. 1

[2] Rui GUO, Shuangjiang LI, and Li. HE. Pervasive and unobtrusive emotion sensing for human mental health. In *In : 2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, pages 436–439. IEEE, 2013. 1

[3] Bruno VERSCHUERE, Geert CROMBEZ, and Ernst KOSTER. Psychopathy and physiological detection of concealed information: A review. *Psychologica Belgica*, 46(1-2):99–116., 1927. 1

[4] Yu-Dong ZHANG, Zhang-Jing YANG, and Hui-Min LU. Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access*, 4:8375–8385., 2016. 1

[5] Qirong MAO, Ming DONG, and Zhengwei HUANG. Learning salient features for speech emotion recognition using convolutional neural networks. *IEEE transactions on multimedia,*, 16(8):2203–2213., 2014. 1

[6] Walter B CANNON. The james-lange theory of emotions: A critical examination and an alternative theory. *The American journal of psychology*, 39(1/4):106–124, 1927. 2

[7] Andrius DZEDZICKIS, Artūras KAKLAUSKAS, and Vytautas BUCINSKAS. Human emotion recognition: Review of sensors and methods. *Sensors,*, 20(3):592, 2020. 6

[8] María MALFAZ and Miguel A SALICHS. A new architecture for autonomous robots based on emotions. *IFAC Proceedings Volumes*, 37(8):805–809, 2004. 6

[9] Stefan SOSNOWSKI, Ansgar BITTERMANN, and Kolja KUHNLENZ. Design and evaluation of emotion-display eddie. In *In : 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems.*, pages 3113–3118. IEEE, 2006. 6

[10] Mohammad Ali DELKHOON and Fereshteh LOTFIZADEH. An investigation on the effect of gender on emotional responses and purchasing intention due to advertisements. *Journal of Social Sciences and Humanities Research,*, 2(1):6–11, 2014. 6

[11] Stefano SCOTTI, Maurizio MAURI, and Riccardo BARBIERI. Automatic quantitative evaluation of emotions in e-learning applications. In *In : 2006 International Conference of the IEEE Engineering in Medicine and Biology Society.*, pages 1359–1362. IEEE, 2006. 7

[12] Beverly WOOLF, Toby DRAGON, and Ivon ARROYO. Recognizing and responding to student affect. In *In : International conference on human-computer interaction.*, pages 713–722. Springer, Berlin, Heidelberg,, 2009. 7

[13] Georgios N. YANNAKAKIS and John HALLAM. Real-time game adaptation for optimizing player satisfaction. *IEEE Transactions on Computational Intelligence and AI in Games*, 1(2):121–133, 2009. 7

[14] Moataz EL AYADI, Mohamed S. KAMEL, and Fakhri KARRAY. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern recognition,*, 44(3):572–587, 2011. 7, 8, 18

[15] Stamos KATSIGIANNIS and Naeem RAMZAN. Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. *IEEE journal of biomedical and health informatics*, 22(1):98–107, 2017. 7, 33, 35, 65

[16] Patrícia BOTA, Chen WANG, and Ana FRED. Emotion assessment using feature fusion and decision fusion classification based on physiological data: Are we there yet?. *Sensors*, 20(17):4723, 2020. 7

[17] Foteini AGRAFIOTI. *ECG in biometric recognition: Time dependency and application challenges*. PhD thesis, University of Toronto, 2011. 7, 12

[18] C. GODIN, F. PROST-BOUCLE, and A. CAMPAGNE. Selection of the most relevant physiological features for classifying emotion. *Emotion*, 40:20, 2015. 7, 21

[19] Boris REUDERINK, Christian MÜHL, and Mannes. POEL. Valence, arousal and dominance in the eeg during game play. *International journal of autonomous and adaptive communications systems*, 6(1):45–62, 2013. 7

[20] Juan Abdon MIRANDA-CORREA, Mojtaba Khomami ABADI, and Nicu SEBE. Amigos: A dataset for affect, personality and mood research on individuals and groups. *IEEE Transactions on Affective Computing*, 12(2):479–493, 2018. 7, 33, 34

[21] Mohammad SOLEYMANI, Jeroen LICHTENAUER, and Thierry PUN. A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing*, 3(1):42–55, 2011.

[22] Mojtaba Khomami ABADI, Ramanathan SUBRAMANIAN, and Seyed Mostafa KIA. Decaf: Meg-based multimodal database for decoding affective physiological responses. *IEEE Transactions on Affective Computing,*, 6(3):209–222, 2015.

[23] Sander KOELSTRA, Christian MUHL, and Mohammad SOLEYMANI. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31., 2011.

[24] Saskia KOLDIJK, Maya SAPPELLI, and Suzan VERBERNE. The swell knowledge work dataset for stress and user modeling research. In *In : Proceedings of the 16th international conference on multimodal interaction.*, pages 291–298. ACM Digital Library,, 2014. 7

[25] Panagiotis TZIRAKIS, George TRIGEORGIS, and Mihalis A NICOLAOU. End-to-end multimodal emotion recognition using deep neural networks. *IEEE Journal of Selected Topics in Signal Processing*, 11(8):1301–1309, 2017. 8, 21

[26] Nickolaos FRAGOPANAGOS and John G TAYLOR. Emotion recognition in human–computer interaction. *Neural Networks,*, 18(4):389–405, 2005. 8

[27] Paweł TARNOWSKI, Marcin KOŁODZIEJ, and Andrzej MAJKOWSKI. Emotion recognition using facial expressions. *Procedia Computer Science*, 108:1175–1184, 2017.

[28] Kathrin KAULARD, Douglas W CUNNINGHAM, and Heinrich H BÜLTHOFF. The mpi facial expression database—a validated database of emotional and conversational facial expressions. *PloS one*, 7(3):e32321, 2012. 8

[29] Foteini AGRAFIOTI, Dimitris HATZINAKOS, and Adam K ANDERSON. Ecg pattern analysis for emotion detection. *IEEE Transactions on affective computing*, 3(1):102–115., 2011. 8, 9, 26

[30] Somchanok TIVATANSAKUL and Michiko. OHKURA. Emotion recognition using ecg signals with local pattern description methods. *International Journal of Affective Engineering*, pages IJAE–D–15–00036., 2015. 8, 16

[31] Gil KEREN, Tobias KIRSCHSTEIN, and Erik MARCHI. End-to-end learning for dimensional emotion recognition from physiological signals. In *In : 2017 IEEE International Conference on Multimedia and Expo (ICME).*, pages 985–990. IEEE, 2017. 8

[32] Atefeh GOSHVARPOUR, Ataollah ABBASI, and Ateke GOSHVARPOUR. An accurate emotion recognition system using ecg and gsr signals and matching pursuit method. *Biomedical journal,*, 40(6):355–368, 2017. 8, 15, 16

[33] Yu-Liang HSU, Jeen-Shing WANG, and Wei-Chun CHIANG. Automatic ecg-based emotion recognition in music listening. *IEEE Transactions on Affective Computing*, 11(1):85–99, 2017. 8, 21, 32, 35

[34] Nazmi Sofian SUHAIMI, James MOUNTSTEPHENS, and Jason TEO. Eeg-based emotion recognition: a state-of-the-art review of current trends and opportunities. *Computational intelligence and neuroscience,*, 2020, 2020. 8, 16

[35] Siddharth SIDDHARTH, Tzyy-Ping JUNG, and Terrence J SEJNOWSKI. Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing. *IEEE Transactions on Affective Computing,*, 2019. 8, 16

[36] Hany FERDINANDO, Tapio SEPPÄNEN, and Esko ALASAARELA. Comparing features from ecg pattern and hrv analysis for emotion recognition system. In *In : 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB).*, pages 1–6. IEEE, 2016. 8, 16, 25

[37] Catarina C KORDSACHIA, Izelle LABUSCHAGNE, and Sophie C ANDREWS. Diminished facial emg responses to disgusting scenes and happy and fearful faces in huntington's disease. *Cortex,*, 106:185–199., 2018. 8, 16

[38] Lin SHU, Jinyan XIE, and Mingyue YANG. A review of emotion recognition using physiological signals. *Sensors,*, 18(7):2074., 2018. 8, 10, 12, 13, 14, 22, 29

[39] João Ribeiro PINTO, Jaime S CARDOSO, and André LOURENÇO. Evolution, current challenges, and future possibilities in ecg biometrics. *IEEE Access*, 6:34746–34776, 2018. 9, 10, 23

[40] Jing CHEN, Bin HU, and WANG. Subject-independent emotion recognition based on physiological signals: a three-stage decision method. *BMC medical informatics and decision making,*, 17(3):45–57., 2017. 9, 11

[41] Emotion | definition of emotion by oxford dictionary on lexico.com also meaning of emotion. `https://www.lexico.com/definition/emotion.`, 2021. Accessed 02/02/2021. 10

[42] Christine L LISETTI. Affective computing. *Springer,*, pages 71–73., 1998. 10, 11

[43] M STRAUSS, C REYNOLDS, and S HUGHES. Affective computing and intelligent interaction. In *In First International Conference, ACII*, volume 3784, pages 699–706, 2005. 11, 16

[44] Jasleen KAUR and Jatinderkumar R SAINI. Emotion detection and sentiment analysis in text corpus: a differential study with informal and formal writing styles. *International Journal of Computer Application, ISSN*, pages 0975–8887, 2014. 11

[45] C. Darwin, J. Murray, and A. Street. Expression of the emotions in man and animals. *John Murray, London, Albemarle Street,*, 1872. 12

[46] Paul EKMAN. An argument for basic emotions. *Cognition emotion*, 6(3-4):169–200, 1992. 12

[47] Robert PLUTCHIK. A psychoevolutionary theory of emotions. *Lippincott Williams Wilkins,*, 1982. 12

[48] Plutchik's wheel of emotion - system concepts. .https://www.system-concepts.com/insights/emotion-ai-part-1/plutchiks-wheel-of-emotion/, 2021. Accessed on 01/11/2021). 13

[49] Carroll E IZARD. Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on psychological science*, 2(3):260–280, 2007. 12

[50] James A RUSSELL, Anna WEISS, and Gerald A MENDELSOHN. Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology*, 57(3):493, 1989. 12

[51] Peter J LANG. The emotion probe: Studies of motivation and attention. *American psychologist*, 50(5):372, 1995. 12

[52] Albert MEHRABIAN. Comparison of the pad and panas as models for describing emotions and for differentiating anxiety from depression. *Journal of psychopathology and behavioral assessment*, 19(4):331–357, 1997. 13

[53] Gyanendra K VERMA and Uma Shanker TIWARY. Affect representation and recognition in 3d continuous valence–arousal–dominance space. *Multimedia Tools and Applications*, 76(2):2159–2183, 2017. 15

[54] Amy Beth WARRINER, Victor KUPERMAN, and Marc BRYSBAERT. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior research methods*, 45(4):1191–1207., 2013. 13

[55] Roland KEHREIN. The prosody of authentic emotions. In *In : Speech Prosody 2002, International Conference*, 2002. 14

[56] Frank H WILHELM, Julina A RATTEL, and Melanie WEGERER. Attend or defend? sex differences in behavioral, autonomic, and respiratory response patterns to emotion–eliciting films. *Biological psychology*, 130:30–40, 2017. 15

[57] Elizabeth RENDON-VELEZ, P. M. VAN LEEUWEN, and Riender HAPPEE. The effects of time pressure on driver performance and physiological activity: A driving simulator study. *Transportation research part F: traffic psychology and behaviour,*, 41:150–169, 2016. 15

[58] Megan Patricia BARNARD and Peter CHAPMAN. The effects of instruction and environmental demand on state anxiety, driving performance and autonomic activity: Are ego-threatening manipulations effective? *Transportation research part F: traffic psychology and behaviour,*, pages 123–135, 2018. 15

[59] Bridie SCOTT-PARKER. Emotions, behaviour, and the adolescent driver: A literature review. *Transportation research part F: traffic psychology and behaviour*, 50:1–37, 2017. 15

[60] Krystyna RYMARCZYK, Cezary BIELE, and Anna GRABOWSKA. Emg activity in response to static and dynamic facial expressions. *Transportation research part F: traffic psychology and behaviour*, 79(2):330–333, 2011. 15

[61] Bryan REIMER, Bruce MEHLER, and Joseph F COUGHLIN. The impact of a naturalistic hands-free cellular phone task on heart rate and simulated driving performance in two age groups. *Transportation research part F: traffic psychology and behaviour,*, 14(1):13–25, 2011. 15

[62] Sihem NITA, Salim BITAM, and Mellouk Abdelhamid. A body area network for ubiquitous driver stress monitoring based on ecg signal. In *In : 2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob).*, pages 1–6. IEEE, 2019. 15, 37

[63] Khairun Nisa MINHAD, Sawal Hamid Md ALI, and Mamun Bin Ibne REAZ. Happy-anger emotions classifications from electrocardiogram signal for automobile driving safety and awareness. *Journal of Transport Health,*, 7:75–89., 2017. 15

[64] Sihem NITA, Salim BITAM, and Mellouk Abdelhamid. An enhanced random forest for cardiac diseases identification based on ecg signal. In *2018 14th International Wireless Communications Mobile Computing Conference (IWCMC).IEEE*, pages 1339–1344, 2018. 15, 58

[65] Wei-Long ZHENG, Jia-Yi ZHU, and Bao-Liang LU. Identifying stable patterns over time for emotion recognition from eeg. *IEEE Transactions on Affective Computing,*, 10(3):417–429.., 2017. 16

[66] Guo XIANHAI. Study of emotion recognition based on electrocardiogram and rbf neural network. *Procedia Engineering,*, 15:2408–2412., 2011. 16

[67] Mingyang LIU, Di FAN, and Xiaohan ZHANG. Human emotion recognition based on galvanic skin response signal feature selection and svm. In *In : 2016 international conference on smart city and systems engineering (ICSCSE)*, volume 15, pages 157–160. IEEE, 2016. 16

[68] Zhong YIN, Mengyuan ZHAO, and Yongxiong WANG. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer methods and programs in biomedicine,*, 140:93–110., 2017. 16, 35

[69] Bo CHENG and Guangyuan LIU. Emotion recognition from surface emg signal using wavelet transform and neural network. In *In : Proceedings of the 2nd international conference on bioinformatics and biomedical engineering (ICBBE)*, pages 1363–1366, 2008. 16

[70] MinSeop LEE, Yun Kyu LEE, and Myo-Taeg LIM. Emotion recognition using convolutional neural network with selected statistical photoplethysmogram features. *Applied Sciences,*, 10(10):3501, 2020. 16

[71] Jeff F COHN and Fernando DE LA TORRE. Automated face analysis for affective computing. *Handbook of affective computing*, page 131–150, 2015. 16

[72] Paul EKMAN and Wallace V FRIESEN. Facial action coding system: Investigator's guide. *Consulting Psychologists Press,*, 1978. 17

[73] Shishir BASHYAL and Ganesh K VENAYAGAMOORTHY. Recognition of facial expressions using gabor wavelets and learning vector quantization. *Engineering Applications of Artificial Intelligence,*, 21(7):1056–1064, 2008. 17

[74] Byoung Chul KO. A brief review of facial emotion recognition based on visual information. *sensors,*, 18(2):401, 2018. 17, 18

[75] Joy NICHOLSON, Kazuhiko TAKAHASHI, and Ryohei NAKATSU. Emotion recognition in speech using neural networks. *Neural computing applications*, 9(4):290–296., 2000. 18

[76] Rainer BANSE and Klaus R SCHERER. Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, 70(3):614., 1996. 18

[77] Tin Lay NWE, Say Wei FOO, and Liyanage C DE SILVA. Speech emotion recognition using hidden markov models. *Speech communication,*, 41(4):603–623, 2003. 18

[78] Björn SCHULLER, Gerhard RIGOLL, and Manfred LANG. Hidden markov model-based speech emotion recognition. In *In : 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).*, pages II–1. IEEE, 2003. 18

[79] Vaibhav V KAMBLE, Ratnadeep R DESHMUKH, and Anil R KARWANKAR. Emotion recognition for instantaneous marathi spoken words. In *In : Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*, pages 335–346. Springer, Cham,, 2015. 19

[80] Mehrabian ALBERT. Nonverbal communication. *Routledge: London, Uk,*, 2017. 19

[81] J. F. Iaccino. Left brain-right brain differences: Inquiries, evidence, and new approaches. *Lawrence Erlbaum Associates, Inc,*, 1993. 19

[82] Pavlo MOLCHANOV, Shalini GUPTA, and Kihwan KIM. Hand gesture recognition with 3d convolutional neural networks. In *In : Proceedings of the IEEE conference on computer vision and pattern recognition workshops.*, pages 1–7. IEEE, 2015. 19

[83] Tomasz SAPIŃSKI, Dorota KAMIŃSKA, and Adam PELIKANT. Emotion recognition from skeletal movements. *Entropy,*, 21(7):646., 2019. 19

[84] Fatemeh NOROOZI, Ciprian Adrian CORNEANU, and Dorota KAMIŃSKA. Survey on emotional body gesture recognition. *IEEE transactions on affective computing,*, 12(2):505–523., 2018. 20

[85] Sylvia D KREIBIG. Autonomic nervous system activity in emotion: A review. *Biological psychology*, 84(3):394–421, 2010. 20

[86] Pramila RANI, Changchun LIU, and Nilanjan SARKAR. An empirical study of machine learning techniques for affect recognition in human–robot interaction. *Pattern Analysis and Applications*, 9(1):58–69, 2006. 21

[87] Gert van der VLOED and Jelle BERENTSEN. Measuring emotional wellbeing with a non-intrusive bed sensor. In *In : IFIP Conference on Human-Computer Interaction*, pages 908–911. Springer, Berlin, Heidelberg, 2009. 21

[88] A. J FRIDLUND and Carroll E IZARD. Electromyographic studies of facial expressions of emotions and patterns of emotions. *Social psychophysiology: A sourcebook*, pages 243–286, 1983. 21

[89] Kyung Hwan KIM, Seok Won BANG, and Sang Ryong KIM. Emotion recognition system using short-term monitoring of physiological signals. *Medical and biological engineering and computing*, 42(3):419–427., 2004. 21

[90] Gaetano VALENZA, Luca CITI, and Antonio LANATÁ. Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics. *Scientific reports*, 4(1):1–13, 2014. 21

[91] Ross HARPER and Joshua SOUTHERN. A bayesian deep learning framework for end-to-end prediction of emotion from heartbeat. *IEEE Transactions on Affective Computing*, pages 1–1, 2020. 21

[92] João Tiago Ribeiro PINTO. *Continuous biometric identification on the steering wheel.* PhD thesis, University of Porto, 2017. 22, 25

[93] Philip TATE. Seeley's principles of anatomy and physiology. *McGraw-Hill Higher Education*, 2011. 22, 25

[94] E. N. Marieb and K. Hoehn. Human anatomy physiology. *Pearson, Glenview, IL, 9 edition,*, 2013. 22, 24, 25

[95] Valerie C. SCANLON and Tina SANDERS. Essentials of anatomy and physiology fifth edition. *F. A. Davis Company, Philadelphia, PA, 5 edition,*, 2007. 23

[96] Alan KENNEDY, Dewar D FINLAY, and Daniel GULDENRING. The cardiac conduction system: generation and conduction of the cardiac impulse. *Critical Care Nursing Clinics*, 28(3):269–279, 2016. 23

[97] Michael SAMPSON and Anthony MCGRATH. Understanding the ecg. part 1: Anatomy and physiology. *British Journal of Cardiac Nursing*, 10(11):548–554, 2015. 23, 25

[98] J. G. Betts, P. DeSaix, E. Johnson, O. Korol J. E. Johnson, D. H. Kruse, B. Poe, J. A. Wise, M. Womble, and K. A. Young. Anatomy and physiology, chapter 19. *OpenStax College, Rice University, Houston, Texas, 9 edition*, 2017. 24

[99] A. VAN OOSTEROM, R. HOEKEMA, and G. J. H UIJEN. Geometrical factors affecting the interindividual variability of the ecg and the vcg. *Journal of electrocardiology*, 33:219–227., 2000. 25

[100] Rudi HOEKEMA, Gérard JH UIJEN, and Adriaan VAN OOSTEROM. Geometrical aspects of the interindividual variability of multilead ecg recordings. *IEEE Transactions on Biomedical Engineering*, 48(5):551–559, 2001. 25, 26

[101] Foteini AGRAFIOTI and Dimitrios BUI, Francis M.and HATZINAKOS. Secure telemedicine: Biometrics for remote and continuous patient verification. *Journal of Computer Networks and Communications*, 2012, 2012. 25

[102] B. J. A. Schijvenaars. *Intra-individual Variability of the Electrocardiogram computerized EEG analysis*. PhD thesis, University of Rotterdam, 2000. 25, 26

[103] Foteini AGRAFIOTI, Jiexin GAO, and Dimitrios HATZINAKOS. Heart biometrics: Theory, methods and applications. *Biometrics*, 3:199–216, 2011. 25, 26, 27

[104] J. Catalano. Guide to ecg analysis. *ISBN 9780781729307*, 1993. 26

[105] I. Michael REVINA and WR Sam EMMANUEL. A survey on human face expression recognition techniques. *Saud University-Computer and Information Sciences*, 33(6):619–628., 2021. 29

[106] Noman NASEER and Hammad NAZEER. Classification of normal and abnormal ecg signals based on their pqrst intervals. In *In : 2017 International Conference on Mechanical, System and Control Engineering (ICMSC).*, pages 388–391. IEEE, 2017. 29

[107] Mayank Kumar GAUTAM and Vinod Kumar GIRI. An approach of neural network for electrocardiogram classification. *APTIKOM Journal on Computer Science and Information Technologies*, 1(3):119–127., 2016. 30, 42

[108] S. Udhaya KUMAR and H. Hannah INBARANI. Neighborhood rough set based ecg signal classification for diagnosis of cardiac diseases. *Soft Computing*, 21(16):4721–4733, 2017. 30

[109] IS Siva RAO and T. Srinivasa RAO. Performance identification of different heart diseases based on neural network classification. *Neural Computing and Applications*, 11(6):3859–3864., 2016. 30, 42

[110] Tahmida TABASSUM and Monira ISLAM. An approach of cardiac disease prediction by analyzing ecg signal. In *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT).*, pages 1–5. IEEE, 2016. 30, 42

[111] Zerina MASETIC and Abdulhamit SUBASI. Congestive heart failure detection using random forest classifier. *Computer methods and programs in biomedicine*, 130:54–64, 2016. 30

[112] João Tiago Ribeiro PINTO. Continuous biometric identification on the steering wheel. *Lippincott Williams Wilkins,*, 2002. 31

[113] Martina S MÜLLER, Alexei L. VYSSOTSKI, and Maki YAMAMOTO. Heart rate variability reveals that a decrease in parasympathetic ('rest-and-digest') activity dominates autonomic stress responses in a free-living seabird. *Comparative Biochemistry and Physiology Part A: Molecular Integrative Physiology*, 212:117–126, 2017. 31

[114] Pritam SARKAR and Ali ETEMAD. Self-supervised learning for ecg-based emotion recognition. In *In : ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).*, pages 3217–3221. IEEE, 2020. 32, 33, 35

[115] Adam K ANDERSON. Ecg pattern analysis for emotion detection. *IEEE TRANSACTIONS ON AFFECTIVE COMPUTING*, 3(1):117–126, 2012. 32

[116] Jonghwa KIM and Elisabeth ANDRÉ. Emotion recognition based on physiological changes in music listening. *IEEE transactions on pattern analysis and machine intelligence,*, 30(12):2067–2083, 2008. 32

[117] Mimma NARDELLI, Gaetano VALENZA, and Alberto GRECO. Recognizing emotions induced by affective sounds through heart rate variability. *IEEE Transactions on Affective Computing,*, 6(4):385–394, 2015. 33

[118] Pritam SARKAR and Ali ETEMAD. Self-supervised ecg representation learning for emotion recognition. *IEEE Transactions on Affective Computing*, 2020. 33, 35

[119] Luz SANTAMARIA-GRANADOS, Mario MUNOZ-ORGANERO, and Gustavo RAMIREZ-GONZALEZ. Using deep convolutional neural network for emotion detection on a physiological signals dataset (amigos). *IEEE Access*, 7:57–67, 2018. 34

[120] Abdul Rehman ASLAM and Muhammad Awais Bin ALTAF. An 8 channel patient specific neuromorphic processor for the early screening of autistic children through emotion detection. In *In : 2019 IEEE International Symposium on Circuits and Systems (ISCAS).*, pages 1–5. IEEE, 2019. 34

[121] Abdul Rehman ASLAM and Muhammad Awais Bin ALTAF. An on-chip processor for chronic neurological disorders assistance using negative affectivity classification. *IEEE Transactions on Biomedical Circuits and Systems*, 14(4):838–851, 2020. 34

[122] Fang WANG, Sheng-hua ZHONG, and Jianfeng PENG. Data augmentation for eeg-based emotion recognition with deep convolutional neural networks. In *In : International conference on multimedia modeling.*, pages 82–93. Springer, Cham, 2018. 34

[123] Siao Zheng BONG, Khairunizam WAN, and M MURUGAPPAN. Implementation of wavelet packet transform and non linear analysis for emotion classification in stroke patient using brain signals. *Biomedical signal processing and control,*, 36:102–112, 2017. 35

[124] Yang WEI, Yue WU, and TUDOR. A real-time wearable emotion detection headband based on eeg measurement. *Sensors and Actuators A: Physical*, 263:614–621., 2017. 35

[125] Peixiang ZHONG, Di WANG, and Chunyan MIAO. Eeg-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*, 2020. 35

[126] H LENG, Y. LIN, and L. A ZANZI. An experimental study on physiological parameters toward driver emotion recognition. In *In : International Conference on Ergonomics and Health Aspects of Work with Computers.*, pages 237–246. Springer, Berlin, Heidelberg, 2007. 35

[127] Feng-Tso SUN, Cynthia KUO, and Heng-Tze CHENG. Activity-aware mental stress detection using physiological sensors. In *In : International conference on Mobile computing, applications, and services.*, pages 282–301. Springer, Berlin, Heidelberg, 2010. 36

[128] Dhananjay KUMAR, Kumar Gaurav SHANKAR, and Anshul SARVATE. Stress recognition in physiological signals from drivers. *International Journal of Advanced Research in Computer Science*, 3(6), 2012. 36

[129] N. KESHAN, P. V PARIMI, and Isabelle BICHINDARITZ. Machine learning for stress detection from ecg signals in automobile drivers. In *In : 2015 IEEE International conference on big data (Big Data).*, pages 2661–2669. IEEE, 2015. 36

[130] Jennifer A HEALEY and Rosalind W PICARD. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on intelligent transportation systems,*, 6(2):156–166, 2005. 36

[131] Isabelle BICHINDARITZ, Cassie BREEN, and Ekaterina COLE. Feature selection and machine learning based multilevel stress detection from ecg signals. In *In : International Conference on Innovation in Medicine and Healthcare.*, pages 202–213. Springer, Cham, 2017. 37, 53

[132] George B. MOODY and Roger G MARK. The impact of the mit-bih arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine,*, 20(3):45–50, 20O1. 37, 48

[133] Qing PAN, Xinyi LI, and Luping FANG. Data augmentation for deep learning-based ecg analysis. In *In : Feature engineering and computational intelligence in ECG monitoring*, pages 91–111. Springer, Singapore, 2020. 37

[134] Zhicheng CUI, Wenlin CHEN, and Yixin CHEN. Multi-scale convolutional neural networks for time series classification. *arXiv preprint arXiv:1603.06995, ,*, 2016. 37

[135] Terry T UM, Franz MJ PFISTER, and Daniel PICHLER. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In *In : Proceedings of the 19th ACM international conference on multimodal interaction.*, pages 216–220, 2017. 37, 38

[136] Ping CAO, Xinyi LI, and Kedong MAO. A novel data augmentation method to enhance deep neural networks for detection of atrial fibrillation. *Biomedical Signal Processing and Control*, 56:101675, 2020. 37, 38

[137] Arthur LE GUENNEC, Simon MALINOWSKI, and Romain TAVENARD. Data augmentation for time series classification using convolutional neural networks. In *In : ECML/PKDD workshop on advanced analytics and learning on temporal data.*, 2016. 37

[138] Tae Joon JUN, Hoang Minh NGUYEN, and Daeyoun KANG. Ecg arrhythmia classification using a 2-d convolutional neural network. *IEEE Access*, 2018. 37

[139] Runnan HE, Yang LIU, and Kuanquan WANG. Automatic cardiac arrhythmia classification using combination of deep residual network and bidirectional lstm. *arXiv preprint arXiv:1804.06812*, 7:102119–102135., 2019. 37

[140] Qihang YAO, Ruxin WANG, and Xiaomao FAN. Multi-class arrhythmia detection from 12-lead varied-length ecg using attention-based time-incremental convolutional neural network. *Information Fusion*, 53:174–182., 2020. 37

[141] U. Rajendra ACHARYA, Shu Lih OH, and Yuki HAGIWARA. A deep convolutional neural network model to classify heartbeats. *Computers in biology and medicine*, 89:389–396., 2017. 38

[142] Shota HARADAL, Hideaki HAYASHI, and Seiichi UCHIDA. Biosignal data augmentation based on generative adversarial networks. In *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC).*, pages 368–371., 2018. 38

[143] Naoki NONAKA and Jun SEITA. Data augmentation for electrocardiogram classification with deep neural network. *arXiv preprint arXiv:2009.04398 ,*, 2020. 38

[144] https://www.cdc.gov/., 2019. Accessed 23.08.2017. 41

[145] David H SPODICK. Normal sinus heart rate: sinus tachycardia and sinus bradycardia redefined. *American heart journal*, 124(4):1119–1121, 1992. 42

[146] Olga MALGINA, Jana MILENKOVIC, and Emil PLESNIK. Ecg signal feature extraction and classification based on r peaks detection in the phase space. In *2011 IEEE GCC Conference and Exhibition (GCC).*, pages 381–384. IEEE, 2011. 42

[147] Leif SÖRNMO and Pablo LAGUNA. Bioelectrical signal processing in cardiac and neurological applications. *Academic Press*, 2005. 42

[148] S KARPAGACHELVI, M ARTHANARI, and M SIVAKUMAR. Classification of electrocardiogram signals with support vector machines and extreme learning machine. *Neural Computing and Applications*, 21(6):1331–1339, 2012. 42

[149] Manpreet KAUR and A. S ARORA. Unsupervised analysis of arrhythmias using k-means clustering. *JCSIT) International Journal of Computer Science and Information Technologies*, 1(5):417–419., 2010. 42

[150] Nahit EMANET. Ecg beat classification by using discrete wavelet transform and random forest algorithm. In *In : 2009 Fifth International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control.*, pages 1–4. IEEE, 2009. 42

[151] R. Ganesh KUMAR and Y KUMARASWAMY. Investigating cardiac arrhythmia in ecg using random forest classification. *International Journal of Computer Applications*, 37(4):31–34., 2012.

[152] Emina ALICKOVIC and Abdulhamit. SUBASI. Medical decision support system for diagnosis of heart arrhythmia using dwt and random forests classifier. *Journal of medical systems*, 40(4):1–12., 2012. 42

[153] Andy LIAW and Matthew WIENER. Classification and regression by random forest. *R news,*, 2(3):18–22., 2002. 43

[154] Ladjel BELLATRECHE and Sharma (ed.) CHAKRAVARTHY. Big data analytics and knowledge discovery. In *19th International Conference, DaWaK 2017, Lyon, France*, page 28–31. Proceedings. Springer,, 2017. 43

[155] https://machinelearningmastery.com/bagging-and-random-forest., 2019. Accessed 01.03.2018. 43

[156] https://en.wikipedia.org., 2019. Accessed 06.12.2017. 43

[157] Ke-Lin DU, SWAMY, and M. N. S. Search and optimization by metaheuristics. *Techniques and Algorithms Inspired by Nature*, pages 1–10, 2016. 43, 44, 58

[158] Ary L. GOLDBERGER, Luis AN AMARAL, and Leon GLASS. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Journal of Educational Technology Society*, 101(23):e215–e220, 2000. 44, 48, 55, 56, 59

[159] https://www.physionet.org., 2019. Accessed 02.03.2019. 44, 55

[160] https://archive.ics.uci.edu/ml/datasets/Heart+Disease., 2017. Accessed 05.11.2017. 44, 48

[161] https://archive.ics.uci.edu/ml/datasets/Arrhythmia., 2017. Accessed 05.11.2017. 44, 48

[162] Hari Mohan RAI, Anurag TRIVEDI, and Shailja SHUKLA. Ecg signal processing for abnormalities detection using multi-resolution wavelet transform and artificial neural network classifier. *Measurement,*, 46(9):3238–3246, 2013. 45, 49, 57, 59

[163] Leo BREIMAN. Random forests. *Machine learning*, 45(1):5–32, 2001. 46, 54

[164] Federico MUSCIOTTO, Luca MAROTTA, and Salvatore MICCICHÈ. Bootstrap validation of links of a minimum spanning tree. *Physica A: Statistical Mechanics and its Applications*, 512:1032–1043, 2018. 46

[165] https://en.wikipedia.org/wiki/RStudio/., 2018. Accessed 02.03.2018. 48

[166] https://physionet.org/physiotools/., 2019. Accessed 02.03.2019. 49, 59

[167] https://www.physionet.org/physiotools/wfdb.shtml., 2017. Accessed 10.11.2017. 49

[168] https://www.codeproject.com/Articles/20995/ECG-Annotation-C?Library., 2017. Accessed 12.12.2017. 49

[169] Yvonne M ULRICH-LAI and James P HERMAN. Neural regulation of endocrine and autonomic stress responses. *Nature reviews neuroscience*, 6:397–409, 2020. 53

[170] Hans SELYE. Selye's guide to stress research. *Van Nostrand Reinhold*, 1980. 53

[171] Jonathan RUBIN, Hoda ELDARDIRY, and Rui ABREU. Towards a mobile and wearable system for predicting panic attacks. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 529–533. ACM digital library, 2015. 54

[172] Arindam RAY and Amlan CHAKRABARTI. Design and implementation of technology enabled affective learning using fusion of bio-physical and facial expression. *Journal of Educational Technology Society*, 19(4):112–125, 2016. 54

[173] Ahmet AKBAS. Evaluation of the physiological data indicating the dynamic stress level of drivers. *Journal of Educational Technology Society*, 6(2):430–439, 2011. 56

[174] Yingchun ZHOU and Nell. SEDRANSK. Functional data analytic approach of modeling ecg t-wave shape to measure cardiovascular behavior. *The Annals of applied statistics*, pages 1382–1402, 2009. 57

[175] Sihem NITA, Salim BITAM, Matthieu HEIDET, and Mellouk Abdelhamid. A new data augmentation convolutional neural network for human emotion recognition based on ecg signals. *Biomedical Signal Processing and Control*, 75:103580, 2022. 64

[176] Hongyu SHI, Licai YANG, and Lulu ZHAO. Differences of heart rate variability between happiness and sadness emotion states: a pilot study. *Journal of Medical and Biological Engineering*, 37(4):527–539, 2017. 64, 69

[177] Muhammad Najam DAR, Muhammad Usman AKRAM, and Sajid Gul KHAWAJA. Cnn and lstm-based emotion charting using physiological signals. *Sensors*, 20(16):4551, 2020. 65

[178] George E BILLMAN. Heart rate variability–a historical perspective. *Frontiers in physiology*, 2:86, 2020. 65

[179] Shiying DONG, Fangfei XU, and Barbara LINGWOOD. R-wave detection: A comparative analysis of four methods using newborn piglet ecg. In *10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010)*, pages 320–323. IEEE, 2010. 65

[180] ouraring. `https://ouraring.com/what-is-heart-rate-variability/.`, 2020. Accessed 09.06.2020. 68

[181] harvard. `https://www.health.harvard.edu/blog/heart-rate-variability.`, 2020. Accessed 19.06.2020. 68

[182] Gaetano VALENZA, Antonio LANATA, and Enzo Pasquale SCILINGO. The role of nonlinear dynamics in affective valence and arousal recognition. *IEEE transactions on affective computing*, 3(2):237–249, 2011. 68

[183] Hye-Geum KIM, Eun-Jin CHEON, and et al. BAI, Dai-Seg. Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry investigation*, 15(3):235, 2018.

[184] U. RAJENDRA ACHARYA, K. PAUL JOSEPH, and Natarajan KANNATHAL. Heart rate variability: a review. *Medical and biological engineering and computing*, 44(12):1031–1051, 2006. 68

[185] Lulu ZHAO, Licai YANG, and Hongyu SHI. Evaluation of consistency of hrv indices change among different emotions. In *2017 Chinese Automation Congress (CAC)*, pages 4783–4786. IEEE, 2017. 68

[186] whoop. `https://www.whoop.com/thelocker/heart-rate-variability-hrv/.`, 2020. Accessed 06.05.2020. 68

[187] elitehrv. `https://help.elitehrv.com/article/159-hrv-frequency/.`, 2020. Accessed 09.06.2020. 69

[188] HRVAS Software. `https://sourceforge.net/projects/hrvas/.`, 2020. Accessed 01.09.2020. 69

[189] John T RAMSHUR. *Design, evaluation, and application of heart rate variability analysis software (HRVAS)*. PhD thesis, Memphis, 2010. 69

[190] J. Sztajzel. Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss medical weekly*, 134(35-36):514–522, 2004. 70

[191] V.Nair and G.E.Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010. 70

[192] A. Bhardwaj and A. Tiwari. Breast cancer diagnosis using genetically optimized neural network model. *Expert Systems with Applications*, 42(10):4611–4620, 2015. 73

[193] D. Acharya, N. Varshney, and A. Vedant. An enhanced fitness function to recognize unbalanced human emotions data. *Expert Systems with Applications*, 166:114011, 2021. 73

[194] C. Xu, A.S Ding, and S.S Liao. A privacy-preserving recommendation method based on multi-objective optimisation for mobile users. *International Journal of Bio-Inspired Computation*, 16(1):23–32, 2020. 73