UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO
Escola de Ciências e Tecnologia

Facial Image Processing to Monitor Physical Exercise Intensity

Doctoral Thesis in Informatics

Salik Ram Khanal

Advisor:

Professor Doutor Vítor Manuel de Jesus Filipe

Co-Advisor:

Professor Doutor João Manuel Pereira Barroso

Professor António Jaime da Eira Sampaio

Vila Real, 2019

UNIVERSIDADE DE TRÁS-OS-MONTES E ALTO DOURO
Escola de Ciências e Tecnologia

Facial Image Processing to Monitor Physical Exercise Intensity

Doctoral Thesis in Informatics

Salik Ram Khanal

Advisor:

Professor Doutor Vítor Manuel de Jesus Filipe

Co-Advisors:

Professor Doutor João Manuel Pereira Barroso

Professor António Jaime da Eira Sampaio

Tese de doutoramento apresentada por Salik Ram Khanal à Universidade de Trás-os-Montes e Alto Douro para o cumprimento dos requisitos necessários à obtenção do grau de Doutor em Informática, sob a orientação do Professor Doutor Vítor Manuel de Jesus Filipe e coorientação do Professor Doutor João Manuel Pereira Barroso e Professor Doutor António Jaime da Eira Sampaio

Vila Real, 2019

RESUMO

Os avanços registados na tecnologia de visão por computador têm possibilitado a sua aplicação em várias áreas, incluindo as ciências do desporto. Nas últimas décadas várias tecnologias foram desenvolvidas para monitorizar a intensidade do exercício físico, no entanto ainda se sente a falta de técnicas confiáveis e não invasivas (contactless). Existem vários parâmetros medíveis para quantificar a intensidade do exercício, sendo a frequência cardíaca uma das mais utilizadas. Contudo o esforço também pode ser revelado pela expressão facial do praticante, sem recorrer a sensores de contato direto.

Nesta tese é proposta uma abordagem baseada em técnicas de visão por computador para quantificar o nível de intensidade do exercício físico, usando caraterísticas faciais como a cor e a expressão. O problema de investigação é abordado usando técnicas de aprendizagem automática: classificação e regressão. No primeiro modelo proposto, a intensidade do exercício é classificada em quatro níveis de intensidade, usando algoritmos tradicionais de aprendizagem automática e uma rede neuronal convolucional. No segundo modelo, a intensidade do exercício é estimada usando modelos de regressão linear e multivariada, em que as variáveis independentes são os valores de intensidade média de cor, calculado num recorte da imagem situada na região da testa.

O conjunto de dados para as experiências foram recolhidos em laboratório com vinte indivíduos familiarizados com os protocolos de cicloergómetro, durante a realização de um exercício submáximo de ciclismo. Os dados de frequência cardíaca foram adquiridos com um sensor de contacto e uma câmara estática fez a captura do vídeo facial. As imagens foram obtidas e pré-processadas a partir do vídeo, de acordo com os parâmetros necessários aos modelos de classificação e de regressão.

O primeiro modelo obteve um grau de precisão até 99% sobre um conjunto de teste obtido aleatoriamente do conjunto inicial. No segundo modelo obteve-se um coeficiente de correlação de Pearson acima de 0.98. Estes resultados sugerem que as abordagens propostas são adequadas para estimar a intensidade do exercício físico.

ABSTRACT

Advances in computer vision technology allow to develop applications is several areas including sports sciences. In the last few decades, various technologies were developed to monitor exercise intensity, however, there are still lacking reliable non-invasive techniques. There are many measurable parameters to quantify exercise intensity, among them, heart rate is one of the most used. However, the physical exercise intensity can be quantified by analysing facial expression without using any contact sensor.

The present thesis proposes an approach based on computer vision techniques to quantify intensity of physical exercise using facial features such as colour and expression. The research problem is formulated and addressed using supervised machine learning models such as classification and regression. In the first model, the classification of physical exercise intensity up to four different levels of exertion has been done using traditional machine learning techniques and Convolutional Neural Network. In the second model, the level of exertion has been predicted using linear univariate and multivariate regression, where the input variables were average colour intensity values extracted from a small forehead patch.

The dataset for the experiments were collected in laboratorial normalized settings from twenty subjects familiarized with a cycloergometer protocols during submaximal cycling exercise. Heart rate data were collected using a contact-sensor device and a frontal stationary camera captured the facial video. The separate image dataset was prepared according to the goal of the experiment (classification and regression problem) from the facial video.

In the first model, we obtained up-to 99% accuracy from a randomly split dataset. In the second model, we obtained the Pearson correlation coefficient more than 0.98. These results reveal that the proposed approaches are reliable to quantify physical exercise intensity.

ACKNOWLEDGEMENT

This thesis was carried out, from 2016 to 2019, at the **School of Science and Technology (ECT), Department of Engineering, University of Trás-os-Montes and Alto Douro, in Vila Real, Portugal** with collaboration from the Department of Sports, Exercise and Health Sciences of the same university.

A very special gratitude goes to two projects**, NanoSTIMA – Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics** belonging to **Institute for Systems and Computer Engineering, Technology and Science (INESC TEC)** and **INDTECH 4.0 - New Technologies for Smart Manufacturing** belonging to **Universidade de Trás-os-Montes e Alto Douro** for helping me financially and technically to carry out my PhD research. This thesis and its contribution to knowledge would not exist if these projects and its collaborators had not believed in this work.

I am deeply grateful **to Professor Doctor Vitor Filipe**, for your dedicated hours, all the opportunities provided, the sharing of knowledge, scientific and personal suggestions, and the encouragement in every activity. To **Professor Doctor João Barroso and Professor Doctor Jaime Sampaio**, I would like to thank you for your every contribution which makes my dream success.

I am also very grateful to all the teaching and research staffs that have taken some time to discuss and enrich my work. To my professors, **Arsénio Reis** and **Hugo Paredes**, a big thanks for all the supports given at various times of this period. I would also like to thank all other professors from the **University of Trás-os-Montes de Alto Douro** with whom I interacted in various ways.

With a special mention to my laboratory colleagues **Hugo Fernandes, Dennis Paulino, André Sousa, André Silva, Nuno Lopes, Ana Teresa, and Prabin Sharma, Juliana Exel** in general. It was fantastic to have opportunity to work the majority of my research in your facilities. What a cracking place to work!

I very special word of thanks goes to my parents, **Bedi Ram Khanal** and **Dim Kumari Khanal** who allowed me to go abroad for my study. To my grandfathers, grandmothers, family members **Sagar Mani Khanal** (Brother), **Pushpawati Khanal**, **Shrijana Khanal**, and **Durga Khanal** (Sisters), special thanks for never-ended love and support. I would also like to remember my parents-in-law, **Narayan Prasad Tiwari** and **Sita Tiwari**, Sister-in-law **Devi Tiwari**, Brothers-in-law **Janak Tiwari** and **Dipesh Tiwari** for your support in various ways.

To my parents and all family members, wife **Tulasi Tiwari Khanal**, for your never-ending love and support which helps me in every aspect of my live. You always play major role in my every success of my life.

The last word goes for **Sahasi Sarathi Tiwari Khanal** my little Bunny, who has been the full bright light of my life for the last two and half years and who has given me the extra strength and motivation to get things done. This thesis is dedicated to you my baby.

Thanks all for your encouragement!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

Table of Acronyms

Acronyms	Descriptions
AI	Artificial Intelligence
ANN	Artificial Neural Networks
AUC	Area Under Curve
BPM	Beat Per Minutes
BVP	Blood Volume Pulse
CCD	Charged Coupled Device
CMOS	Complementary Metal-Oxide Semiconductor
CNN	Convolutional Neural Network
DTFT	Discrete Time Fourier Transform
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
FACS	Facial Action Coding System
FFT	Fast Fourier Transform
FPR	False Positive Rate
HOOF	Histogram of Oriented Optical Flow
HR	Heart Rate
HRV	Heart Rate Variability
HSV	Hue Saturation Value
IBI	Inter-Bit Intervals
ICA	Independent Component Analysis
IR	InfraRed
kNN	k Nearest Neighbor
Lab	Lightness-Blue: Red, Blue: Yellow
LDA	Linear Discriminant Analysis
LOOCV	Leave One Out Cross Validation
LOSO	Leave One Subject Out
MAP	Maximum A Posteriori Probability
MSE	Mean Square Error
OPI	Opto-Physiological Imaging
PAT	Pulse Arrival Time
PCA	Principle Component Analysis

PPG	Photoplethysmography
PSV	Puple Size Variation
PTT	Pulse Transit Time
QDA	Quadratic Discriminant Analysis
RGB	Red Green Blue
ROI	Region of Interest
RPE	Rate of Perceived Exertion
RR	Respiratory Rate
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine
TICS	Tensor Independent Color Space
TPR	True Positive Rate
TT	Telemetry Thermistor
VGG	Visual Geometry Group
vPPG-MA	Video Photoplethysmography and Motion Analysis
VT	Ventilation Threshold
YCbCr	Luminance -Chroma: Blue-Chroma: Red

Table of Abbreviations

e.g.	For Example
et al.	And Other Authors
etc.	And Other More
i.e.	In Other Words

1. INTRODUCTION

This chapter presents an overview of the present doctoral thesis. The motivation for the work, as well as the objectives to be achieved and the contributions to the area will be presented. The list of peer-reviewed publications derived from the thesis are included in this chapter. Finally, the general structure of this thesis is stablished and a brief description of the content of each chapter of this document is given.

1.1. Monitoring Intensity of Physical Exercise

Physical exercise is a key determinant of health status; therefore, organizations worldwide recommend its incorporation as part of one's daily routine. Adequate physical fitness relates to reduced risks of physical and mental disease development (Reed & Pipe, 2016; Schuch et al., 2016) and improvements in health span (Das P Fau - Horton & Horton, 2012). Exercise testing and prescription are fundamental to optimize the dose-response relationship of physical exercise and, therefore, successfully cope with the hypokinetic-related diseases (Bayles & Swank, 2018). Thus, guidelines point out to the need of monitoring the main exercise variables, such as frequency, mode, duration, and intensity, so that it is possible to reach adequate levels of physical activity. This process is particularly improved with the help of technology.

Several technological advances have brought a variety of monitoring techniques to the sport sciences field, turning the physical exercise more intuitive and effective. The primary purpose of exercise monitoring is to measure psychophysiological changes yield by physical exercise, which helps on self/automatic controlling of the effort or load. In the case of self-controlled exercise, the exercise load can be monitored with respect to heart rate changes, where the target heart rate is pre-defined (Hunt & Fankhauser, 2016). Another purpose is to collect information such as not only on the heart rate, but also about body temperature, blood pressure, before, during, and after exercise. Therefore, most of the exercise monitoring systems consist of a tool to measure physiological but not limited to, it can also be monitored without measuring these parameters.

Exercise intensity is a particularly important parameter to exercise prescription, once different levels of intensity lead to specific body adaptations. It is related to how hard a subject feels to carry out the physical exercise (Robertson & Noble, 1997). The intensity is estimated based on physical and physiological gold standard responses as heart rate (Eka Adi Prasetyo Joko, Chia-Chun, Yi-Sheng, Cheng-Hung, & Yuan-Hsiang, 2014; Irani, Nasrollahi, & Moeslund, 2014; Spinsante, Ricciuti, & Scalise, 2018), maximal oxygen consumption (Meyer, Gabriel Hh Fau - Kindermann, & Kindermann, 1999; Scribbans, Vecsey, Hankinson, Foster, & Gurd, 2016), blood lactate (Cerezuela-

Espejo, Courel-Ibáñez, Morán-Navarro, Martínez-Cava, & Pallarés, 2018; Wiles, 2008), and blood pressure (Wiles, 2008). The heart rate is the most popular physiological response used to define and control exercise intensity (Eka Adi Prasetyo Joko, Chia-Chun, Yi-Sheng, Cheng-Hung, & Yuan-Hsiang, 2014; Irani, Nasrollahi, & Moeslund, 2014; Spinsante, Ricciuti, & Scalise, 2018), especially because it can be measured objectively and continuously during exercise than other parameters. Thus, several methods regarding the use of heart rate have been proposed to assign levels of physical exertion.

A common practice to assign the level of exercise exertion is based on the maximum heart rate percentage (Meyer et al., 1999). In general, one's maximum heart rate is calculated by subtracting age from 220. In the process of exercise prescription, specific ranges of heart rate are targeted, depending on the body adaptations aimed (Nikolaidis, Rosemann, & Knechtle, 2018; Sarzynski et al., 2013). In the recent days, heart rate measurement is widespread through both contacts (Hannan et al., 2019; Henriksen et al., 2018; Koehler & Drenowatz, 2017) and contactless sensors (Monkaresi, Calvo, & Yan, 2014; Thomson et al., 2019; Wang, Tao, Burghardt, & Mirmehdi, 2018; Wang, den Brinker, Stuijk, & de Haan, 2017; Xie, Fu, Liang, Hong, & Zhu, 2019).

Although perceived exert (Chen, Chiou, Tzeng, Lu, & Chen, 2017) ion health scales are a contactless validated way of obtaining data on physical exercise intensity, it demands from the tested participants a deep familiarization and understanding of the protocol beforehand (Chen et al., 2017). Thus, a limitation to the use of perceived exertion scales is the degree of subjectivity associated with it (Reed & Pipe, 2016). Besides, there might also be a problem in the generalization of its application, specially to people cognitively impaired.

1.2. Scope of the thesis

There is a recent trend in the use of objective methods for the measurement of physical exercise intensity, which has been facilitated due to the great availability of low-

cost contact sensors and increasingly reliable contactless technology. In line with such trend, the present study investigates various approaches using facial image processing to monitor exercise intensity, as schematically depicted in Figure 1. Facial image processing was applied to classify the exercise intensity where the ground truth value of exercise intensity was based on heart rate data recorded during experimental protocol. The research problems were formulated into two ways: (1) classification problem and (2) regression problem. In the classification problem, the level of physical exercise intensity is classified into four different levels from the facial features extracted. The classification of exercise intensity was carried out using thirteen conventional machine learning algorithms and using deep neural network. The results were validated using K-fold and leave-one-subject-out cross validation. In the regression problem, changes in facial expression and facial color as a function of exercise intensity were analyzed. The color values of a patch on the forehead were tracked throughout the exercise, and the correlation between color value and heart rate was calculated. The color analysis was carried out in various color models (RGB, HSV, YCbCr, Lab, and YUV).

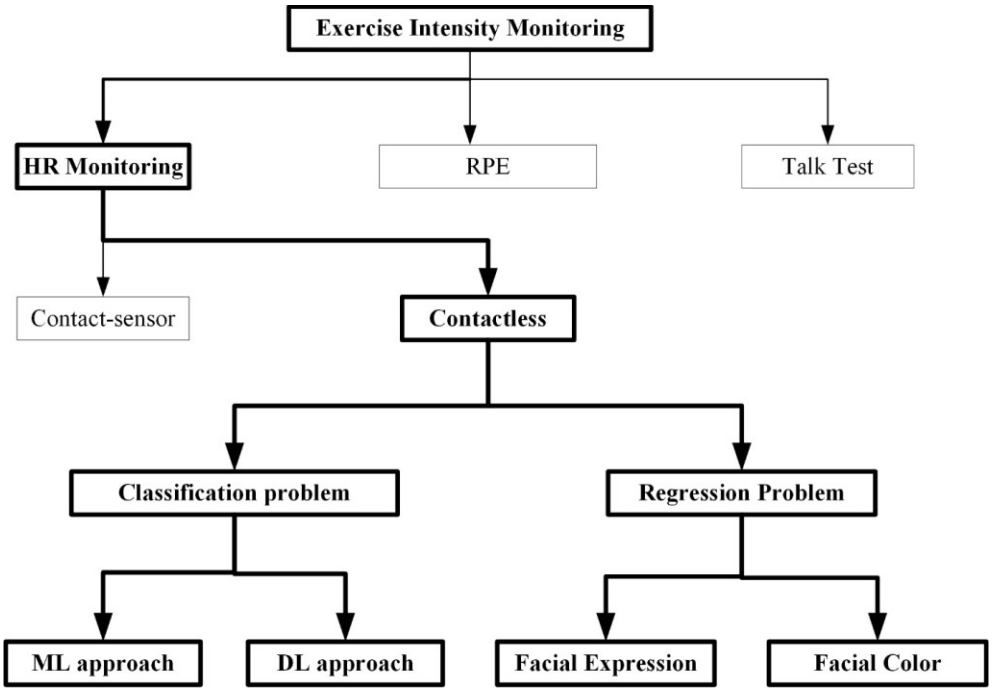


FIGURE 1 - HIERARCHICAL DIAGRAM OF SCOPE OF THESIS.

1.3. Motivations and Objectives

In self-assessment methods for physical exercise monitoring, one of the noticeable limitations is that inexperienced individuals are often unable to pace themselves. For this reason, objective assessments as direct physiological measurements are preferred (Washington, 2004). The recent trends in exercise monitoring relies on objective assessments, whether it is contact-sensor or non-contact technique (Xie et al., 2019). However, although the objective methods might overcome some limitations compared to subjective methods, there are still some significant restraints associated with the measurement of physiological parameters in terms of accuracy. Usually, the prediction of physiological parameters is based on average samples of subjects, which is inefficient to predict the performance of an individual (Aboagye et al., 2017).

The main problems with contact-sensor technology are associated with a possible discomfort that the attachment to the body might yield to the user, and the fact that it often requires a power supply (Seshadri et al., 2019). These limitations might be overcome by

the contactless technology, including computer vision techniques. According to the literature, the prediction of intensity in a particular instance of physical exercise can be performed by extracting and analyzing physiological parameters using both (contact and contactless) types of technology (García-García, Benito, & Hernando, 2016; Montoye, Westgate, Fonley, & Pfeiffer, 2018; Pernek, Kurillo, Stiglic, & Bajcsy, 2015; Seshadri et al., 2019; Um, Babakeshizadeh, & Kulić, 2017). Particularly, it can be measured or monitored continuously throughout the exercise (regression problem) or discretely (classification problem). Facial expression analysis through the monitoring of facial expressions or skin color for quantification of the exercise intensity is one of the candidates to measure psychophysiological parameters (Khanal, Reis, Barroso, & Filipe, 2018; Khanal, Barroso, Sampaio, & Filipe, 2018; Khanal, Sampaio, Barroso, & Filipe, 2019). Therefore, the present thesis proposes the use of discrete and continuous facial color and expression analysis to predict the physical exercise intensity. To the best of our knowledge, literature presents quantification methods for physical exertion, but never proposed the prediction of physical exercise intensity using facial features.

1.4. Research Questions

Towards the end of the present thesis, the following research questions will be discussed and answered:

a) Is it possible to classify human faces to indicate the level of exercise intensity during fatiguing exercise? If so, which machine learning techniques would present the most accurate classification results? [Chapter IV]

b) Is there any correlation between facial color changes and different exercise intensity levels? If so, which color models or color components are correlated to changes in the exercise intensity levels? [Chapter IV and V]

c) Is it possible to quantify the level of exertion by the analysis of facial expression? [Chapter V].

d) Is it possible to obtain a good regression model to predict the level of exertion from increments in heart rate and facial color or expression during exercise? [Chapter V].

1.5. Outline of Thesis

The remaining chapters of the thesis will be presented in the following structure:

Chapter Two: Detailed literatures will be reviewed related to the exercise monitoring using image processing techniques. This chapter is divided into sub-chapters according to the type of physiological features central to each monitoring technique. For instance, it will cover heart rate, body temperature, respiratory rate, blood pressure, calorie consumption, and without measuring any parameters. It will also be presented a comparison among the methods and its major limitations.

Chapter Three: This chapter will present the related theoretical background for the use of supervised machine learning classification algorithms and regression, deep neural network architecture, and evaluation of classification for the quantification of exercise intensity.

Chapter Four: Describes the experiments performed to test the classification of exercise intensity using traditional machine learning approaches and deep learning. It involves the use of digital images to extract facial features using these methods for classification. The experiments and analysis of results are carried out using machine learning algorithms and deep learning architecture separately. The results of each algorithm are validated using leave-one-subject-out and 10-fold cross validation and evaluated using matrices such as confusion matrix and receiver operating characteristic curve.

Chapter Five: It is centered on the experiment for tracking facial color to monitor physical exercise. Different color models are applied and correlated to the heart rate.

Univariate and multivariate regression models are created for individuals and the group of participants tested.

Chapter Six: Presenting of the thesis conclusions and the future directions of the research work.

1.6. Published and Submitted Articles/Papers

1.6.1. Papers published related to thesis work

Khanal, S. R., Barroso, J., Sampaio, J., and Filipe, V. "Classification of physical exercise intensity by using facial expression analysis," *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, 2018, pp. 765-770. doi:10.1109/ICCMC.2018.8488080.

Khanal, S. R., Fonseca, A., Marques, A., Barroso , J., and Filipe, V. "Physical exercise intensity monitoring through eye-blink and mouth's shape analysis," *2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW)*, Thessaloniki, 2018, pp. 1-5. doi: 10.1109/TISHW.2018.8559556.

Khanal S. R., Sampaio J., Barroso J., Filipe V. (2019) Classification of Physical Exercise Intensity Based on Facial Expression Using Deep Neural Network. In: Antona M., Stephanidis C. (eds) Universal Access in Human-Computer Interaction. Multimodality and Assistive Environments. HCII 2019. Lecture Notes in Computer Science, vol. 11573. Springer, Cham.

Khanal, S. R., Sampaio, J., Exel, J., Barroso, J., Filipe, V. (2019). Using Computer Vision to Track Facial Color Changes and Predict Individual's Intensity of Physical Exercise. Manuscript sumbitted for publication.

Khanal, S. R., Sampaio, J., Barroso, J., Filipe, V. (2019). Individual's neutral emotional expression tracking for monitoring physical exercise. *22nd International Conference on Human-Computer Interaction,* Manuscript accepted for publication.

Khanal, S. R., Sampaio, J., Paulino, D., Barroso, J., Filipe, V. (2019). Physicale Exercise Monitoring using computer vision techniques – A Review. Manuscript in preparation.

1.6.2. Publications with Collaboration

Reis, A., Paredes, H., Barroso, I., Monteiro, M. J., Rodrigues, V., Khanal, S. R., Barroso, J., "Autonomous systems to support social activity of elderly people a prospective approach to a system design," *2016 1st International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW)*, Vila Real, 2016, pp. 1-5. doi: 10.1109/TISHW.2016.7847773.

Reis, A., Barroso, I., Monteiro, M. J., V. Filipe, Paredes, H., Rodrigues, V., Khanal, S. R., Barroso, J., "Designing autonomous systems interactions with elderly people," in International Conference on Universal Access in Human-Computer Interaction, pp. 603–611, Springer, Cham, 2017.

Reis, A., Paredes, H., Barroso, I., Monteiro, M. J., Rodrigues, V., Khanal . S. R., and Barroso, J. "A prospective design of a social assistive electronic system for the elderly," Advances in Science, Technology and Engineering Systems Journal (ASTESJ), vol. 2, no. 3, pp. 1665–1669, 2017.

Khanal, S., Reis, A., Barroso, J., and Filipe, V. "Using emotion recognition in intelligent interface design for elderly care," in World Conference on Information Systems and Technologies, pp. 240–247, Springer, Cham, 2018.

Lopes, N., Silva, A., Khanal, S. R., Reis, A. Barroso, J. ,Filipe, V., and Sampaio, J. "Facial emotion recognition in the elderly using an SVM classifier," in 2018

2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW) , pp. 1–5, IEEE, 2018.

Khanal, S. R., Barroso, J., Lopes, N., Sampaio, J., and Filipe, V. "Performance analysis of Microsoft's and Google's emotion recognition API using pose-invariant faces," in 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, pp. 172–178, 2018.

Sharma, P., Eseng ̈on ̈ul, M., Khanal, S. R., Khanal, T. T., Filipe, V., and Reis, M. J.. "Student concentration evaluation index in an e-learning context using facial emotion analysis," in International Conference on Technology and Innovation in Learning, Teaching and Education, pp. 529–538, Springer, Cham, 2018.

Sharma, P., Upadhaya,M. D., Twanabasu, A., Barroso, J., Khanal, S. R., and Paredes, H. "express your feelings": An interactive application for autistic patients," in International Conference on Human-Computer Interaction, pp. 160–171, Springer, Cham, 2019.

2. LITERATURE REVIEW

This chapter presents a literature review that covers the state-of-art related to the current methodological techniques and technologies applied to important issues in sports sciences. This chapter covers topics related to image processing techniques to monitor and assess physical exercise.

2.1. Background

Physical activity is essential to reduce the risks of many diseases, and can improve overall health (Paterson & Warburton, 2010). Currently, many technologies capable of controlling important parameters for physical activity, particularly the ones based on mobile or wearable devices (Allen, Stephens, Dennison Himmelfarb, Stewart, & Hauck, 2013; Glynn et al., 2014; Shin et al., 2017). The use of wearables is a way to facilitate the supervision of physical activities, especially to frail people, as the elderly (Borg, 1962; Faulkner & Eston, 2008; Hasselmann, Oesch, Fernandez-Luque, & Bachmann, 2015; Huanga, Chioua, & Chenb, 2015; Leat, Mei, & Susan, 2008). Intelligent models to predict intensity during the physical exercise provide, through computer vision techniques, the level of exertion while performing the physical activity. The level of exercise intensity determines its effectiveness and controls whether the intensity is in accordance to what has been prescribed.

In the recent years, many studies proposed various methodologies to estimate exercise intensity (Chan et al., 2018; Gade, Larsen, & Moeslund, 2017; Kim, Son, Ko, & Yoon, 2013; Lin & Lin, 2018; Pichierri, Coppe, Lorenzetti, Murer, & de Bruin, 2012; Schoene et al., 2013; Silveira, van het Reve, Daniel, Casati, & De Bruin, 2013) based on physiological variables as heart rate, blood pressure or body temperature, using different type of sensors (Faulkner & Eston, 2008; Huanga et al., 2015; Leat et al., 2008). These sensors can be grouped as a contact (Chan et al., 2018; Pichierri et al., 2012; Schoene et al., 2013; Silveira et al., 2013) and contactless technology (Gade et al., 2017; Kim et al., 2013; Lin & Lin, 2018). The physiological feature extraction could be carried out before, after, and during exercise. In contact sensor techniques, the sensor is attached in the body and the data are interpreted using machine learning or statistical approaches (Dias & Cunha, 2018; Hannan et al., 2019; Seshadri et al., 2019). For contactless sensor techniques, various types of cameras are used to capture video/images which are processed through various techniques (Lopez, del-Blanco, & Garcia, 2017; Sun et al., 2012; Sun et al., 2011). These extraction techniques must be followed by the use of

computer vision and artificial intelligence techniques to interpret the results (Lopez et al., 2017).

In this section, research studies on the topic are presented according to the physiological features and its methods of extraction during physical exercise using computer vision techniques applied on face.

2.2. Heart Rate Extraction

Heart rate (HR) is the rate at which the heart muscle veins perform contraction and expansion to pump and absorb the blood from various parts of the body. It is an important indicator that reflects the level of exertion during exercise (Eka Adi Prasetyo Joko et al., 2014; Wiles, 2008). There are several non-contact techniques to measure HR at rest or during physical exercise with avoiding the motion artifact.

2.2.1. Photoplethysmography

Photoplethysmography (PPG) is a simple, low-cost, and noninvasive optical bio-monitoring technique used for a few decades to monitor the blood volume changes that occur in the human body vessels due to the blood circulation system. This technology consists one light emitter and one light detector. With advances in PPG-based physiological parameter monitoring system, the contact based PPG was extended to a contactless system using images (Li et al., 2019). It is not only applied for heart rate measurement, but also for measuring blood oxygen saturation and respiratory rate (Kumar, Veeraraghavan, & Sabharwal, 2015). The implementation of such system is widespread using various types of image capturing devices as mobile phone cameras, webcams, and thermal cameras (Cui, Fu, Hong, Zhang, & Shu, 2015; Kumar et al., 2015; Sun et al., 2012; Sun et al., 2011). A common way to capture images is using Opto-Physiological Imaging (OPI) with Complementary Metal-Oxide Semiconductor (CMOS) camera and webcam (Sun et al., 2011). Various research directions (scientific camera-

based and webcam-based imaging) have been applied to monitor heart rate and blood pressure during rest, exercise and recovery using PPG. However, only a few studies from the large amount of literature presenting PPG to measure heart rate (Ahmadi Ak Fau - Moradi, Moradi P Fau - Malihi, Malihi M Fau - Karimi, Karimi S Fau - Shamsollahi, & Shamsollahi, 2015; Fallet & Vesin, 2017; Zhang, 2014) focus on HR monitoring during exercise (Bosi, Cogerino, & Bazzani, 2016; Wang et al., 2017).

One of the well-known devices to monitor exercise using PPG technology is the Microsoft Kinect sensor, which can sense the motion using image processing techniques on photoplethysmography images captured by its RGB camera mounted in it. The Kinect RGB camera detects changes of blood flow by mixing PPG signals and volumetric changes of the vessels. For a decade, these devices are heavily used for monitoring parameters during exercise and rest (Bosi et al., 2016; Gambi et al., 2017). Bosi (Bosi et al., 2016) proposed a study using the Microsoft Kinects to detect HR at rest and exercise. From the photoplethysmography images captured by the Microsoft Kinect RGB camera, the average color intensity of each color channel of the RGB region of interest (ROI) selected from the forehead was processed through a bandpass digital filter (Mitra, Li, Lin, & Yu, 1991), Independent Component Analysis (ICA) (Comon, 1994; Hyvärinen & Oja, 2000), and Principle Component Analysis (PCA) (Flury, 1988; Jackson, 1991). The spectral analysis using Fast Fourier Transform (FFT) of the principal components was applied. Results were compared with a pulse oximeter attached in the body simultaneously to the motion sensing device. It was found a good correlation with the oxygen values in the rest, but not for the exercise situation, where participants performed rehabilitation exercises.

Remote PPG have also been proposed to improve heart rate measurement during physical exercise (Wang et al., 2017). The main idea behind this version of PPG was to increase the degrees-of-freedom for noise reduction by decomposing the n-wavelength camera signals into multiple orthogonal frequency bands and extracting the pulse-signal per band. The videos of seven subjects were recorded in various practical challenges such as different skin colors of subjects, various light sources, illumination intensities and exercising modes. The authors suggested that the proposed method increases the

robustness of HR measurement in various fitness applications. Other research presented estimations body surface temperature using Infrared (IR) camera during rest and various stages of exercise (sustained exercise at 50%, 60%, and 70% of age predicted maximum heart rate). HR and Respiratory Rate (RR) were measured with photoplethysmography and motion analysis (vPPG-MA). The results of body surface temperature using IR, and HR and RR using vPPG-MA showed the high level of agreement between the two measurements (Capraro et al., 2018).

2.2.2. Video-based Image Processing

With the rapid advancement in the application of computer vision, researchers have also developed methods for the extraction of reliable heart rate measures (Balakrishnan, Durand, & Guttag, 2013; Poh, McDuff, & Picard, 2010; Poh, McDuff, & Picard, 2011). One of them includes the use of face detection and feature extraction to obtain estimates of heart rate. For this, images or videos are recorded using digital cameras that are placed to frame the person's face during exercise performance or at rest, as the example showed in Figure 2. The most used technique to detect face during heart rate estimation is the Viola Jones object detection algorithm (Viola & Jones, 2001). The blind-source separation of the digital signal algorithm (Pal, Roy, Basu, & Bepari, 2013) to measure heart rate during rest was then purposed by (Poh et al., 2010). This study was extended to subsequently measure respiratory rate and HR variability (Poh et al. (2011).

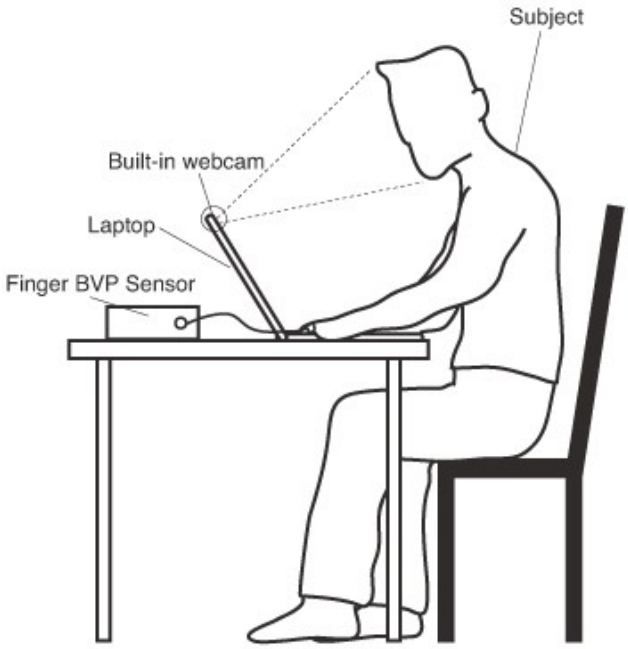


FIGURE 2 - EXPERIMENTAL SETUP FOR VIDEO DATA CAPTURING FOR HEART RATE MEASUREMENT USING BLIND-SOURCE SEPARATION ((POH ET AL., 2010).

Digital image color analysis have also been proposed to help in the HR estimation, and several digital filters has been proposed to increase the detection accuracy related to the color channels (More, Wakankar, & Gawande, 2019; Wang et al., 2017; Xie et al., 2019; Yang, Yang, Jin, & Wu, 2019; Yu, Kwan, Lim, Wong, & Raveendran, 2013). A case-study article introduced the use of ICA of each color channel of RGB in the raw images, followed by the application of short-time Fourier Transform (STFT) (Yu et al., 2013). Using proposed methods, root mean square error was less than 2.5 beats per minute (BPM) for heart rate variation between 80 BPM and 130 BPM.

Monkaresi (Monkaresi et al., 2014) proposed the advancement of work proposed by Poh (Poh et al., 2010; Poh et al., 2011) using the application of the kNN algorithm after extracting features by ICA. The experiments were carried out in various intensities of exercise performed in stationary bicycle. Other non-contact techniques to measure heart rate is Blood Volume Pulse (BVP), which detects blood volume in cardiovascular dynamics using optical sensor (Yang et al., 2019). Cui (Cui et al., 2015) proposed using facial image processing during exercise to measure BVP. The image pre-processing includes face recognition (Viola & Jones, 2001), band-pass filter (Mitra et al., 1991), trend

removal, and reconstruction of source signals to then retrieve BVP. Authors selected only the Green channel for the ROI for the further signal processing. This signal is applied to ICA and yield three independent signals. A series of image filters, Discrete Time Fourier Transform (DTFT), Discrete Short Time Fourier Transform (Discrete STFT), Discrete Wavelet Transform (DWT), Peak Counting, and Mean Value of Interbit Intervals (IBI) were applied to obtain the frequency spectrum of the raw signal.

The recent development of an exergame joins physical exercise to entertainment, typical of a gaming activity. That could provide solution to issues regarding treatments of age-related diseases (Vojciechowski et al., 2017). Proper extraction and implementation of physiological features while using exergames is critical. Spinsante (Spinsante et al., 2018) presented a contactless HR measurement system based on motion-compensated video signals which can be implemented in the remote controllers used for exergames.

A completely new approach was presented by Lin (Lin, 2018), that measured heart rate and step count during exercise, detected by facial image processing (see Figure 3). The authors implemented a chrominance-based adaptive filter and normalization method which enhanced accuracy. The detection rate was 99.52% and 99.7% for stepping and pulse rate using treadmill exercise, respectively. The estimated heart rate was compared to cardiofrequencimeter data during all the activities, showing positive correlation. The authors concluded that step count and heart rate can be measured synchronously during exercise.

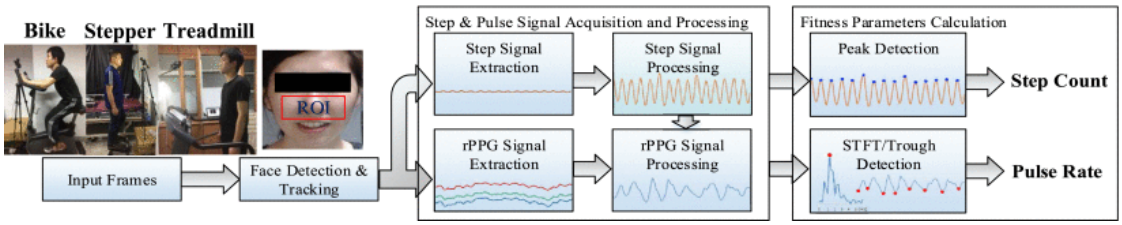


FIGURE 3 - EXPERIMENTAL SETUP FOR VIDEO DATA CAPTURING FOR HEART RATE MEASUREMENT USING BLIND-SOURCE SEPARATION (M.-Z. POH ET AL., 2010; M. Z. POH ET AL., 2011).

2.2.3. Facial Expression

The facial expression is an important cue to monitor physical exercise, since people express feelings through the face, and it changes due to exertion (Khanal et al., 2018; Khanal et al., 2018; Khanal, Fonseca, Marques, Barroso, & Filipe, 2018). Wu (Wu, Lin, Huang, Lin, & Chung, 2017) proposed a method to estimate Rate of Perceived Expression automatically without any wearable devices and questionnaires. Camera-based heart rate and fatigue expression feature extractor were fused. The Rate of Perceived Exertion (RPE) was considered roughly ten times the heart rate in the calculations, and it was correlated with various approaches such as PPG and fatigue feature, rPPG and fatigue feature, PPG, and rPPG. The authors reported high correlation coefficient of RPE and heart rate for the ground truth value of RPE. From the results, it was noted that if the information of facial expression related to fatigue is considered, the error in the estimation can be reduced.

2.3. Heart Rate Variability

Heart rate variability (HRV) is the physiological phenomenon of variation of time interval between heartbeats. It is an important physiological parameter to be considered to monitor intensity during exercise and many approaches has been proposed to measure HRV using contact and noncontact sensor technologies. (Hung & Zhang, 2006; Hunt & Fankhauser, 2016; Li et al., 2019; Nakamura et al., 2015)

A completely new approached was proposed by Hung (Hung & Zhang, 2006) to measure HRV by continuous monitoring of the Pupil Size Variability (PSV) during physical activity. The authors also measure blood pressure variability using the same approach. It was hypothesized that pupil size variability is highly correlated to both physiological parameters which are also an indicator of exercise intensity. Many parameters including electrocardiogram, respiration effort, finger arterial pressure, and pupil images were recorded from ten subjects before and after five minutes of exercise on a treadmill. The pupil images were captured by a 1/3’’ Charged Coupled Device (CCD)

camera connected to an 8-bit monochrome video frame-grabber, set to capture at a resolution of 512x512 pixels and capture speed of 2 frames per second. All the signals, including pupil images were acquired by a data acquisition unit (Li, Li, Lv, & Zhang, 2018) during exercise. The experiments underwent in six phases of 5 minutes recording sessions. The findings suggested that PSV may be a valid indicator of cardiovascular variability.

2.4. Blood Pressure

The blood pressure level can signpost the effort level when an individual is performing exercise, helping to estimate its workload (Inder et al., 2016). The blood pressure is also important to determine some patterns that could correspond to cardiac diseases, thus, significant on the monitoring of health output during exercise (Palatini, 1988).

2.4.1. Photoplethysmography and Pulse Arrival Time

The Pulse Arrival Time (PAT) is the time between the peak of the electrocardiogram (ECG) to the arrival of the pulse detected by the PPG. Fatemeh Shirbani (2018) investigated the correlation between image based photoplethysmography – pulse arrival time (iPPG-PAT) and diastolic BP (DBP) during one-minute seated rest and three minutes of isometric handgrip exercise. The video was recorded from the face using standard web camera and estimates were compared to a ground-truth device. It was found that the beat-to-beat iPPG-PAT and DBP were negatively correlated.

2.4.2. Photoplethysmography and Pulse Transmit Time

Pulse Transit Time (PTT) is the time a pulse wave takes to travel between two different arterial points, which is an important cue to estimate blood pressure. Several studies carried out to obtain the correlation between PTT and blood pressure. Jeong

(Jeong & Finkelstein, 2016) introduced a contactless approach to estimate blood pressure using PPT, with seven healthy subjects. Image based (iPTT) and image-based PPG were recorded using high speed camera at 426 Hz during physical exercise in a stationary bicycle. The exercises were carried out at three different times: rest, peak exercise, and recovery. The study found a high positive correlation between iPTT and iPPG during exercise, concluding that measuring BP using PTT is reliable. It was shown that the skin color changes due to blood pulsation such changes could be identified by the three-color component processing in the facial images.

2.5. Body Temperature

Body temperature rises due to physical exertion and relates to other physiological parameters, therefore, can also be considered when monitoring physical exercise intensity (Sawka, Wenger, Young, & Pandolf, 1993). Thermal imaging is the technology that allows for measuring the body temperature without any contact sensors.

Thermal imaging captures images of an object based on the infrared radiation emanated from it (Lopez et al., 2017; McFarlin, Venable, Williams, & Jackson, 2015). Its use in medical and sports environment is widespread (Ludwig, Formenti, Gargano, & Alberti, 2014). However, it seems that the body location is important in terms of assessment quality, once thermal emission might differ depending on the body part aimed to be analyzed (Fernandes, Amorim, Brito, Sillero-Quintana, & Marins, 2016). Ludwig (Ludwig et al., 2014) presented a critical comparison between the main methods used to obtain body temperature from images, and proposed also an alternative. It was found that temperature obtained within an ROI selection of a well-defined area can be considered as the most reliable.

James (James, Richardson, Watt, and Maxwell (2014) proposed an approach to investigate the validity and reliability of skin temperature measurement using a Telemetry Thermistor system (TT) and thermal camera during exercise in a hot and humid environment. Another similar study was presented by (McFarlin et al., 2015) to compare

data loggers (skin adhesive), thermal imaging, and wired electrodes for the measurement of skin temperature during exercise in a similar environment. The authors concluded that data logger and thermal imaging can be used as alternative measures for skin temperature in exercising, especially on higher temperatures and humidity.

Another study explored temperature in several different body parts pre, during and after moderate aerobic exercise using infrared thermography, was proposed by Fernandes (Fernandes et al., 2016) (see Figure 4). The authors concluded that there are significant distinctions in the skin temperature distribution during exercise depending on the body part.

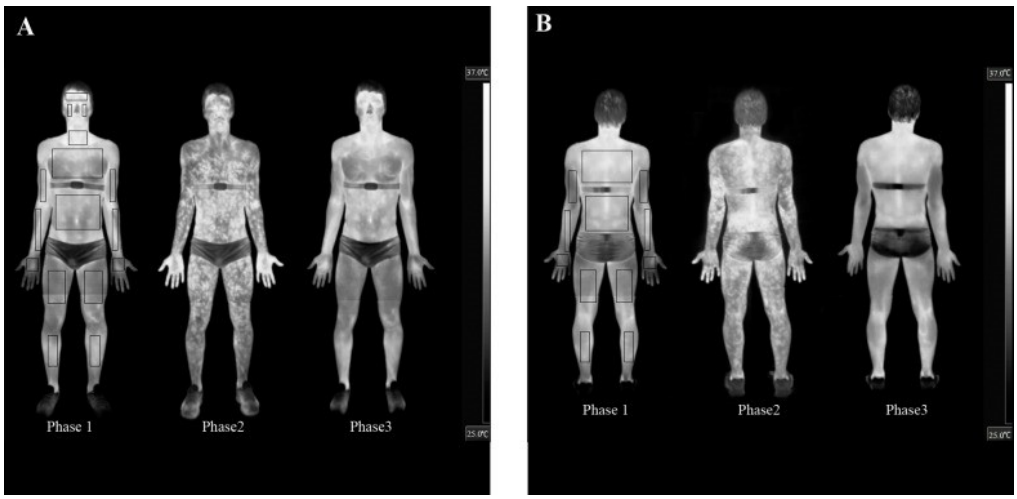


FIGURE 4 - THERMAL IMAGES OF WHOLE BODY TAKEN IN THREE PHASES OF EXERCISE (FERNANDES ET AL., 2016).

The changes in body temperature during endurance work-out in highly trained male sprinters were analyzed by Korman (Korman, Straburzynska-Lupa, Kusy, Kantanista, & Zielinski, 2016), using a thermal camera. The aim was to characterize athletes' warming up by comparing body temperatures in the four phases of the experiment: before the session (pre-exercise), warm-up, specific drills for sprinting techniques, and endurance exercise. It was found significant differences between the temperature of athletes' back and front body profiles, and significant changes before and after exercise. However, the thermography results were not compared to ground-truth measurements of temperature.

2.6. Energy Expenditure

Energy expenditure is highly correlated to exercise intensity, thus is essential to the planning, prescription and monitoring of physical exercise programs (Jensen et al., 2016; Koporec, Vučković, Milić, & Perš, 2018; Tao et al., 2016). Although its estimation is challenging, the oxygen uptake has been considered the best direct way to measure energy expenditure (Gade et al., 2017; Koehler & Drenowatz, 2017). Indirectly, it can be performed through heart rate and body acceleration (Koporec et al., 2018), which involves the development and application of non-contact technology as presented in the following sections.

2.6.1. Thermal Imaging

The introduction of thermal imaging in the exercise monitoring also allowed for the development of reliable contactless techniques for energy expenditure measurements. Thermal cameras capture infrared radiation in the mid or long-wavelength infrared spectrum, depending on sensor type, emitted from any object. Therefore, the pixel values in the images are converted to temperature values and finally mapped with energy expenditure.

Jenson (Jensen et al., 2016) validated the thermal imaging method to estimate energy expenditure using the oxygen uptake as a comparative. Fourteen endurance-trained subjects completed an incremental exercise test on a treadmill. Heart rate, gas exchange, and mean accelerations of the ankle, thigh, wrist, and hip were measured throughout the exercise. A linear correlation was found between the energy expenditure calculated using the optical flow of the thermal imaging and the oxygen uptake values. The contactless measurement of energy expenditure during exercise was also presented by Gade (Gade et al., 2017). The authors used thermal video analysis to automatically extract the cyclic motion pattern in walking and running. The results indicated a linear correlation between the proposed method and oxygen uptake.

2.6.2. RGB-Depth

One of the noticeable technologies used for image capturing is RBG-Depth camera, which captures objects in three dimensions. Tao (Tao et al., 2016) presented a framework for estimation of vision-based energy expenditure using depth camera and validating the method with oxygen uptake measures. The method was found suitable for monitoring in a controlled environment, showing advantages as pose-variant and individual-independent way of measuring energy expenditure, in a real time and remotely.

Deep learning technique has been considered one of the best tools to estimate, classify, and analyze quantitative data. Its application in sports has been increasing in the recent years. Wang (Wang et al., 2018) proposed a method for deriving calorific expenditure based on deep learning Convolutional Neural Network (CNN) features. This method is suitable for implementation in controlled environments, where the system first detects the presence of humans and then track the human body. This process is followed by a CNN-based feature extraction, then activity recognition and, finally, prediction of the calories produced, as shown in Figure 5 (Tao et al., 2017).

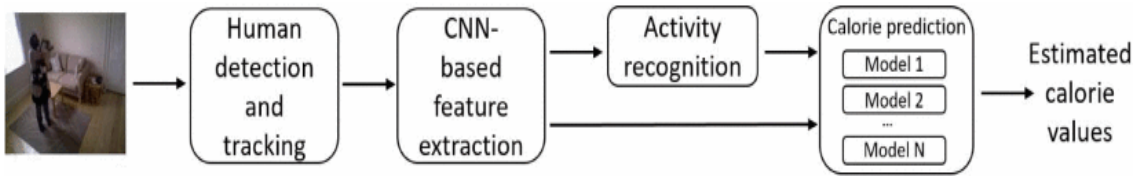


FIGURE 5 - OVERVIEW OF THE FRAMEWORK. THE PROPOSED PIPELINE USES VIDEO FOOTAGE AS INPUT AND EXTRACTS RELEVANT FEATURES USING A CNN (TAO ET AL., 2017).

A novel approach using a fully contactless and automatic method, based on computer vision algorithm, was presented by Koporec (Koporec et al., 2018). The RGB-Depth images are captured using Microsoft Kinect during exercise and a histogram of Oriented Optical Flow (HOOF) descriptors are extracted from the depth images and are used to predict heart rate. It feeds a regression model that finally estimates the energy consumption.

2.7. Respiratory Rate

Breathing is the process of taking the air inside the body so the oxygen can be absorbed, and then expelling carbon dioxide out of the body. Physical exertion not only increases the frequency of breathing, but also demands the exchange of a higher volume of air. Respiratory rate is directly linked to exertion, thus, to energy expenditure as well (Aliverti, 2016; Aoki, Ichimura, Kiyooka, & Koshiji, 2007; Persinger, Foster, Gibson, Fater, & Porcari, 2004; Sun et al., 2018).

2.7.1. Video-based Image Processing

The ventilation threshold is an important variable to investigate physical exertion. It is related to the anaerobic threshold, an event characterized by the increase in ventilation at a faster rate than what the body is capable of absorbing. In the recent decades, many methods have been proposed to measure the respiratory rate during exercise using contactless technology. Aoki (Aoki et al., 2007) proposed measuring respiratory rate during pedal stroke using optical techniques, as shown in Figure 6. A dot matrix optical element was arranged in front of the participant's face and a laser was emitted to be captured by the camera CCD. After using low-pass digital filtering, a sinusoidal wave which vibrates at the respiratory frequency was calculated. The results showed high correlation when compared to data obtained using gas analyzer.

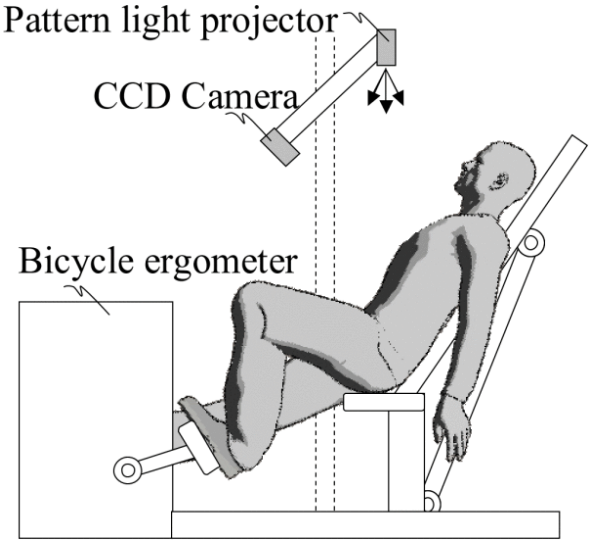


FIGURE 6 - MEASUREMENT OF RESPIRATORY RATE USING CCD CAMERA DURING EXERCISE IN BICYCLE ERGOMETER (AOKI ET AL., 2007).

2.7.2. RGB-Depth

Aoki and colleagues not only proposed and evaluated the use of CCD cameras for respiratory rate measurements, but also authored a series of publications exploring new trends in contactless sensor as Kinect. In 2015 the use of Kinect camera for respiratory rate measurement was validated (Aoki, Nakamura, Fumoto, Nakahara, & Teraoka, 2015). RGB depth images were captured and processed to obtain sinusoidal waves during exercise performed in a stationary cycle ergometer. The frequency found in the set of waves represented the respiratory rate.

Later, the authors investigated whether the new method was capable of providing good estimates of the ventilation threshold (VT). The experimental setup was maintained, but was applied to an incremental test, specific to identify this variable. The authors found that respiratory rate measures are possible from increments above 160 W and ventilatory threshold values can be estimated with ± 10 W of deviation from the VT calculated by gas analyzer (Aoki & Nakamura, 2018).

2.8. Muscle Fatigue

Fatigue is a subjective symptom of malaise and aversion to perform the activity or to objectively impaired performance (Sharpe & Wilks, 2002). It can be assessed by either self-report scales or performance based measures (Krupp). Fatigue can be either physical or mental and both types are important to assess due to its high correlation to health related parameters (Hulme et al., 2018; Karlsen, Larsen, Tandberg, & Jørgensen, 1999).

Facial expression is effective in assessing physical and mental fatigue (Haque, Irani, Nasrollahi, & Moeslund, 2016; Irani et al., 2014; Uchida et al., 2018). Irani (Irani et al., 2014) proposed an approach to measure fatigue by tracking facial features during exercise. The main hypothesis of the study was that, towards a fatiguing state, the points of interest in the image would increase vibration, thus, it could be identified in the power spectral analysis of the signal. This model was tested in maximal and submaximal dumbbell lifting test, against force measures obtained by a dynamometer. The results showed that the temporal point of interest in the face could be easily found using the method.

Deep learning and thermal imaging were fused to automatically detect in the face exercise-induced fatigue (Lopez et al., 2017). Different devices captured RGB, near infra-red, and thermal images, while the pre-trained CNN, Alexnet (Krizhevsky, Sutskever, & Hinton, 2012), and Visual Geometry Group - 16 (VGG16)/ VGG19 (Simonyan & Zisserman, 2014) were the deep learning methods used for the classification of different regions in the face according to fatigued/rested state. The authors found that the Alexnet applied to the region around the mouth showed best classification of the fatiguing state.

2.9. Other Approaches

2.9.1. Muscle Oxygenation

Muscles need oxygen supply to work, thus, aerobic muscle performance increases muscle oxygenation. This parameter is related to heart rate and blood pressure during exercise. It is also closely related to muscle fatigue, thus, might bring important information on the research about physical exercise intensity (Ellwein et al., 2017).

2.9.2. Facial Expression

The human face is a door for expressing feelings yield either by physical or mental condition. Pain, tiredness, and illness due to exertion is reflected in facial expressions (Khanal et al., 2018), therefore, to monitor the exercise intensity level, the analysis of facial expression analysis might be an interesting cue.

Khanal (Khanal et al., 2018; Khanal et al., 2018; Khanal et al., 2018) explored various methods of automatic classification of exercise intensities using computer vision techniques of a subject performing sub-maximal incremental exercise in cycle ergometer. The facial expression was analyzed by extracting 70 facial feature points. The exercise intensity was classified according to the distance between points and stage of the incremental exercise. The intensity was classified into two, three, and four classes using kNN, Support Vector Machine (SVM), and discriminant analyses. The results showed that facial expression is a good method to identify exercise intensity levels.

A different way of using facial expression to analyze the physical effort was presented by Uchida (Uchida et al., 2018). The facial images were analyzed at different levels of resistance training. The authors evaluated the changes in facial expression changes using Facial Action Coding System (FACS) and the facial muscle activity using

surface electromyography (see Figure 7). The association of these parameters was mild, however, statistically significant.

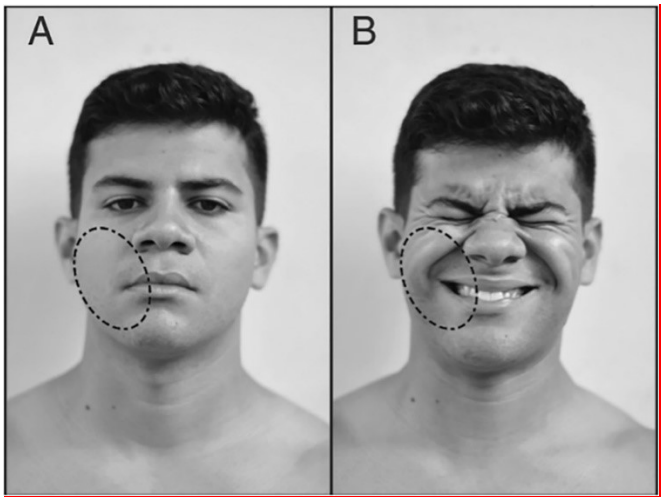


FIGURE 7 - CALCULATION OF MUSCLE FATIGUE BY ANALYSIS OF FACIAL EXPRESSION (M. C. UCHIDA ET AL.).

Miles (Miles, Clark, Periard, Goecke, & Thompson) presented also an analysis of the reliability in tracking data from facial features across incremental exercise on cycle ergometer. The results differed according to the face parts analyzed, but higher reliability was found for the lower face. Non-linear relationship between facial movement and power output were also determined. The power output, heart rate, RPE, blood lactate, positive and negative effects in corresponding exercise intensity were satisfied in the two blood lactate thresholds and maximum a posteriori probability MAP. These results show the potential in using the tracking of facial features as a non-invasive way of obtaining psychophysiological measures to access exercise intensity.

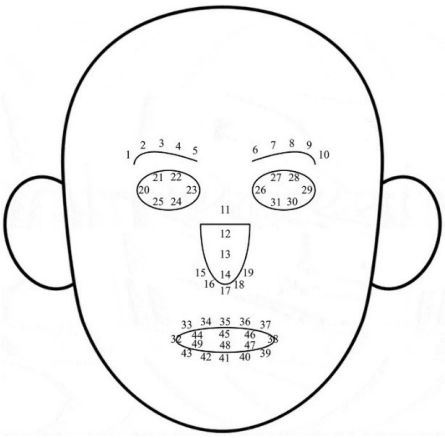


FIGURE 8 - FACIAL FEATURE POINTS EXTRACTION FOR EXPRESSION ANALYSIS (MILES ET. AL.)

Still regarding the use of facial features to evaluate levels of exertion, mouth and eyes are particularly interesting parts that express information by muscle actions. Thus, there is a variety of facial expressions and emotions heavily oriented by the eyes and mouth. Therefore, tracking the movement of these parts could be a key idea to analyze exercise intensity (Khanal et al., 2018). Recently, the eye-blink rate and open-close rate of the mouth were tracked using Viola and Jones algorithm for image processing (Viola & Jones, 2001) during sub-maximal exercise in a cycle ergometer (Khanal et al., 2018). The eye-blinking rate was correctly identified with 96% of accuracy. Additionally, the higher the exercise intensity, the higher the eye and mouth movement.

2.10. Conclusion

In the last two decades the use of image and video processing to monitor physical exercise has evolved and brought attention to contactless technology. Most of the research work presented in this review contributed to the development of methods capable of assessing important variables related to physical exertion, but there is still a gap in the implementation of such methods and technology. This review intended to provide a current and useful summary of the recent technology available for contactless devices and its application in sports sciences.

3. AN OVERVIEW OF MACHINE LEARNING ALGORITHMS AND ISSUES

This chapter presents the basic concepts and functioning of common machine learning prediction algorithms. We introduce classification and regression algorithms, which are used in our experiments. Regarding to classification techniques, thirteen machine learning classifiers and a deep learning architecture are explained. We also present a first benchmark study, which compares the accuracy of the classifiers. Beside these, some matrices for evaluation of the classifiers and the classification validation method are also explained.

3.1. Machine Learning

Machine learning is a branch of Artificial Intelligence (AI) that allows software applications to become more accurate in predicting outcomes without being explicitly programmed. In other words, it allows the computer to train the software by using data inputs and use statistical analysis on the data in order to output accurate values. There is a wide variety of different approaches within machine learning, with many algorithms requiring either training data to train the model, or to infer the rules. (Bishop, 1995). The machine learning can be divided into three different categories based on the way of learning and predicting the outcomes (Bishop, 1995) (see Figure 9).

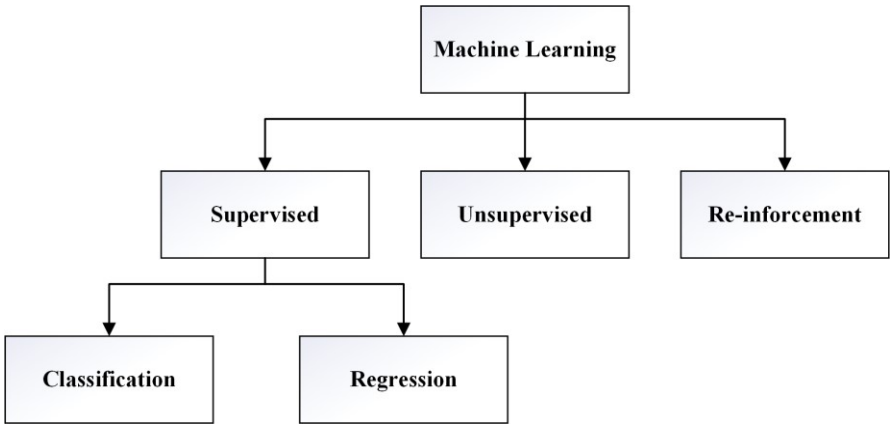


FIGURE 9 - CLASSIFICATION OF MACHINE LEARNING ALGORITHMS BASED ON WAY OF LEARNING AND PREDICTION.

In sports science, there is a strong presence of machine learning algorithms as a suitable data analysis tool. In the field of exercise monitoring, these algorithms have the potential to predict the level of exercise intensity, the physiological responses collected from sensor data, or level of fatigue (Balakrishnan et al., 2013; Capraro et al., 2018; Docampo, 2012; Dooley, Golaszewski, & Bartholomew, 2017; Khanal et al., 2019; Mackingnon, 1999; Thomson et al., 2019; Wiles, 2008). The learning process can be done as follows:

Supervised Learning: The algorithm learns from example data, which consist of numerical values or string labels, such as classes or tags, in order to later predict the

correct response. This form of learning is like human learning, where the human mind can predict the outcomes based on similar examples. It is further divided into two categories: Classification and Regression. The classification is a discrete type of prediction and analysis whereas as a regression continuous. The detail of these two algorithms will be described in the next section.

Unsupervised Learning: The algorithm learns from the plain examples without associated responses, leaving the algorithm to determine the data patterns on its own. This example, usually re-structures the data which represents a new series of un-correlated data.

Re-enforcement Learning: The algorithm is trained with examples without any labels, but either positive or negative feedback is provided, according to the solution the algorithm proposes. The machine learning model lets the algorithm know the outcome of actions it takes, with learning occurring as it tries to avoid negative results.

3.2. Machine Learning Classifiers

3.2.1. K-Nearest Neighbour Algorithm

It is assumed that a data instance is determined by the popularity vote of its neighbors (Aha, Kibler, & Albert, 1991; Duda, Hart, & Stork, 2000). An important parameter to be considered using this algorithm is the number of neighbors that perform the popularity voting. Mathematically, the nearest neighbors are determined by calculating the Euclidean distance from the test objects to all other objects, as shown by the expression 3.1. The training phase of this algorithm is very simple and fast, without the need to calculate special parameters.

$$\mathrm{d}(\mathrm{p},\mathrm{q}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_1)^2 + \cdots + (q_n - p_n)^2} \tag{3.1}$$

3.2.2. Support Vector Machine (SVM)

It is a popular machine learning classification and regression algorithm. Two classes are separated by a hyper plane, which has maximum margin, or maximum distance, from the data points of both classes (Pernek et al., 2015), as shown in Figure 10.

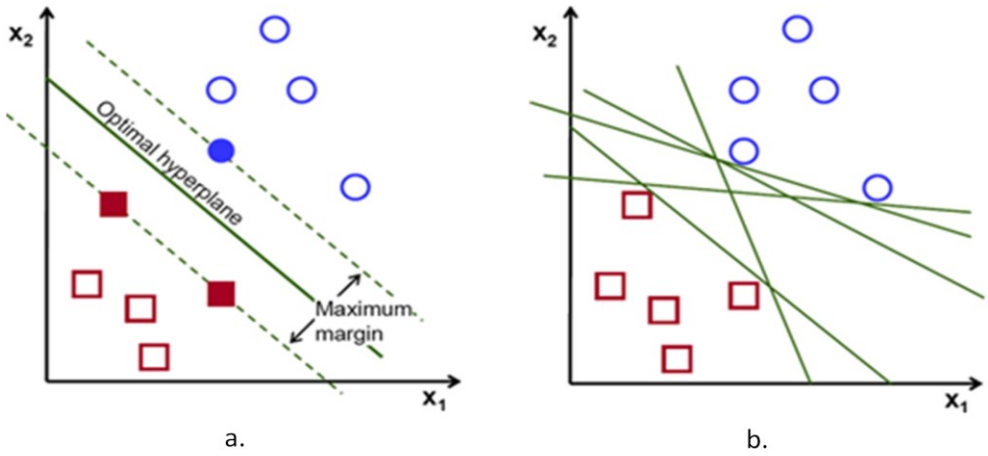


FIGURE 10 - SEPARATING HYPERPLANES FOR CLASSIFICATION USING SVM. RED RECTANGLE AND BLUE CIRCLES REPRESENT THE FEATURES OF TWO CLASSES. (IMAGE DOWNLOADED FROM HTTPS://DOCS.OPENCV.ORG/2.4/DOC/TUTORIALS/ML/INTRODUCTION_TO_SVM/INTRODUCTION_TO_SVM.HTML IN SEPTEMBER 2019).

The dimension of the hyperplane depends on the number of features. The hyperplanes are decision boundaries to classify data points. However, the basis of this algorithm is the classification into two classes, it can be used for more classes using the one vs rest classification.

The hypothesis function or classifier function h can be defined as

$$h(x_i) = \begin{cases} +1 \; if \; w.x + b \; \geq 0 \\ -1 \; if \; w.x + b \; < 0 \end{cases} \tag{3.2}$$

Where, $h(x_i)$ is hypothesis function to classify the classes. Two classes are supposed to be +1 and -1. The x (x_1, x_2, … x_n) is feature vector and w is a matrix to optimize to separate hyperplane.

The hyperplane can be defined as:

$$w^T x = 0 \tag{3.3}$$

The main optimization problem of this algorithm is the maximization of the margin by using an existing hyperplane as shown in Figure 10. Mathematically, the SVM classifier amount to minimize an expression of the form.

$$\left[\frac{1}{n}\sum_{i=1}^{n} \max(0, 1 - yi(w.xi - b))\right] + \lambda\|w\|^2 \tag{3.4}$$

In the training phase, the optimal hyperplane is determined, and, in the testing phase, the test features are tested with respect to the hyperplane.

3.2.3. Decision Tree

A decision tree indicates a flowchart-like tree structure, where features are attributes and the branches of a tree represent a decision rule. A decision rule is a simple expression or a threshold value to compare with input parameters. The decision is carried out through an iterative process. In this algorithm, the number of decision nodes depends on the number of features. A general decision tree flowchart is shown on Figure 11.

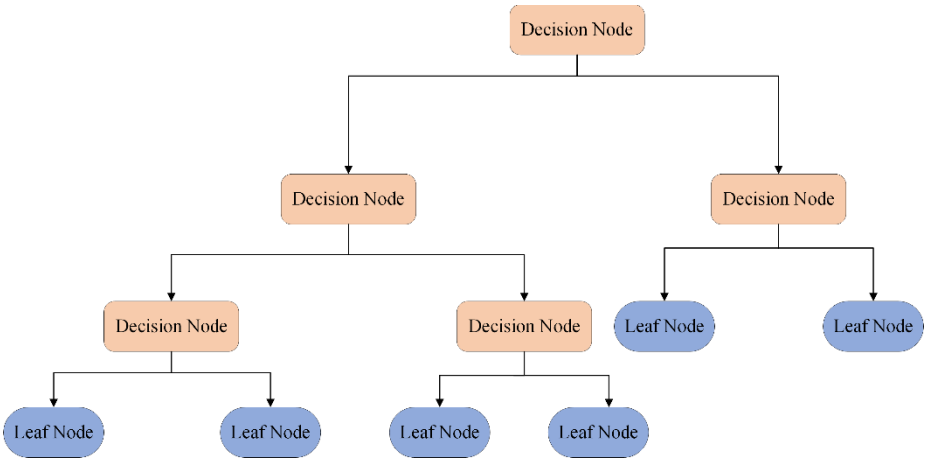


FIGURE 11 - GENERAL FLOWCHART REPRESENTATION OF A DECISION TREE. THE DECISION NODE REPRESENTS THE CONDITIONAL EXPRESSION AND THE LEAF NODE REPRESENT A CLASS.

3.2.4. Neural Network

It is a machine learning algorithm inspired by the human nervous system. It is made up of a number of interconnected processing units called neurons, which process on various inputs and provide a dynamic state response to external inputs (Caudill, 1989). It is typically organized in layers which include input layer, hidden layers, and output layer. The input layer presents patterns or data input to the network with communicating one or more hidden layers which are actually the neurons, and the hidden layers link to output layer which provides the output or results. The hidden layer could be one or many. Figure 12 shows a basic illustration of a neural network, where x_1, $x_2 \ldots x_n$ are features, w_1, $w_2 \ldots w_n$ are weights and $\sum$ represents the weighted sum adding also the bias (see expression 3.5). The weighted sum result is passed to activation function which provides the input to the next layer. Various activation functions can be applied to decide the output such as *Sigmoid*, *tanh*, *ReLU*, *Leaky ReLU* etc.

$$\text{act} = \sum_{i=1}^{n} xi \times wi + \text{b} \tag{3.5}$$

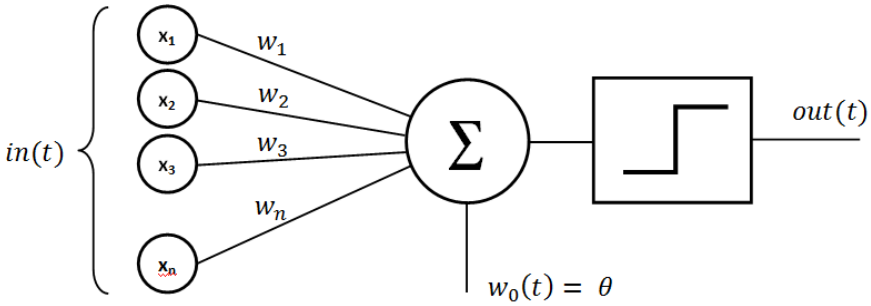


FIGURE 12 - BASIC BLOCK DIAGRAM OF NEURAL NETWORK.

Consider an activation function *ReLU* for detail illustration. It stands for a rectified linear unit which is mathematically defined as y = *max (0, x)*. The *ReLU* is the commonly used activation function in neural networks. It works based on a linear equation, as shown in Figure 13. Consider an output of an activation function *y* and an input to the function *x* (the result of expression ∑). If $x < 0$, then $y = 0$ or output is classified as class 0, otherwise $y = x$ as illustrated in Figure 13. Similar to *ReLu* activation function, all other activation functions decide the output class based on the expression they have used.

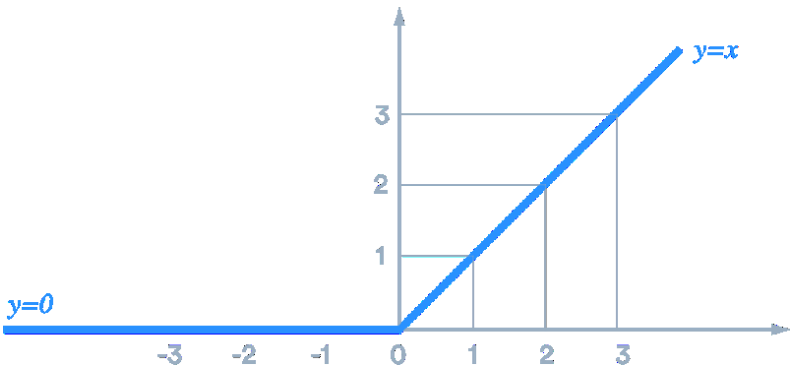


FIGURE 13 - AN ILLUSTRATION OF A RELU ACTIVATION FUNCTION.

3.2.5. Discriminant Analysis

It builds a predictive model to classify an observation based on probability of belonging to a particular class. Mathematically, consider an observation X = (x_1, x_2, x_3, … x_n) which belongs to class $i \in \{0,1, \dots k\}$, with the largest posterior probability $P(y_k/X)$. The posterior probability is calculated using Bayes' theorem (see expression 3.6).

$$p(k/X) = \frac{p(X/yk)p(yk)}{p(X)} \tag{3.6}$$

Where, k is the number of classes, $p(X/y_k)$ is the likelihood, $p(y_k)$ is the prior probability, and p(X) is the prior probability of predictor (Bishop, 1995).

Let's suppose, the class conditional distributions $p(y_k/x)$ are multivariate Gaussian, it can be calculated as:

$$P(X \mid \pi_i) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} exp\left[-\frac{1}{2}(X-\mu_i)'\Sigma^{-1}(X-\mu_i)\right] \tag{3.7}$$

Where, π_i is population function, μ_i is mean vector and Σ is variance-covariance matrix.

During training phase, μ_i *and* Σ of the conditional probability density function $P(X|\pi_i)$ are calculated to use in test phase. Two forms of discriminant analysis are the Linear Discriminant Analysis (LDA) and the Quadratic Discriminant Analysis (QDA). The LDA assumes the observations are normally distributed (Gaussian Distribution) and the different classes have class-specific means and equal variance. Mathematically, $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_K = \Sigma$ Whereas, in QDA the covariance matrix can be different for each class. Mathematically, $\Sigma i \neq \Sigma j$ for some $i \neq j$. Therefore, LDA is better for a small number of training samples, while QDA is recommended for large datasets.

3.2.6. Naïve Bayes Classifier

It is a probabilistic machine learning classifier to classify observations based on features. An observation x is classified through the Bayes' rule by calculating the posterior probability of each class (Duda et al., 2000). The classifier model learns from the conditional probabilities $p(x_i|y)$ (see expression 3.5) of each variable x_i, where yk is

given class labels. It is assumed that the variables are conditionally independent of the class label. Hence,

Let us suppose X is the features, X = (x_1, x_2, x_3,…., x_n)

$$P(X|yk) = P(x1, x2, \ldots . xn|yk) = \prod_{i=1}^{n} p(xi|yk) \quad (3.8)$$

So, the posterior probability can be written as:

$$p(yk/X) = \frac{\prod_{i=1}^{n} p(xi|yk)}{p(X)} \quad (3.9)$$

The classification problem is formulated for a class k as the maximum of $\prod_{i=1}^{n} p(xi|yk)$. Mathematically,

$$C = argmax_{yk} \prod_{i=1}^{n} p(xi|yk) \quad (3.10)$$

Where, C is the estimated class variable with outcomes with x_i given its features (x_1, x_2, x_3…x_n).

Regarding to the assumption of distribution of $P(x_i/y_k)$, different Naïve Bayes classifiers are used such as, Gaussian Naïve Bayes, Multinomial Naïve Bayes, Bernoulli Naïve Bayes etc. These algorithms are fast and easy to implement, but they require the predictors be independent. Due to the independent assumption about the feature data, it can quickly learn from high dimensional feature data, especially, high dimensional image data. The run time complexity is $O(n_k)$, where n is the number of features and K in number of label classes.

3.2.7. Logistic Regression

It is a probabilistic classifier that classifies observations using a logistic sigmoid function, which returns the probability value according to an observation of a particular class. The sigmoid function serves as a complex cost function (see Figure 14) and is a logistic instead of a linear function. The output of logistic function represented by σ(z) is between 0 and 1 or a probability estimation, z is input to the function or it is a feature vector of an observation. Usually, a threshold value above which an observation is classified as class 1 and below will be classified as class 2 is 0.5. For multiclass classification, we could select the class with highest predicted probability.

The input function z can be represented as shown in expression 3.11.

$$z = \sum w_i x_i + bais \tag{3.11}$$

$$\mathrm{P(class} = 1) = \frac{1}{1 + \mathrm{e}^{-z}}$$

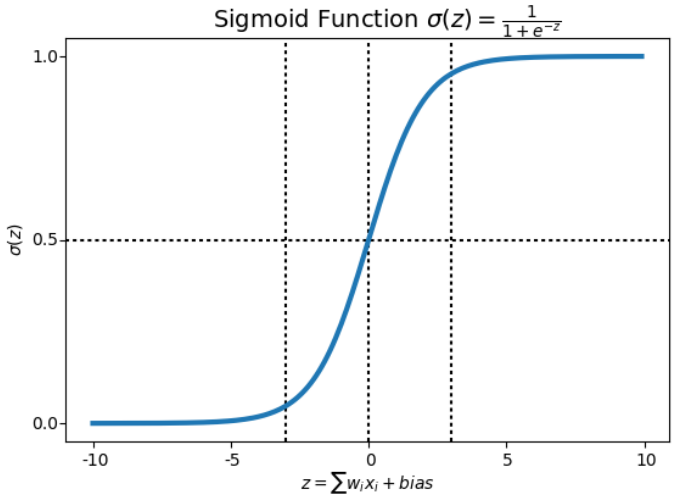


FIGURE 14 - SIGMOID FUNCTION GRAPH FOR LOGISTIC REGRESSION.

Let us generalize the model. Suppose, we have n observations and each observation consist p features. So, the hypothesis to predict the output can be represented as shown in expression 3.11.

Hence, the hypothesis can be represented in the form:

$$h(x_i) = \frac{1}{1 + e^{-w^T x_i}} \tag{3.12}$$

We can define the conditional probability for two classes (two labels: 0 and 1) for i^{th} observation.

$$P(y_i = 1|x_i; w) = h(x_i) \tag{3.13}$$

$$P(y_i = 0|x_i; w) = 1 - h(x_i) \tag{3.14}$$

Combining both expressions in a single expression as shown in expression 3.15

$$P(y_i|x_i; w) = (h(x_i))^{y_i} (1 - h(x_i))^{1 - y_i} \tag{3.15}$$

The likelihood of parameter w can be calculated as:

$$(w) = \prod_{i=1}^{n} (h(x_i))^{y_i} (1 - h(x_i))^{1 - y_i} \tag{3.16}$$

For easier calculations, we use log likelihood:

$$l(w) = log(L(z)) \tag{3.17}$$

$$l(w) = \sum_{i=1}^{n} y_i log(h(x_i)) + (1 - y_i) log(1 - h(x_i)) \tag{3.18}$$

The cost function for logistic repression is inversely proportional to likelihood of parameters. Therefore, the cost function can be represented as:

$$J(w) = \sum_{i=1}^{n} -y_i log(h(x_i)) - (1 - y_i) log(1 - h(x_i)) \tag{3.19}$$

The primary aim of logistic model training is to minimize the value of cost function. It can be carried out using Gradient descent algorithms (Bishop, 2006).

3.2.8. Ensemble Algorithms

Ensemble classification algorithms are machine learning algorithms that combine multiple but homogeneous weak models at the level of individual's output using various ways of combination methods to obtain the optimal predictive model (Cheng, Wu, & Li, 2018; Maclin & Opitz, 2011; Rokach, 2005). Following are the list of ensemble methods used in our classification problem. Ensemble algorithms are usually distinguished into two families:

Averaging methods - the basic principle is to build several estimators independently and average their predictions. Bagging, Random Forest, and Extra Tree classifiers fall into this family.

Boosting algorithm - the basic principle is to compile and build a strong estimator. Adaboost, Gradient boosting, and Voting fall into this family.

3.2.8.1. Bagging

It combines the predictions from multiple machine learning algorithms together to enhance the classification accuracy of an individual algorithm (Breiman, 1996; Breiman, 2001). This approach is used as a way to reduce the variance of a base estimator by randomizing the construction procedure and ensemble them. Usually, the observations and features are sub-sampled and create a classification model with a sub-sample of the observations and a subset of the features, and finally compile the output of each model. Suppose a feature dataset is divided into N number of sub-samples, then the probability

of being selected in some particular trial is 1/N, the probability of not being selected particular sample is 1-1/N, and the probability of not being selected M trials is $(1\text{-}1/N)^M$. Then, we build a model for each sample and calculate the majority vote for each class. A simple bagging classifier can be applied any basic classifier.

A specific case of bagging method is based on a divide-and-conquer approach, which combines more than one randomized decision tree of the same or different kind of classifying objects is called Random Forest. A diverse set of classifiers is created by introducing randomness in the classifier construction and averaging their predictions. It typically exhibits high variance, which results in overfitting. It creates a set of decision trees from a randomly selected subset of the training set. The individual decision trees are generated by using an attribute selection indicator, such as information gain, gain ratio, and the Gini index for each attribute (Breiman, 2001). Mathematically, assume a set of randomly slit training dataset $D = \{(X_1, y_1), \ldots. (X_n, y_n)\}$ with a probability distribution $(Xi, yi) \sim (X, Y)$.

Let us suppose, ensemble of weak classifiers: $h = \{h_1(x), \ldots h_k(x))\}$, where, $h_k(x)$ is set of decision tree. The parameters of the decision trees are supposed as:

$$\Theta_{kp} = (\theta_{k1}, \theta_{k2}, \theta_{k3}, \ldots \theta_{kp})$$

Therefore, $h_k(x) = h(x| \Theta_{kp})$

According to (Breiman, 2001), a random forest can be define as:

A random forest is a classifier based on a family of classifiers $h(x| \Theta_1)$, ... , $h(x| \Theta_k)$ based on the classification tree with parameter Θ_1 randomly chosen from a model random vector Θ.

Geurts (Geurts, Ernst, & Wehenkel, 2006) proposed some advancements in random forest with reducing the variance of predictor. As in the random forest, a random set of features is split such that the threshold value of each candidate features is randomly selected instead of looking for most discriminating one. This method is an estimator from a randomized decision tree. Basically, on various sub-samples of the dataset, it uses

averaging to improve the predictive accuracy and control the over-fitting. This ensemble technique is very powerful and has a consistent prediction rate.

3.2.8.2. Boosting

The basic principle behind a Boosting algorithm is to combine multiple weak predictor models to produce a strong predictor by repeatedly modifying the input features (Freund & Schapire, 1997). The predictions from all the predictors are combined with a weighted majority vote. The data modification in each iteration is subject to optimize the values of weights. As in a multilayer perception, in each iteration the weights are individually modified and learning algorithms are repeated. The processing to fit different models cannot be parallel since it consists in fitting sequentially multiple weak learners such that each model in the sequence is fitting which improves from the previous models. Therefore, comparing to bagging, this technique exhibits high computational complexity. Adaboost (Adaptive Boosting) and Gradient Boosting are two important boosting algorithms.

Consider the ensemble model:

$$\mathrm{SL}(.) = \sum_{l=0}^{L} c_l w_l (.) \tag{3.20}$$

where c_l is coefficients and w_l are weak learners.

The ensemble model is interactively optimized to get more optimal values of the coefficients, and weak learners such that ensemble model provides best accuracy.

The iterative procedure to get s is defined such that,

$$s_l(.) = s_{l-1} + c_l \times w_l(.) \tag{3.21}$$

Where, s_l and c_l are chosen such that the model fit the best values for training that and it improves in each iteration.

Unlike in Adaboost, it trains many models in a gradual, additive, and sequential manner. It computes a loss function (y = ax+b+e), where y is the loss function, x is the input features, b is the additive parameter and e is the error). The loss function is a measure that indicates how good the model is at fitting data. The problem is formulated such that the loss function is optimized, and a better model is created. The main advantage of this algorithm is it allows to optimize a user specific cost function. The gradient boosting algorithm casts the problem into a gradient descent one: a weak learner is fitted in each iteration, such that it is opposite of gradient of current fitting error with respect to the current ensemble model. The gradient descent process over the ensemble model can be written as:

$$Sl(.) = sl - 1(.) - cl \times \Delta_{sl-1} E(sl-1)(.) \quad (3.22)$$

Where E(.) is fitting error of the given model.

3.2.8.3. Voting

The basic principle behind this algorithm is to compile various machine learning classifiers and use the majority vote, or average predicted probabilities, to predict the class labels. The concept of this approach is very simple. A predicted class label for a particular observation is the class label that represents the majority vote from the class labels from the individual classifiers.

Suppose, we have a set of classifiers f1, f2, f3, .. fn. Using each classifier, the prediction of each class is carried out and the final decision is based on voting.

3.3. Deep Learning Classifiers

It is a sub-category of machine learning which offers a flexible and powerful approach to solve complex problems. The deep learning is made up of Artificial Neural Networks (ANN), which were inspired by human biological neural networks. It is a cascade of several processing units as the layers which consist input features, nodes (processing unit), weights, output etc.(Schmidhuber, 2015) Usually, deep learning models contain three basic types of layers: Input, hidden, and Output. A basic block diagram of neural network is shown in Figure 15. It has a wide area of application, including medical imaging, facial expression analysis, sport, etc.(Chen et al., 2019; Hernndez-Blanco, Herrera-Flores, Toms, & Navarro-Colorado, 2019; Khan & Yairi, 2018; Zhu et al., 2019; Zhu, Lu, You, & Chiang, 2019).

Input layer: It is a feature data. For image, it can be pixels of an image or a time series data.

Hidden layer: Multiple number of layers that consist of parameters called weights, which are learned or optimized, while the neural network is trained to obtain an accurate classification.

Output layer: Final layer provides class prediction from the input you fed into your network.

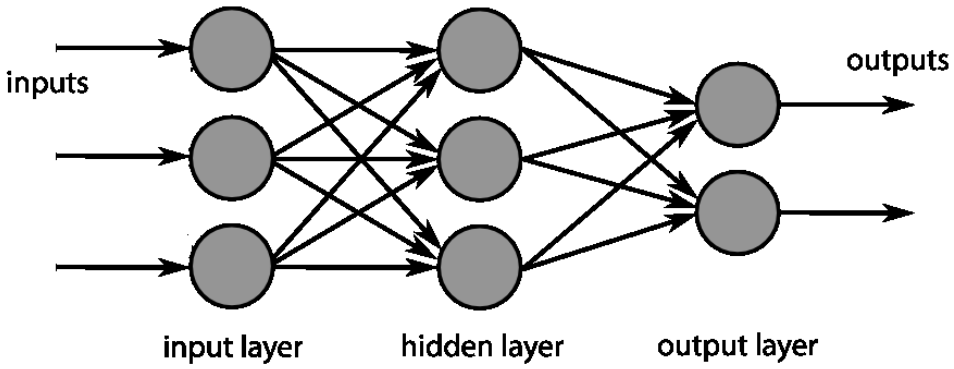


FIGURE 15 - BASIC BLOCK DIAGRAM OF A NEURAL NETWORK, WHERE THE CIRCLE REPRESENTS THE NODES OR THE PROCESSING UNIT.

So, the neural network is an approximation function in which the network tries to learn the parameters (weights) in the hidden layers, which, when multiplied with the input gives a predicted output close to the desired output.

Mathematically,

$$Y = \sum (weight * input) + bias \tag{3.23}$$

In a journey of less than a decade, several deep learning architectures have been proposed, but the basis of these architectures is the Convolution Neural Network. It is a deep neural network architecture that has proved to be very effective in terms of classification accuracy from a large set of observations. The CNN typically has three main layers: A Convolution layer, a Pooling or Sub Sampling layer, and a fully connected layer (see Figure 16).

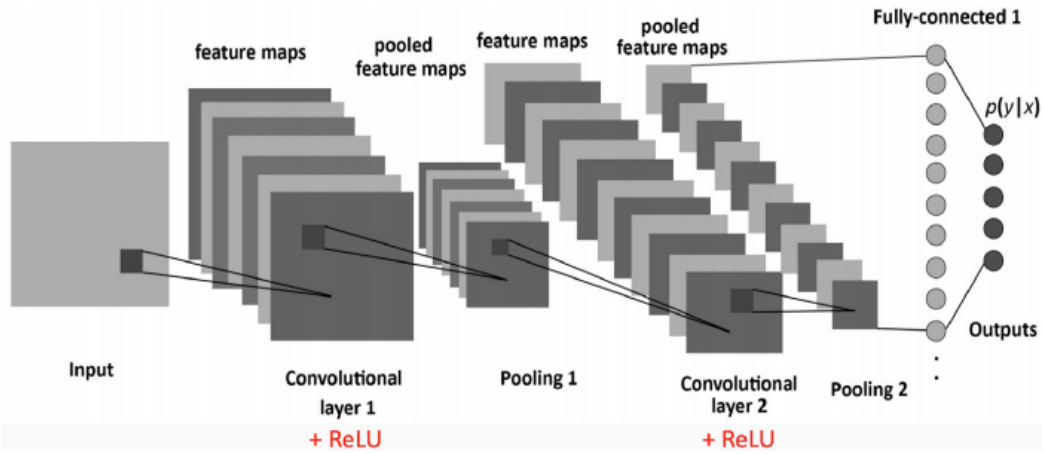


FIGURE 16 - A BASIC DIAGRAM OF A CONVOLUTIONAL NEURAL NETWORK WITH INPUT, VARIOUS LAYERS, AND OUTPUT. (IMAGE DOWNLOADED FROM HTTPS://TOWARDSDATASCIENCE.COM/UNDERSTANDING-NEURAL-NETWORKS-FROM-NEURON-TO-RNN-CNN-AND-DEEP-LEARNING-CD88E90E0A90, IN SEPTEMBER, 2019)

Convolutional Layer: The CNN is named by the convolution layer which computes dot product between patch of input image (input feature matrix) and a learnable parameter called a kernel. The kernel is smaller in size than the input images, but it can

be more in-depth. Generally, the input images are represented in the matrix which are the input feature matrix of the neural network. Therefore, every operation in the CNN is carried out by a matrix operation. Suppose, x is an input matrix, w is a convolutional kernel and m is the kernel width and height. The convolution output can be calculated as:

$$hi,j = \sum_{k=1}^{m}\sum_{l=1}^{m} w_k x_{i+k-1,j+l-1} \tag{3.24}$$

Polling Layer: The primary propose of this layer is to reduce the size of matrices by replacing a small portion of the matrix with a value. Two different types of pooling mechanism are used in practice (max-pooling and average-pooling). For example, if we consider a max polling, then a 2x2 matrix is transformed into one value consisting of the max elements of the sub-matrix. Mathematically, max-pooling can be calculated as:

$$hi,j = max\{xi+k-1, j+l-1 \; \forall \; 1 \leq k \leq m \; and \; 1 \leq l \leq m\} \tag{3.25}$$

Fully connected layer: The neurons are fully connected to all other neurons in the preceding and succeeding layers.

Another important layer of the CNN is the Non-Linearity layers, which consist of certain mathematic expressions that decide which value is chosen for the next iteration. Like in other machine learning algorithms, deep learning is also learning from the training. Technically, training the model is finding out the optimal values of parameters and it is represented by a cost function. Usually, the cost function is the root meaning square or binary cross entropy loss.

3.4. Regression

It is a statistical machine learning approach that determines the strength of the relationship between independent variables. It allows us to predict a continuous outcome

variable (Y) based on one or multiple predictor variables (X), that serve as input variables. The main goal of the repression analysis is to build a mathematical equation that defines Y as a function of X, which is carried out during the training of the model (Bishop, 1995). Based on the input variables, regression can be divided into two categories:

Linear regression: It considers a linear relationship between the outcome and predictor variables by using a mathematical equation. A simple linear regression plot with two variables (one input and one output) is shown in Figure 17.

$$Y = a + bX + u \tag{3.26}$$

Multiple regression: Instead of having only one predictor multiple regression consists more than one predictor to find the outcome.

$$Y = a + b1X1 + b2X2 + b3X3 + \cdots . + bnXn + u \tag{3.27}$$

Where: Y = output variable or a variable that is going to be predicted.

X = the input variable(s) that are used to predict the output variable.

a = the intercept

b = the slope

u = the regression residue.

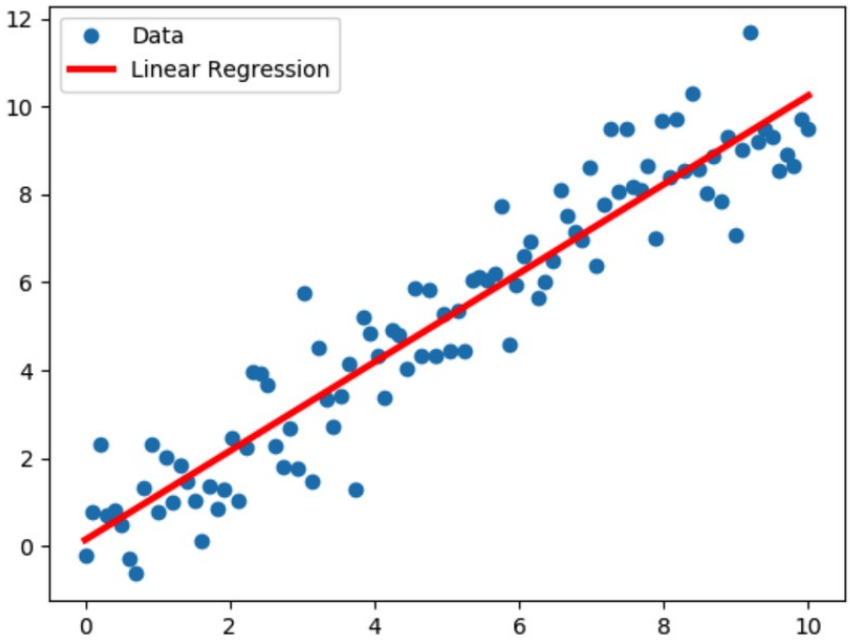


FIGURE 17 - LINEAR REGRESSION EXAMPLE (IMAGE SOURCE: (TRAN, 2019)).

The regression coefficients are determined in such a way that the error in predicting the outcome value is minimized. It is common practice to predict heart rate from various indicators such as facial color, facial expression, RPE, etc. in exercise science (Capraro et al., 2018; Scribbans et al., 2016; Zhang et al., 2017).

3.5. Model Validation and Assessment

The classification performance of a classifier can be evaluated by various methods, where a testing set should be properly chosen so that the results can be generalized for new test samples (Riou, Rioux, Lamothe, & Doucet, 2015). Following is a list of model assessment and evaluation methods we used in our experiments.

3.5.1. K-Fold Crossing-Validation Method

It is a statistical method that evaluates the result of the classification and validates how the accuracy of the classification can be generalized to a new dataset. In the validation procedure, a feature dataset is divided into training, validation, and test datasets, as shown in Figure 18. Generally, the training and testing data are split from a same bundle of observation samples.

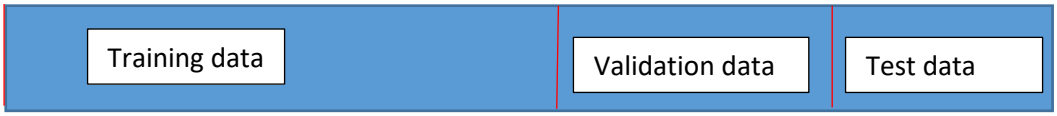


FIGURE 18 - DATA SPLIT FOR TRAINING, VALIDATION, AND TESTING TO TRAIN AND TEST THE MACHINE LEARNING CLASSIFIER.

In the case of k-fold cross validation, the sample dataset is divided into K sub-samples, with one sub-sample being used as the testing set, while the other K-1 samples are used for training. Each sub-sample is validated once against the rest of sub-samples as the training set. The experiment is performed iteratively k times to individually test each sub-sample. Finally, the average accuracy of all iterations is calculated. k=10 is common practice in k-fold cross-validation and is called 10-fold cross validation. The general procedure in each validation method is: shuffle the dataset, split into k groups, and for each group: train the model, evaluate the model and, finally, evaluate the models based on evaluation score. With this method the overlap between training set is lessened, thus the test error estimates are less correlated and exhibit a higher value of error variance.

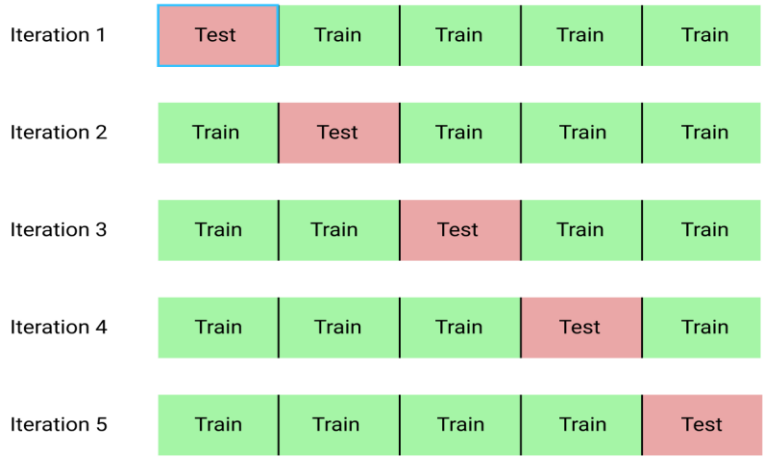


FIGURE 19 – K FOLD CROSS VALIDATION WITH FIVE FOLDS AND FIVE ITERATIONS.

3.5.2. Leave-One-Out-Cross-Validation (LOOCV)

In this approach the feature dataset is partitioned into several parts. Suppose we have n observations, with the training dataset containing n-1 observations and the testing dataset one observation. The training and testing procedure are repeated n times, keeping one observation out as the validation, as shown in Figure 20. The validation process is iterated for each data point and repeated for the training and validation procedure n times, ending with the calculation of the mean square error (MSE). A bias-variance trade-off between doing in leave-one-out and k-fold cross validation is an important factor to be considered. Unlike the K-fold, the test error in LOOCV generally has a with lower bias because the entire training set is used in each iteration and a higher variance due to there being a lot of overlap between the training sets.

If the observations are collected from a few subjects, for example, observations of the facial images to recognize emotion, both the validation methods fail to validate the results due to a huge variation in the facial expression per subject. In such cases, validate the classifier using leave-one-subject-out approach where subjects in the training dataset are not overlapped in testing dataset (Lopes, de Aguiar, De Souza, & Oliveira-Santos, 2017; Montoye et al., 2018).

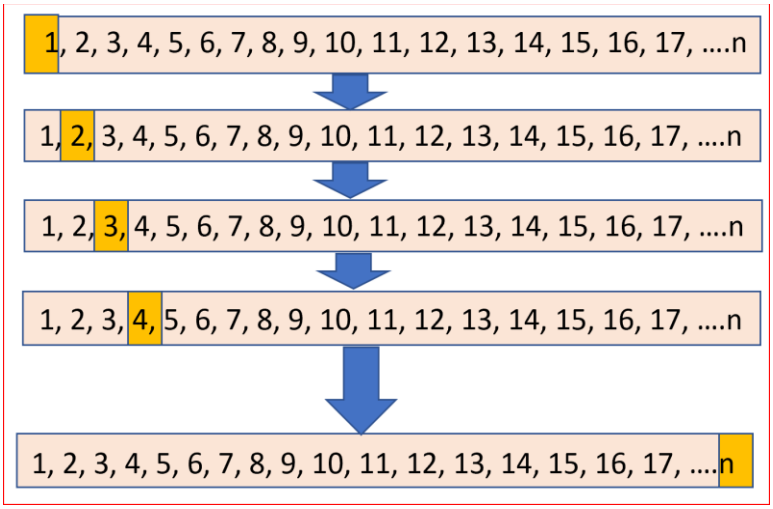


FIGURE 20 - LEAVE-ONE-OUT-CROSS VALIDATION PROCEDURE.(IMAGE DOWNLOADED FROM HTTPS://MEDIUM.COM/DATADRIVENINVESTOR/K-FOLD-AND-OTHER-CROSS-VALIDATION-TECHNIQUES-6C03A2563F1E, IN OCTOBER, 2019).

3.5.3. Confusion Matrix

It is a well-known tool to describe the performance of a classifier for a dataset. The terms to describe the confusion matrix are: True Positive (TP), False Negative (FN), False Positive (FP), and True Negative (TN). These parameters are presented in Table 1.

TABLE 1 - CONFUSION MATRIX WHERE HORIZONTAL VALUES REPRESENT PREDICTED VALUES AND VERTICAL REPRESENT ACTUAL VALUES.

	Predicted (Yes)	**Predicted (No)**
Actual (Yes)	True Positive (TP)	False Negative (FN)
Actual (No)	False Positive (FP)	True Negative (TN)

Let us suppose we have a classification problem where we must predict two fruits, Apple and Orange. Consider an observation in which the label, apple is considered as positive and orange as negative.

True positives (TP): Number of observations predicted correctly as positive. For example, the observation is to predict an apple and the classifier predicted it as an apple.

True negatives (TN): Number of negative observations predicted correctly as negative. For example, the observation is to predict an orange and the classifier predicted it as an orange.

False positives (FP): Number of negative observations predicted incorrectly as positive. For example, the observation is to predict an orange and the classifier predicted it as an apple.

False negatives (FN): Number of positive observations predicted incorrectly as negative. For example, the observation is to predict an apple and the classifier predicted it as an orange.

3.5.4. ROC Curve

It is a graphical tool to evaluate various machine learning classifiers for a specific dataset. It plots on a graph the false positive rate (X-axis) versus the true positive rate (Y-axis). It gives the generalized idea of how good a classifier is at classifying a given observation based on the predicted probability. Another parameter to considered on ROC curve is the Area Under the Curve (AUC). The bigger area represents better machine learning models at distinguishing classes. The ROC of the different classifiers can be plotted in the same graph, which makes it easier to compare the classifier with the pros and cons. The ROC curve closer to the upper left corner of the 2D graph indicates better classifiers. The ROC curve can also be plotted for multiclass classifier.

3.5.5. Performance Evaluation Parameters

Accuracy

Very common parameters to evaluate the classifier are the accuracy and the error rate. The accuracy is the proportion of samples that were correctly classified with respect to the total number of samples, whereas the error rate is the opposite of the accuracy. Mathematically,

$$Accuracy = \frac{Total\ number\ of\ samples\ correctly\ classified}{Total\ number\ of\ samples} \tag{3.28}$$

Precision Rate and Recall Rate

The precision and recall rates can be defined using the following equations:

$$Precision = \frac{TP}{TP + FP} \tag{3.29}$$

$$Recall = \frac{TP}{TP + FN} \tag{3.30}$$

There is an opposite interdependency between recall and precision. If the recall rate of the output increases, its precision rate will reduce and vice versa.

Area Under Curve (AUC-ROC)

AUC–ROC curve is a model selection and evaluation tool for bi–multi class classification problems. A typical ROC curve has the False Positive Rate (FPR) on the X-axis and the True Positive Rate (TPR) on the Y-axis. This covered area is the AUC. In general, the bigger that area is the better that machine learning model is for that particular class. The ideal value for the AUC is 1.

F-measure

$$F1 = \frac{2xRecallxPrecision}{Recall + Precision} \tag{3.31}$$

F1 is defined as the harmonic mean of the precision rate and the recall rate.

4. CLASSIFICATION OF PHYSICAL EXERCISE INTENSITY BASED ON FACIAL EXPRESSION

This chapter presents computer vision techniques to classify the physical exercise intensity by using facial features. Two different models for physical exercise intensity classification are presented, traditional machine learning classifiers and deep learning classifiers, both evaluating with the same image dataset. In a first model thirteen conventional machine learning algorithms were compared to the physical exercise intensity classification into four classes. In the second model, a CNN architecture based on ConvNet was proposed. The classification accuracy of each model was evaluated and validated by using the confusion matrix, k-fold cross validation, leave-one-subject-out, and ROC curve.

4.1. Background

When a subject gets fatigued as an effect of physical exercise, it is common that one of the changes can be detected in facial expressions, therefore, it seems like that exercise intensity can be classified by analyzing facial expressions (Khanal et al., 2018; Khanal et al., 2018; Khanal et al., 2018; Khanal et al., 2019). More specifically, fatigue effects can be detected in the shape of the facial components like the mouth, lips and eyes (Khanal et al., 2018; Miles, Clark, Périard, Goecke, & Thompson, 2017). For instance, if a person gets tired, he needs to breathe faster and often open and close mouth, which changes the position of facial feature points near the mouth. Using these cues, the machine learning techniques can be used to classify different levels of fatigue. One of the well-known factors related with fatigue is exercise intensity. The exercise intensity can be detected or classified by tracking, and analyzing facial expressions during the physical exercise itself (Irani et al., 2014).

In fact, the facial expression can be detected by analyzing the geometric position of facial feature points which changes with respect to the physical exercise intensity (Ekman, Wallace, & Ancoli, 1980). Some of the available studies use facial video recordings for tracking the facial feature points. These feature points can be analyzed by using various supervised machine learning algorithms (Miles et al., 2017). Haque (Haque et al., 2016) presented an efficient non-contact system to detect non-localized physical fatigue by analyzing facial expression changes due to higher levels of physical exertion. The problem formulation on facial expression analysis for physical exercise intensity classification is similar to facial emotion, a well-known classification problem applied in various areas (Miles et al., 2017).

According to state-of-art studies, the facial feature tracking is a reliable approach to detect the level of physical exercise intensity (Marco C. Uchida, 2018; Miles et al., 2017). The findings of past studies suggest the potential utility of facial feature points as a non-contact and psychophysiological measure to determine exercise intensity. However, the impact of this reliability was never tested in real-time data. Another similar

study was presented by Irani (Irani et al., 2014) to measure muscle fatigue by tracking facial feature points.

An important indicator to define the level of exercise intensity is the HR. Therefore, in this study the ground truth class/level of exercise intensity was defined according to the incremental value of HR. The intensity level (class) of exercise at the beginning or the minimum maxHR % is the initial class '*light'* and at the end of the exercise, the maximum heart rate percentage (maxHR%) is the final class or '*hard'* level. Intermediate classes are defined based on maxHR%.

For a decade, deep leaning have become wide-spread technology applied in various fields, such as facial expression recognition for various purposes like emotion recognition (Liliana, 2019; Samadiani et al., 2019), pain detection (Soar, Bargshady, Zhou, & Whittaker, 2018), exercise monitoring (Liao, Vakanski, & Xian, 2019; Um et al., 2017) etc. The deep learning can also be implemented to analyze or monitor physical exercise by analyzing biometric. With development of deep learning techniques in various areas with more accurate results, a couple of research works using them to detect the level of fatigue are presented. Gordienko (Gordienko et al., 2017) proposed a multimodal approach to estimate the fatigue by using the input parameters extracted from wearable sensors.

The primary aim of the present study was to identify how hard a subject is feeling (exercise intensity level) to carry out the exercise during its execution by analyzing facial expression. One noticeable limitation of the deep neural network architecture is training speed. Due to the huge amount of data necessary to train the models, the training time is very high. If you train model directly with the RGB images, the number of features to test and train the network will be immense, which results in a slow execute. The secondary aim of the study was to find out the best color channel among Red, Green, and Blue to determine which values changes most in respect to the exercise intensity level. We evaluate the results from each color component as well as Grayscale to find out the best color component that gives the most accurate result.

To our best knowledge, many authors show the high correlation between facial expression and exercise intensity either by analyzing feature points or facial color, but the classification of exercise intensity using feature point tracking was never examined. In this work, based on reliability, we classified the various levels of exercise intensity based on facial feature position analysis.

In this chapter, the physical exercise intensity level was classified using facial image processing, machine learning, and CNN classifiers. Thirteen well-known machine learning classifiers were used to classify the exercise intensity and the results are validated through K-fold cross validation and leave-one-subject-out validation technique and evaluated via confusion matrix, ROC curve, and accuracy measurements.

4.2. Dataset Preparation

The facial videos were collected in a laboratory normalized setting (SportTech Lab from the University of Trás-OS-Montes and Alto Douro) with twenty subjects (mean age = 26.88 ± 6.01 years, mean weight = 72.56± 14.27 kg, mean height = 172.88 ± 12.04 cm, 14 males and six females, and all white Caucasian) familiarized with cycloergometer protocols. An informed consent form was signed by each participant prior to data collection and they were informed of the study protocol. The subjects underwent a submaximal cycling exercise (Wattbike cycle ergometer, Wattbike Ltd, Nottingham, UK), started pedaling at 50W and the intensity was increased 50W every 2-min (approximate cadence of 70 $rev.min^{-1}$). The heart rate data were collected at 100Hz using the Polar T31 transmitter monitor (Polar Electro, Kempele, Finland) and the power output was calculated by measuring the chain tension over a load cell sampled at 100 Hz (Wattbike cycle ergometer, Wattbike Ltd, Nottingham, UK). A stationary camera (Canon VIXIA HF R800 Camcorder) located approximately one-meter distance from the face and at 90° angle between the face and the camera was used to capture the facial video (see Figure 21). Following are the conditions applied while collecting video data and heart rate.

- The exercise stopped if either a participant could not carry out the exercise more in current setting or a participant stopped himself when he cannot continue the exercise. The duration of the video was different from each subject, depending on the resistance of each individual, with a frame rate of 25 fps (frames per second) and a spatial resolution of 1080×1920 pixel.

- Heart rate, blood pressure, body temperature, height, and weight were measured and asked their date of birth prior to the experiment and each participant should carry out one minute of warmup pedaling exercise with 50W load and approximately cadence of 70 $rev.min^{-1}$.

- The participants were not allowed to talk during the test but could express their feelings freely with facial expression throughout the exercise and suggested to watch camera lenses as far as possible.

- The study protocol was setup inside the room, therefore, experiments were carried out in room temperature with ambient light plus light generated by fluorescent bulb located in the ceiling approximately three meters far from the face.

- Prior to experiment, each participant was asked to abstain from consumption of any caffeinated food or beverage for at least two hours and alcoholic beverages for at least 24 h. The hard-physical exercise was also not allowed 24 h. prior to the experiment.

- The participants were not allowed to have deep makeup and eyeglasses during the experiments.

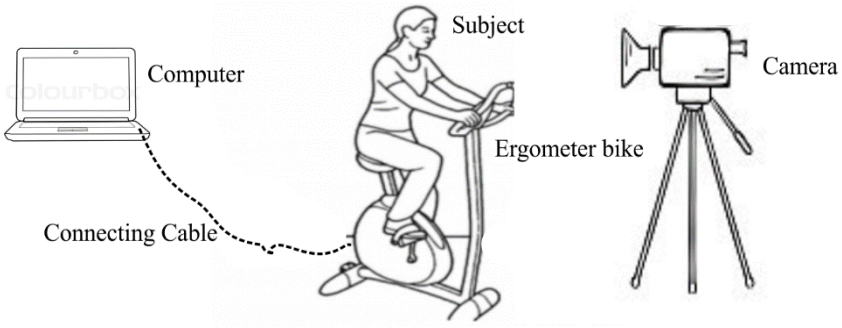


FIGURE 21 - DATA COLLECTION (FACIAL VIDEO AND HEART RATE) SETUP.

As HR is one of the most commonly used parameters to determine exercise intensity. According to (Pollock & Wilmore, 1990), four distinct levels of exercise intensities (light, moderate, vigorous, and very hard) can be determined through maxHR%, where maxHR% is calculated as: (HR/(220-Age)×100) with potential adjustment of ± 10-15 beats per minute (Reynolds, 2014). The image frames were manually assigned to a class, accordingly to the maximum heart rate percentage (maxHR%) (See Figure 20). Suppose, the initial images (first 500) extracted from the video were considered as *light* or class-one, the last 500 images were considered as *very hard* or class-four, and intermediate classes (*moderate* and *vigorous*, each 500 images) were extracted from 70% and 85%, a midway maxHR% of 64-77% and 77-94%. The selection of image frames from the video is illustrated in Figure 22.

TABLE 2 - CLASSIFICATION OF EXERCISE INTENSITY FOR CARDIORESPIRATORY ENDURANCES BASED ON %HRMAX (POLLOCK & WILMORE, 1990).

Exercise intensity	**maxHR%**
Light (Class one)	<64
Moderate (Class two)	64-77
Vigorous/Hard (class three)	77-94
Very Hard (Class four)	>94

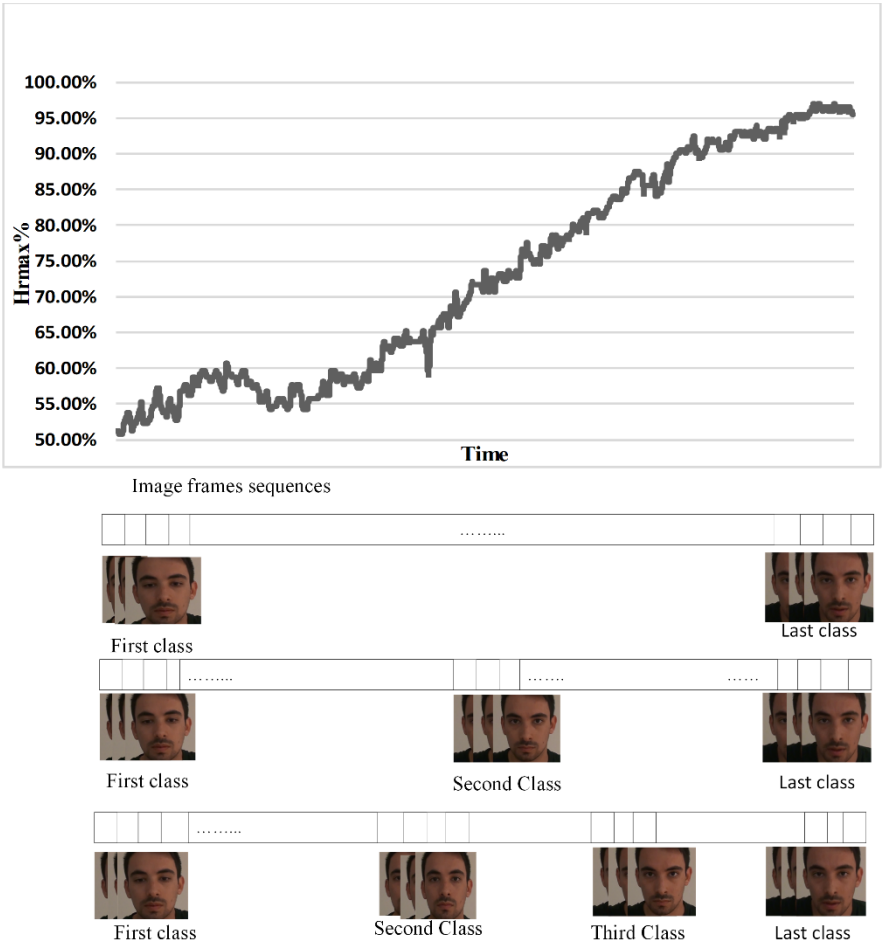


FIGURE 22 - ALLOCATION OF TIME SLOTS TO EXTRACT IMAGES WITH THE INITIAL CLASS (MINIMUM EXERCISE INTENSITY), THE INTERMEDIATE CLASSES, AND THE FINAL CLASS (MAXIMUM EXERCISE INTENSITY).

4.3. Machine Learning Approaches

The classification of facial images was performed using thirteen machine learning algorithms based on facial feature points analysis. All the experiments were carried out in Python 3.7. Most of the algorithms were implemented using the default values of their parameters through *sklearn* python library.

4.3.1. Experiments

The experiments were carried out from the facial images extracted and classified from the video. The basic block diagram is shown in Figure 23. For each image of each class, the face was recognized, cropped, and resized to 400x400 pixels. For face detection, Viola Jones algorithm was used (Viola & Jones, 2001). After resizing, 68 features were extracted using a DLIB library (King, 2009). The DLIB library detects 68 landmarks from various locations of the face as shown in Figure 15. Each landmark represents the specific location of face with respect to the facial components such as eyes, mouth, nose etc. The Euclidian distances between a landmark from a nose tip or landmark-31 (see Figure 24) and all other landmarks were calculated and considered as features to feed the machine learning algorithms. The Euclidian distance between two points is calculated using the expression 4.1:

$$Distance\ (P1, P2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{4.1}$$

Where (x1, y1) and (x2, y2) are two coordinate points. After detecting and calculating the feature points distance for all the images of each class, a feature matrix was created with 67 feature points. Afterwards, each classifier was validated and evaluated using the above mentioned techniques and matrices. The basic block diagram is shown in Figure 23.

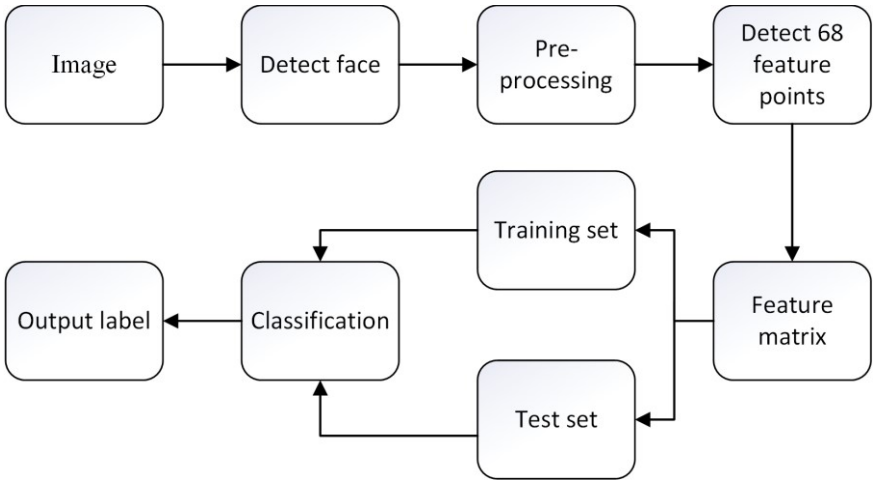


FIGURE 23 - BASIC BLOCK DIAGRAM OF IMAGE CLASSIFICATION USING MACHINE LEARNING ALGORITHM.

Facial feature points are a set of pixel positions representing salient regions of the face such as eyes, mouths, nose, eyebrow etc. The pixel's positions were normalized with relative to detected face. A sample of the facial feature points are shown in Figure 24.

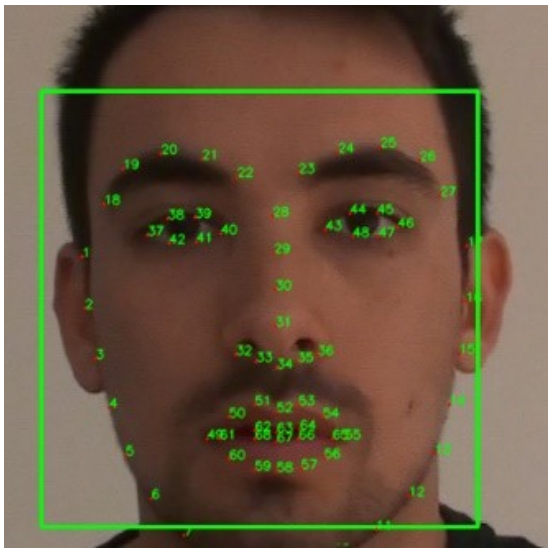


FIGURE 24 - FACIAL FEATURE NUMBER DETECTION FROM THE FACE.

The facial feature points near the mouth and eyes are normally deviated when a person feels fatigue (Khanal et al., 2018), thus, the distances between these points and the fixed point 31 (this point is considered as central point) were considered as feature points. The distances between the central point and other feature points of four classes of exercise

intensity are shown in Figure 25. All the distances were considered as a feature vector to train machine learning algorithm. The classification was validated using the 10-fold cross validation and evaluated using Precision, Recall, and Area Under of Receiver Operating Characteristics (AUROC) curve.

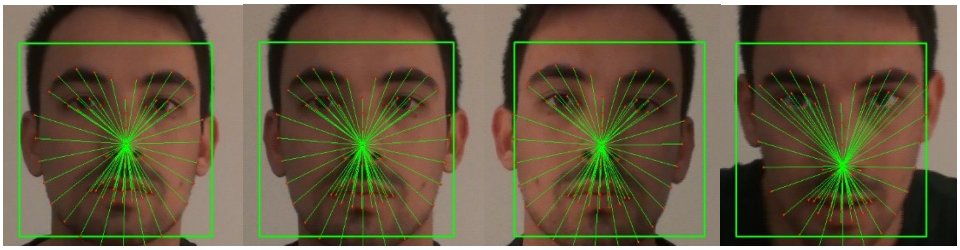


FIGURE 25 - IMAGES AFTER EXTRACTION THE FEATURES FROM THE FEATURE POINTS. THE LENGTHS OF GREEN LINES AROUND THE MOUTH ARE THE FEATURES TO CLASSIFY.

4.3.2. Experimental Results

4.3.2.1. Accuracy and Training Time

The accuracy and execution time in a single trial of each algorithm is presented in Table 3. Accuracy was calculated as the average accuracy for each class; however, the accuracy of each class is also presented using the confusion matrix in the next section.

TABLE 3 - THE AVERAGE CLASSIFICATION ACCURACY FOR EACH CLASSIFIER AND EXECUTION TIME.

Parameter	Accuracy (%)	Training Time (sec)
KNN (k=3)	91.89	0.51
Linear Discriminant Analysis	45.95	0.54
Quadratic Discriminant Analysis	48.65	0.48
Naïve Bayes'	37.83	**0.37**
Logistic Regression	48.65	224.02
SVM	78.39	11708.17
Decision Tree	89.19	4.48
Random forest	89.66	120.20

Multilayer Perceptron (MLP)	81.08	356.23
Voting	64.86	1.062
Adaboost	62.16	54.84
Gradient Boosting	72.97	170.13
Bagging	86.48	426.36
Extra Tree	**94.59**	452.52

According to the results, classifiers based on probability models are weak classifiers, while the classifiers based on ensemble methods are stronger. The KNN algorithm also provided the better result with faster training speed. The ExtraTree classifier, which is based on the ensemble algorithm, provided an accuracy of 94.59%. The classifiers are evaluated and compared using various tools, such as the K-Fold cross validation, ROC curves, confusion matrix and some matrices like precision, recall, F1-measure, and AUROC. The following sub-section (4.3.2.2) presents the results of methods evaluation.

4.3.2.2. K-Fold Cross Validation

Because the value of K is 10 it is called a 10-fold cross validation. The average accuracy of all the folds and standard deviation are presented in Table 4. The best classifier is the ExtraTree with an average accuracy of 94% and a standard deviation of 0.004137. Based on accuracy, the best four algorithms (KNN, Random Forest, Bagging and Extra Tree) were plotted on a comparative graph (See. Figure 26). For further analysis, only the best four classifiers were considered.

TABLE 4 - THE AVERAGE ACCURACY AND STANDARD DEVIATION OF ACCURACIES OF 10-FOLD CROSS VALIDATION.

Classifier (10-fold)	Accuracy (%)	Standard Deviation
KNN (k=9)	0.92	0.004292
Linear Discriminant Analysis	0.54	0.006734
Quadratic Discriminant Analysis	0.58	0.075435
Naïve Bayes'	0.37	0.007431
Logistic Regression	0.54	0.007571
SVM	0.83	0.004525
Decision Tree	0.86	0.005452
Random forest	0.93	0.004847
Multilayer Perceptron (MLP)	0.84	0.028867
Voting	0.68	0.008032
Adaboost	0.60	0.006343
Gradient Boosting	0.81	0.006473
Bagging	0.92	0.004521
Extra Tree	**0.94**	0.004137

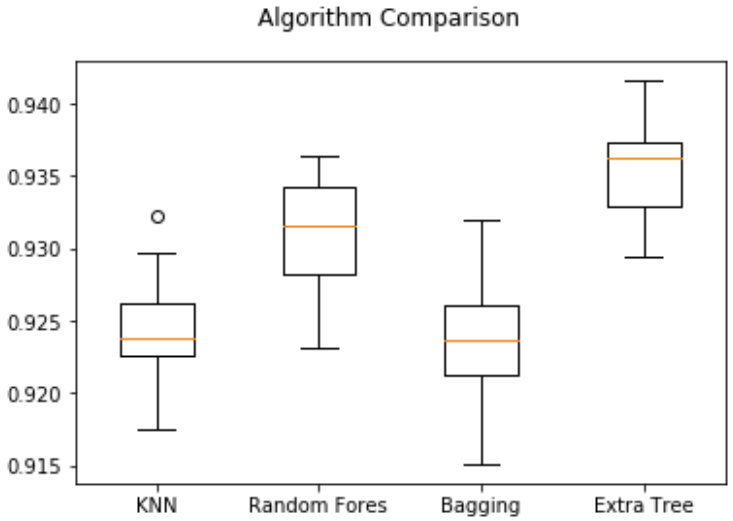


FIGURE 26 - COMPARING BEST FOUR ALGORITHMS BASED ON AVERAGE ACCURACY.

4.3.2.3. Confusion Matrix

The confusion matrix not only evaluates the performance of classifier but also compares the accuracy of individual classes. The confusion matrix of the ExtraTree classifier is shown in Figure 27. In most of classifiers, the true positive rate of the first and last class are high compared to the middle classes. The false negative and true negative values are higher between the nearest classes.

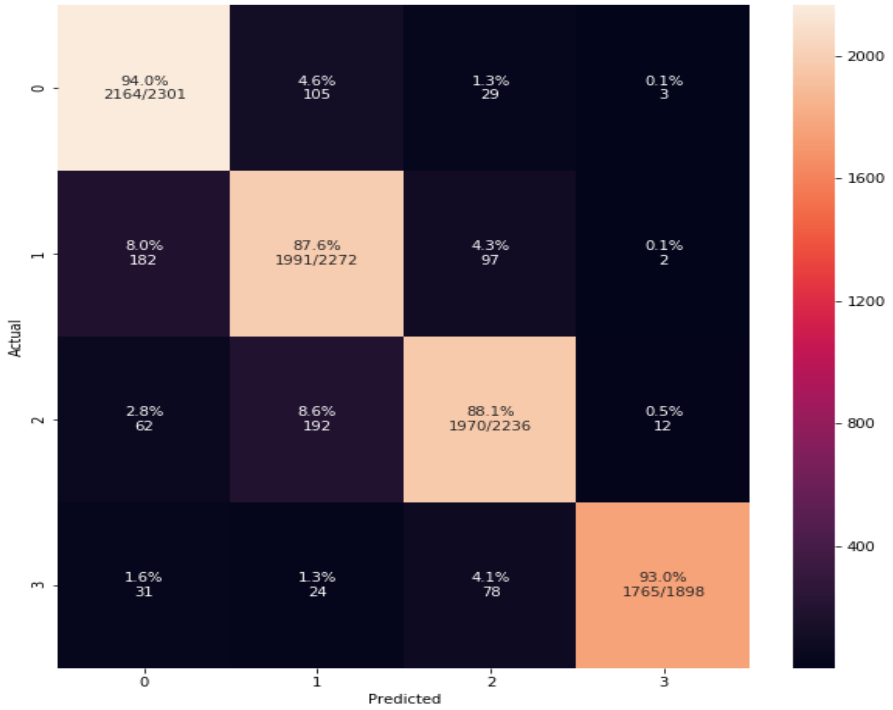


FIGURE 27 - CONFUSION MATRIX OF EXTRA TREE CLASSIFIER.

4.3.2.4. ROC Curve

ROC curve is a plot of the true positive and false positive rates. Only the best classifier, according to 10-fold cross validation, are chosen to plot their ROC curve. The important parameter to consider in evaluating the classifier is the Area under curve

(AUC). A ROC curve of Extra Tree classifier is shown in Figure 28. The average value of the Area Under ROC curve (AUROC) is 0.95 and AUROC curves of class 3 (highest value of exercise intensity) is highest (0.95) which implies that it is more distinct from other classes.

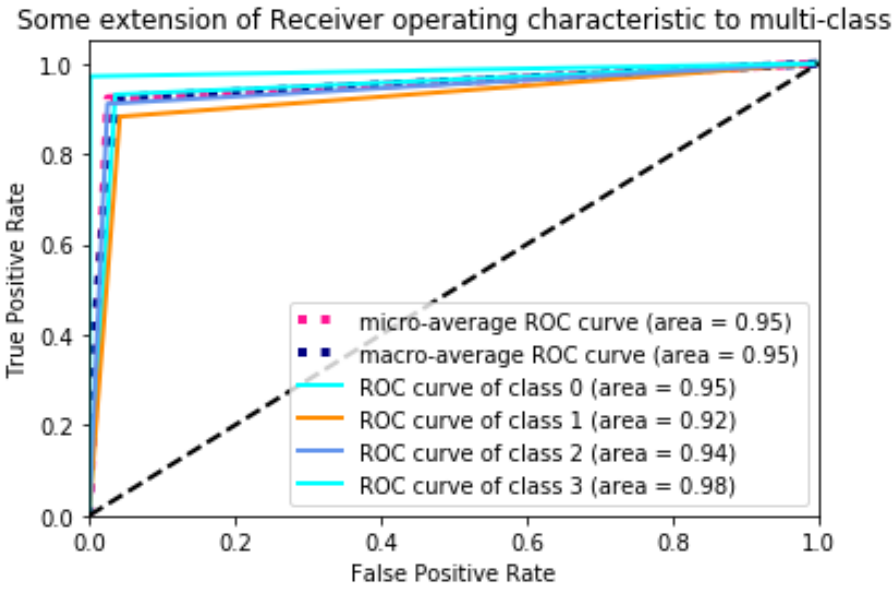


FIGURE 28 - ROC CURVE PLOT OF EXTRA TREE CLASSIFIER.

4.3.2.5. Precision, Recall, AUC, and F-measure

The Precision, Recall, AUC, and F-measures are metrics to evaluate and compare the machine learning classifiers. These score's metrics of all thirteen algorithms are shown in Table 5. According to the result, the Extra tree classifier obtained the best result for individual class and the prediction accuracy is better in the initial and the final classes rather than the middle classes. For instance, a *class-one* and *class-four* have better accuracy, precision, recall, AUC, and F-values than *class-two* and *class-three.* Therefore, it shows the facial expression is quite different in the beginning and end and more similar in the in-between period.

TABLE 5 - PRECISION, RECALL, AREA UNDER CURVE (AUC), AND F-MEASURES OF EACH CLASSIFIER OF 13 MACHINE LEARNING CLASSIFIER.

Classifier	**Precision**					**Recall**					**Area Under Curve**				
	C1	**C2**	**C3**	**C4**	**Avg**	**C1**	**C2**	**C3**	**C4**	**Avg**	**C1**	**C2**	**C3**	**C4**	**Avg**
KNN	0.90	0.87	0.92	0.99	0.92	0.92	0.88	0.90	0.97	0.92	0.94	0.91	0.93	0.98	0.94
LDA	0.49	0.44	0.46	0.86	0.56	0.52	0.43	0.52	0.71	0.55	0.66	0.62	0.65	0.84	0.69
QDA	0.63	0.47	0.78	0.85	0.68	0.69	0.78	0.17	0.92	0.64	0.77	0.74	0.57	0.94	0.76
Naive	0.27	0.40	0.35	0.59	0.40	0.24	0.19	0.62	0.48	0.38	0.50	0.54	0.60	0.69	0.58
Logistic	0.50	0.44	0.45	0.75	0.54	0.53	0.38	0.48	0.75	0.54	0.66	0.61	0.63	0.84	0.69
SVM	0.94	0.91	0.63	1.00	0.87	0.84	0.81	0.95	0.62	0.81	0.91	0.89	0.88	0.81	0.87
DT	0.84	0.78	0.85	0.93	0.85	0.83	0.81	0.84	0.93	0.85	0.89	0.87	0.89	0.95	0.90
RF	0.92	0.88	0.93	0.98	0.93	0.93	0.90	0.91	0.97	0.93	0.95	0.93	0.94	0.98	0.95
NN	0.86	0.70	0.71	0.90	0.79	0.68	0.73	0.81	0.92	0.79	0.82	0.81	0.84	0.94	0.85
Voting	0.55	0.72	0.66	0.94	0.72	0.75	0.47	0.74	0.77	0.68	0.76	0.70	0.80	0.87	0.78
Adaboost	0.54	0.51	0.48	0.82	0.59	0.57	0.50	0.51	0.75	0.58	0.69	0.67	0.66	0.85	0.72
GBoosting	0.80	0.75	0.80	0.93	0.82	0.80	0.78	0.80	0.87	0.81	0.86	0.84	0.86	0.93	0.87
Bagging	0.91	0.87	0.92	0.98	0.92	0.92	0.89	0.90	0.96	0.92	0.94	0.92	0.94	0.98	0.95
Extra Tree	**0.92**	**0.88**	**0.94**	**0.99**	**0.93**	**0.94**	**0.91**	**0.91**	**0.97**	**0.93**	**0.96**	**0.93**	**0.95**	**0.99**	**0.96**

4.4. Deep Learning Approach

4.4.1. Proposed Methodology

The classification problem presented in Section 4.3 is also addressed by using the deep learning approach. As mentioned in Chapter 3, deep learning classifiers do not need a separate feature extraction step. The proposed deep learning architecture is based on a CNN was trained, validated, and tested with manually classified images and the results were validated using K-fold, leave-one-out-subject cross validation while the evaluation was performed with the ROC curve and confusion matrix.

4.4.1.1. Pre-processing

Before feeding the neural network with inputs, various image pre-processing techniques were applied. All those pre-processing steps are shown in Figure 29. Since the frames were extracted from the video recorded with a moving object (head movement), the images may have some motion blurred effects. Therefore, the first pre-processing consists in removing any blurring effect on the image frame. In the second step, the face was detected in the frame so we could analyze the face, specially, rather than the whole frame. The well-known Viola Jones algorithm (Viola & Jones, 2001) was applied to detect the face. After detecting the face, we cropped and down-sampled it into 96x96, which is the size of the input layer. One of the basic purposes of this model is to find out the best color channel to classify the physical exercise; as such, the experiments were performed with a separate raw 2D image representing each color channel (Red, Green and Blue) and Grayscale. So, after cropping the face and resizing, the RGB frames were split into their respective R, G, B channels, and converted to Grayscale image.

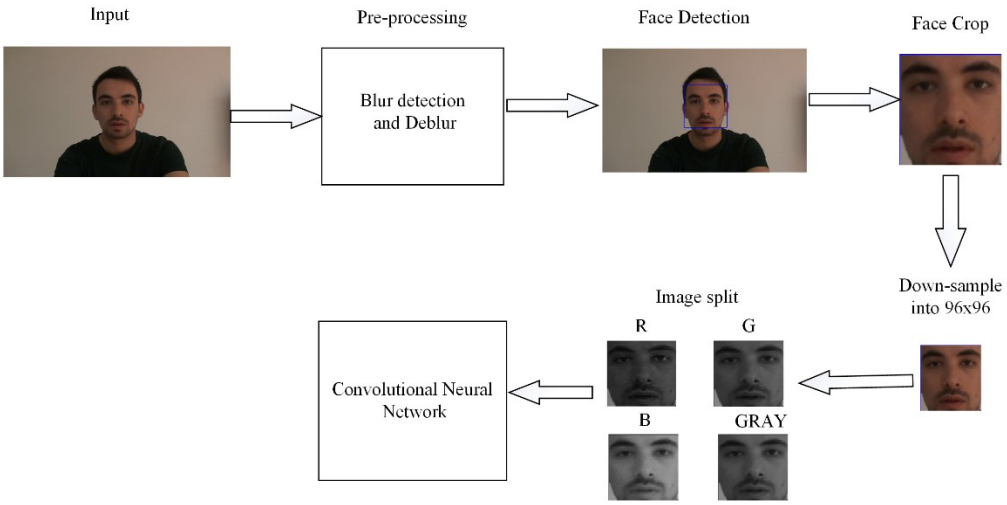


FIGURE 29 - BLOCK DIAGRAM WITH DETAIL PREPROCESSING WITH OUTPUT IMAGES OF EACH STEP.

4.4.1.2. Proposed CNN Architecture

In this study, a deep neural network based on the Convolutional Neural Network (CNN) or ConvNet was designed with five hidden layers and two fully connected layers, as shown in Figure 30. Three main layers used to build the ConvNet architectures

consist Convolutional Layer, Pooling Layer, and Fully-Connected Layer (exactly as seen on a regular Neural Networks). These layers were stacked to form a full ConvNet architecture.

Input Layer [96x96]: holds the raw pixel values of the 2D face image.

Each five hidden layer consists:

- CONV layer [3x3xN]: computes the convolution on output of neurons that are connected to local regions in the input, each computing a convolution of their weights, as well as a small region they are connected to in the input volume. The filter size was 3x3.
- BatchNormalization: Process of resetting the distribution of the output of the previous layer to efficiently process by the subsequent layer.
- The activation function chosen was ReLU.
- Maxpooling with Pool size (2,2).
- 25% dropout results in the maximum amount of regularization.

Fully-connected layer [1x1xN]: computes the class scores, resulting in volume of size [1x1xn], where n is the number of classes. As with ordinary Neural Networks, and as the name implies, each neuron in this layer will be connected to all the neurons in the previous layer.

The first part of each layer consists of a convolutional layer (Conv2d) which can have spatial batch normalization, Maxpooling, dropout and ReLU activation. Each layer consists of these five tasks.

We implemented this architecture in the well-known python library Keras. The experiments were carried out in the Google Colab Graphics Processing Unit (GPU).

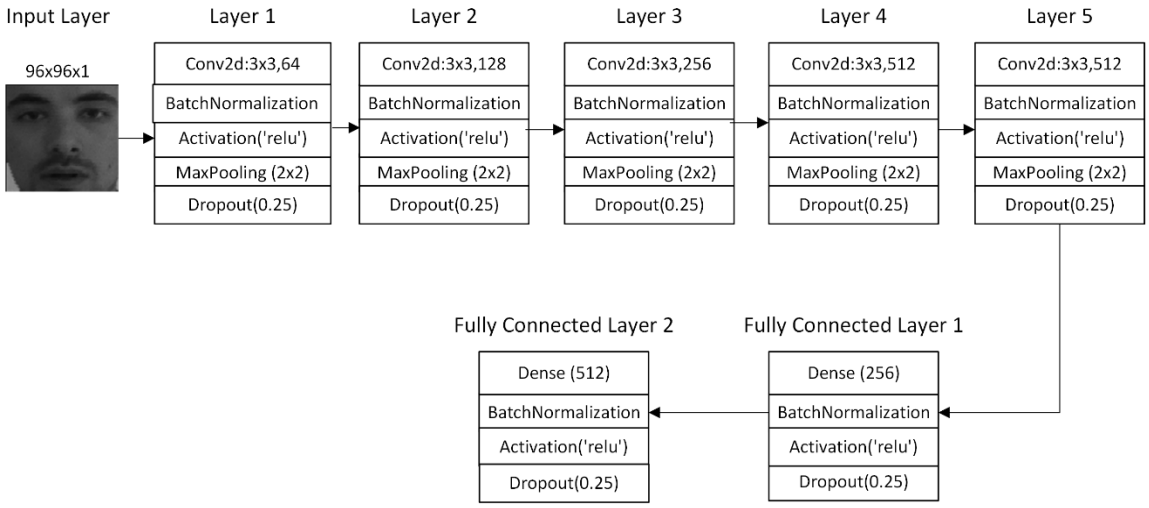


FIGURE 30 - PROPOSED CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE.

4.4.2. Experiments

After all, proposed deep neural networks were trained individually and tested with a random split dataset. The first convolutional layer consists of 64 3x3 filters; the second one had 128 3x3 filters; the third one had 256 3x3; the forth one had 512 3x3 filters; and the last one also had 512 3x3 filters. In all the hidden layers had a stride size of 1, batch normalization, max-pooling of size 2x2, dropout of 0.25 and ReLU as the activation function. These five hidden layers were followed by two fully connected layers with 256 neurons and 512 neurons respectively. Both the fully connected layers had batch normalization, dropout and ReLU with the same parameters. SoftMax is also used as an out-loss function.

The training was performed in 75 epochs with a batch size of 64. From the dataset of 10,000 images for each class, the dataset was randomly split into training, validation and testing set with a ratio of 80:10:10. For two classes (*tired* and *not tired*) the total number of images was 20,000, where 16,000 were for training, 2,000 for validation, and 2,000 for testing. Experiments with two, three, and four classes were also performed. To reduce overfitting, we used dropout and batch normalization in addition to L2 regularization.

In most of state-of-art studies of image classification (mostly face images), the classification results are validated in such a way that subjects in the training set overlap with the test set, which exhibits unfair validation of new test set with a new subject (Lopes et al., 2017; Montoye et al., 2018). In present thesis the results were validated using leave-one-subject-out method in order that, among twenty subjects, eighteen subjects are used for training, one subject for testing, and one for testing. The accuracy was calculated as an average of each iteration. The results were also validated using the 10-fold cross validation. After the cross validation, the results were represented in a confusion matrix, plot ROC, and some parameters were calculated such are precision, Recall, AUC, and F-measure.

4.4.3. Experimental Results

Separate experiments were done for each color space and the accuracy of each case was analyzed. In the experiments, the color images were split into Red, Green and Blue components and the original RGB images were converted into Grayscale. The green color component provided the best classification accuracy. The results were validated using the 10-fold and leave-one-subject-out cross validations and evaluated using various evaluation tools such as confusion matrix, ROC curve, and evaluation indices like, the precision, recall, AUC, and F-measure. The result of each analysis is presented in their respective sub-sections. The results are analyzed individually for each class and the result of each class comparison. The confusion matrix was drawn for each case.

TABLE 6 - THE AVERAGE ACCURACY OF CLASSIFICATION INTO TWO, THREE, AND FOUR CLASSES, USING RED, GREEN, BLUE AND GRAY CHANNELS (TRAINING AND TEST SET WERE RANDOMLY SPLIT).

Color component	Two classes	Three classes	Four classes
Red	100	99.86	95.60
Green	100	99.86	99.75
Blue	100	99.93	97.47
GRAY	100	99.76	99.20

It also shows that the best raw color channel is Green, which obtained an average accuracy of 100%, 99.86%, and 99.75% for the two, three, and four class classification, respectively. From these results it is concluded the level of tiredness or physical exercise intensity is reflected best by the Green color channel.

The accuracy and loss history during training with 75 epochs is shown in Figure 31 (left) and 31 (right) respectively. Only the plot of the Green channel is shown, due to having the best average prediction accuracy among all other color channels.

4.4.3.1. Training Accuracy and Loss History

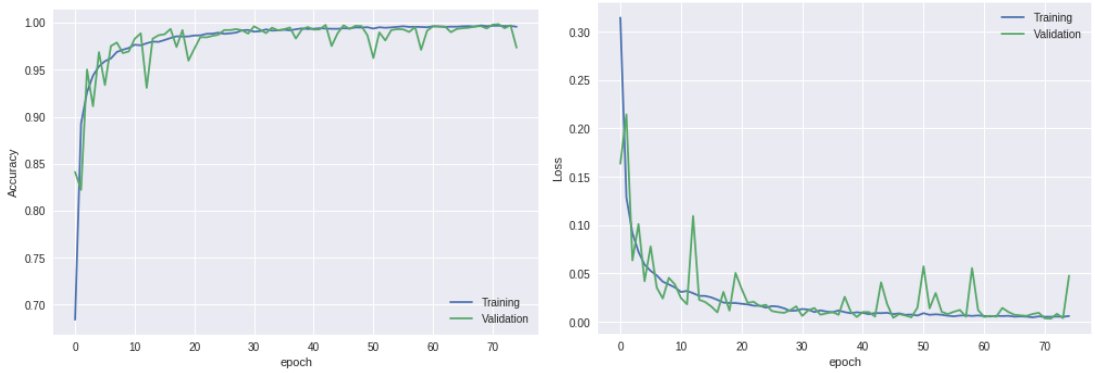


FIGURE 31 - TRAINING AND VALIDATION ACCURACY AND LOSS VS. EPOCH OF GREEN COLOR FOR FOUR CLASS CLASSIFICATION.

The classification accuracy of individual class for each color channel is presented in Table 7, where the training and test set were randomly split. In most of the cases, the classification accuracy is over than 99%. The result also showed that the classification accuracy of the last class is superior to the other classes. This leads us to conclude that fully tired faces are quite distinct from the other classes.

TABLE 7 - CLASSIFICATION ACCURACY OF EACH CLASS IN THE CLASSIFICATION OF PHYSICAL EXERCISE INTENSITY INTO FOUR CLASSES.

Color component	Classes	Accuracy	Average
R	C1	89.6	95.6
	C2	98.9	
	C3	95.2	

	C4	99.7	
G	C1	99.9	99.75
	C2	99.2	
	C3	99.9	
	C4	100	
B	C1	93.9	97.47
	C2	99.5	
	C3	97.8	
	C4	98.9	
GRAY	C1	98.7	99.20
	C2	98.5	
	C3	99.6	
	C4	100	

4.4.3.2. K-fold Cross Validation

The result of average accuracy and standard deviation of each fold by using a 10-fold cross validation is shown in Table 8.

TABLE 8 - AVERAGE ACCURACY AND STANDARD DEVIATION OF 10-FOLD.

Color Model	Accuracy (%)	Standard Deviation
R	0.95	0.0047
G	0.99	0.0054
B	0.96	0.0352
GRAY	0.98	0.0058

In cross fold validation also, the best color channel is Green, which provides a 99% average accuracy with a standard deviation of 0.0054 easily accepted values for real time application. R and B component provide poor result, meaning those channels suffers less variation with tiredness or higher heart rate when compared to Green channel.

4.4.3.3. Leave-One-Subject-Out (LOSO)

The physical exercise intensity was predicted for each participant using the observations (images) from all other subjects. The true leave-one-subject out likely offers

the best accuracy of generalizability of the classifies (Montoye et al., 2018). The accuracy of leave-one-subject-out method for each class is presented in Table 9.

TABLE 9 – AVERAGE ACCURACY OF LEAVE-ONE-SUBJECT-OUT CROSS VALIDATION FOR TWO, THREE, AND FOUR CLASSES.

Color component	**Two classes**	**Three classes**	**Four classes**
Red	86.52	79.52	59.85
Green	89.78	81.56	65.42
Blue	81.95	76.54	61.25
GRAY	84.48	78.26	62.54

As compared with the 10-fold cross validation, the leave-one-subject-out cross validation presented a very poor result. In the 10-fold cross validation, the average accuracy is almost 99% in the best case for four class classification, whereas in the leave-one-subject-out cross validation the result is only 65%. The reason behind this poor result is likely due to cultural and expressiveness differences between the subjects. The results can be improved by increasing the number of subjects.

4.4.3.4. ROC Curve

The ROC curve of each color space is shown in Figure 32, where the performance is represented in the area under the curve. Higher value of AUC implies a better accuracy of the classifier. In most of the classifier, area under curve is almost near to 1.

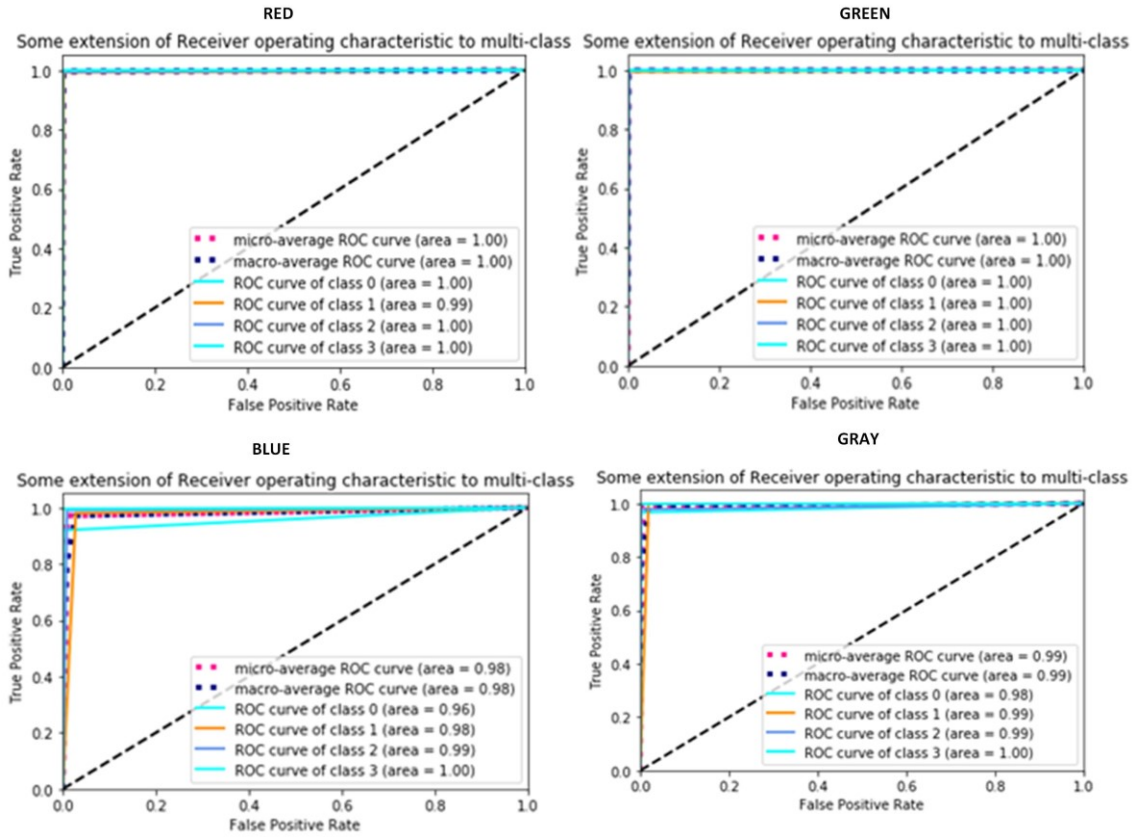


FIGURE 32 - ROC CURVE PLOT OF CLASSIFIER FOR ALL THE COLOR MODELS.

From the ROC curve, the best color space is Green. The area under curve in all the class is approximately 1. In general, the results are almost 100% accurate. On the other hand, the poorest color space is Blue.

4.4.3.5. Parameters (Precision, Recall, AUC, and F-measure)

Here, the performances are evaluated by using Precision, Recall, Area Under Curve (AUC), and F1-Measure. According to the results of these indicators, the best color channel is again Green and easiest class is a *class -four*. In overall, it is concluded that deep neural network is a good candidate technology to classify exercise intensity based on heart rate. In case of the green channel, the average value of each parameter is almost 1 or 100%.

TABLE 10 - PRECISION, RECALL, AUC, AND F1-MEASURE OF EACH CLASS OF FOUR COLORS COMPONENTS.

	BLUE					GREEN				
	C1	**C2**	**C3**	**C4**	**Average**	**C1**	**C2**	**C3**	**C4**	**Average**
Precision	0.9952	0.9215	0.9734	0.9995	0.9724	0.9995	0.9990	0.9950	1.0000	0.9984
Recall	0.9193	0.9805	0.9910	0.9954	0.9715	0.9995	0.9945	0.9995	1.0000	0.9984
AUC	0.9589	0.9763	0.9910	0.9976	0.9810	0.9997	0.9971	0.9989	1.0000	0.9989
F1-Measure	0.9558	0.9500	0.9821	0.9974	0.9713	0.9995	0.9967	0.9972	1.0000	0.9984

	RED					GRAY				
	C1	**C2**	**C3**	**C4**	**Average**	**C1**	**C2**	**C3**	**C4**	**Average**
Precision	0.9941	0.9920	0.9985	0.9990	0.9959	0.9985	0.9478	0.9954	1.0000	0.9854
Recall	0.9966	0.9925	0.9945	1.0000	0.9959	0.9663	0.9980	0.9774	0.9980	0.9849
AUC	0.9973	0.9949	0.9970	1.0000	0.9973	0.9829	0.9898	0.9880	0.9880	0.9872
F1-Measure	0.9954	0.9923	0.9965	0.9995	0.9959	0.9821	0.9722	0.9863	0.9990	0.9849

4.4.3.6. Confusion Matrix

Here, the class wise accuracy is evaluated. Figure 33 shows the confusion matrix of each color space with same classifier. From this tool also, the best color space which varies with tiredness is Green which has almost 100% result.

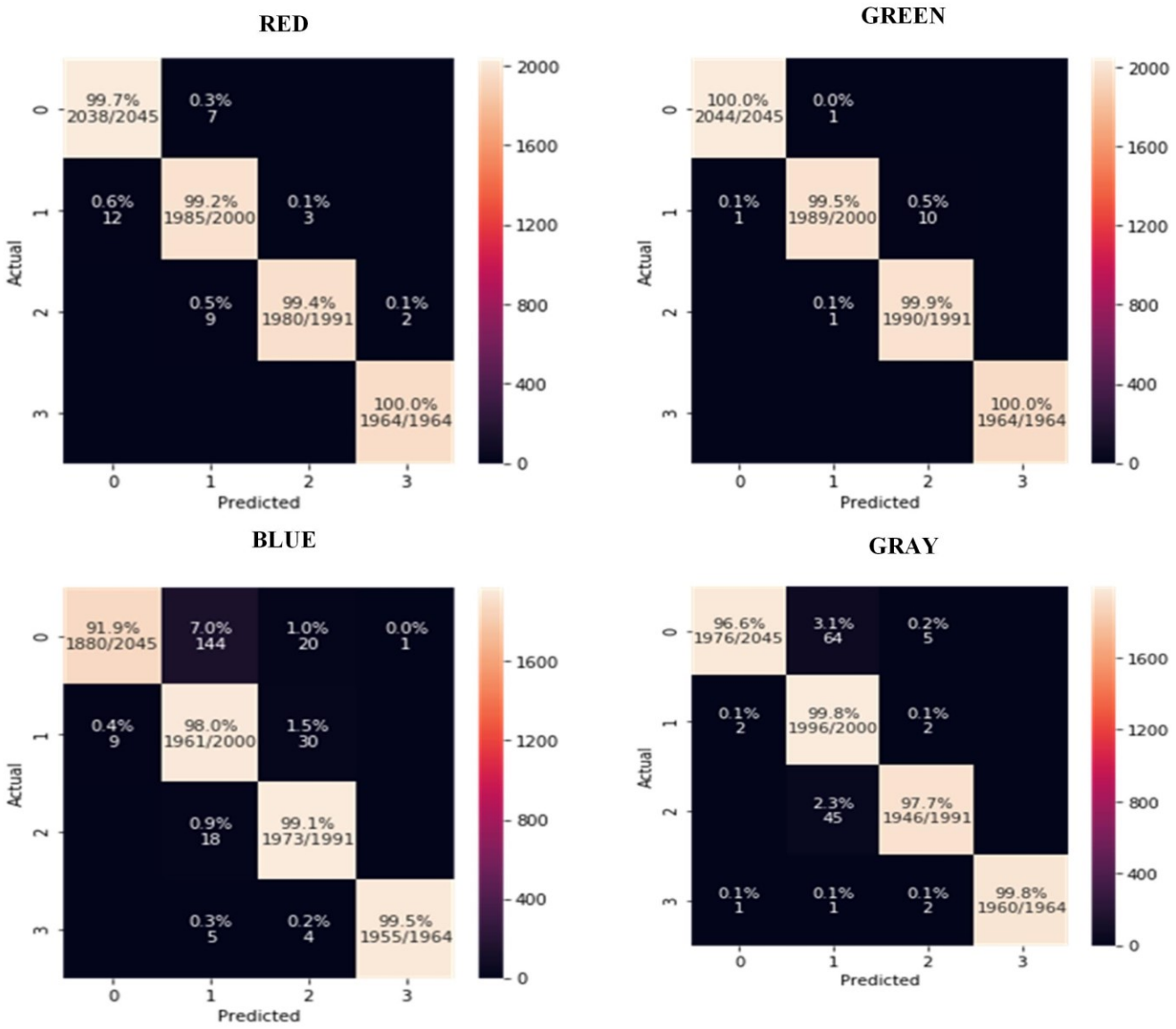


FIGURE 33 - - CONFUSION MATRIX OF THE FOUR-CLASS CLASSIFICATION OF PHYSICAL EXERCISE INTENSITY USING ALL COLOR SPACE.

4.5. Discussion

In the recent years, the facial expression analysis is a widespread technique to measure psychological and physiological indicators (Sheaffer, Golden, & Averett, 2009; Wu, Fu, & Yang, 2012). The facial expression is often analyzed by extracting and analyzing feature points (Alugupally, Samal, Marx, & Bhatia, 2011; Jeong & Ko, 2018;

Ko, 2018). For a few decades, many approaches have been presented in state-of-art articles to monitor physical exercise, with most of them being based on physiological parameters monitoring whether to use contact or contactless sensor technology. As mentioned, facial feature points analysis has been shown to have a high level of co-relation to physical exercise intensity (Khanal et al., 2018; Khanal et al., 2018; Khanal et al., 2018; Miles et al., 2017). To our knowledge, this is the first study to classify faces with various levels of exercise intensities based of feature point analysis. In the present study, a new approach is presented in the field of exercise monitoring using computer vision techniques based on facial expression analysis.

The previous studies (Chen et al., 2017; Marco C. Uchida, 2018) showed the significant level of facial expression change with respect to physical exercise intensity and our results of exercise intensity classification based on facial expression agrees with previous studies. Our findings suggest that feature points analysis has the potential to classify exercise intensity. The results of many state-of-art articles support that better ensemble algorithms have better accuracy than others (Gredell et al., 2019; Zhang et al., 2017). The present study indicates the classification of exercise intensity using ExtraTree classifier provides the best accuracy score among others.

The classification of exercise intensity was also carried out using deep neural networks, which presents more accurate results than traditional machine learning algorithms. We explored a deep learning technique to classify faces by specifying the class or level of the exercise intensity. We achieved very high accuracy classification accuracy in the four classes. The proposed architecture was trained and tested using 2D image or only a single-color space instead of feeding the network with the raw RGB image. One limitation of deep learning is the speed., The deep learning itself is a slow process, and 3D RGB images make the training process slower. Therefore, instead of using a 3-channel images, single channel is good enough to classify. This technique was used to classify only into two classes, whether a person gets tired or not. We extended the state-of-art studies by classifying exercise intensity into more than two classes.

The secondary aim of the present study was to investigate which color channel best reflects the level of exercise intensity or tiredness. The results agreed with some of the state-of-art studies, however, to our knowledge, no existing article research the impact of an individual color channel. Some of the studies used an individual color channel to detect skin (Chaves-González, Vega-Rodríguez, Gómez-Pulido, & Sánchez-Pérez, 2010), face (Shih & Liu, 2005), and facial expression (Lajevardi & Wu, 2012) as well. Our results are in contrast with the result of (Chaves-González et al., 2010) where we got the best classification result from Green channels, while they found a Red color channel more accurate for skin detection results. In this study, we determined which color channel is the most informative to classify the intensity. With reference to (Shih & Liu, 2005), we obtain green channel as the best channels to identify exercise intensity.

Various classifier evaluation techniques are also used to validate the results and to compare each to four input features. Two methods of cross validation are used to split of the dataset and confirm the results can be generalized for new datasets. In case of the 10-fold cross validation, the results have up to an 100% accuracy and a very small standard deviation, which indicate the classification of exercise into four classes is quite accurate and can be extended into more classes. However, with very few subjects, the 10-fold validation cannot be generalized for new subjects. Therefore, we also used the leave-one-subject out validation, but it presented poorer results than 10-fold cross validation.

4.6. Limitations

There were some limitations to this study. (1) The dataset was prepared only with twenty subjects with age ranging from 18-35 years and as such it cannot be generalized to the entire age group. (2) The experiments were carried out only with European people, therefore, the results for other ethnic groups may vary. (3) Only the geometric location of facial feature points was analyzed thus the accuracy could be increased if we consider more parameters such as color, voice, etc.

4.7. Conclusion and Future Work

From all the experiments, the facial image processing using feature point analysis can be a candidate method to measure exercise intensity in four levels. Considering the thirteen machine learning algorithms, Extra Tree algorithm has provided the best results. Preliminary results suggest that distinguishing more levels of exertion might be very useful using more variables. Future work can be extended by using more features such as more color models, eye blinking rate, movement of the head etc. In addition, using more subjects could benefit the external validity of the obtained models. In this thesis exercise intensity levels were classified only up to four classes. It is suggested to classify a greater number of classes in the future work.

Based on various experiments with various types of image datasets, the deep learning approach for exercise intensity classification based on facial expression can be a potential method to classify exercise intensity in two, three, four or more levels. In the case of a two-class classification, the accuracy rate was 100% and even for the three-class classification, it was also around 99%. From all the experiments, it can be concluded the best color channel for the raw input image is Green, in terms of its classification accuracy. The training and testing dataset were randomly prepared from the same subjects; therefore, this approach is more appropriate for personalized physical exercise monitoring.

The experiments were done with only twenty subjects with little diversity in age and origin. To generalize this model, we can train the model with more diversity and a greater number of subjects, in order to improve the test accuracy. Considering that the training and testing were performed from the same subjects, this approach might be more appropriate for personalized exercise monitoring systems, where the system can be trained from the same subject with image datasets taken during various exercise sessions.

5. REGRESSION MODEL FOR PREDICTION OF PHYSICAL EXERCISE INTENSITY

This chapter presents a computer vision technique to track facial color change during exercise to detect levels of exertion at any instant using facial color analysis. Five color models RGB, HSV, YCbCr, Lab, and YUV color models were used to find out the correlation of color component and heart rate changes during exercise. The study presented in this chapter not only found out the correlation of facial color changes with heart rate, but also found out the best color model to predict the level of physical exertion. The data were analysed using Pearson correlation coefficient, Bland-Altman plots, and regression.

5.1. Background

The current technological advances have improved the process of quantifying exercise intensity in sports sciences research. Monitoring physical exercise is essential in the process of planning, applying, and controlling loads for performance optimization and health (Duking, Achtzehn, Holmberg, & Sperlich, 2018; Nelson et al., 2007; Paterson & Warburton, 2010). The physiological responses to exercise depending on the individual and on the task, once exertion evolves in different ways during physical exercise (Borg, 1998). In steady-state continuous exercise, the level of physiological exertion is very low at the beginning and increases linearly with exercise intensity. However, in cases where intensity rises above the ventilatory and/or anaerobic threshold, the increase in the exertion is exponential. The changes of cardiovascular status during exercise and, broadly, the entirety of physiologic function, is reflected in heart rate responses (Dooley et al., 2017; Hensen, 2017), although body temperature, blood pressure, blood lactate concentration can also quantify internal loads (Miles et al., 2017). An important part of the research related to the measurement of physical exercise intensity relies on the extraction of physiological parameters as heart rate (Aoki et al., 2015; Kiviniemi et al., 2015; Nakamura et al., 2015), as one good measure of internal responses to exercise.

Digital images can be represented by the intensity of color components present in each pixel through a range of various color channels. Recent research efforts have described various color models rather than RGB (Red; Green; Blue) which may provide effective information for facial image processing such as face and emotion recognition, face detection, etc. Wang showed that the performance of CIELab, CIELub, and Tensor Independent Color Space (TICS), are better than those of RGB or gray representation to recognize a micro-expression related to emotion (Wang, Yan, Li, Zhao, & Fu, 2014; Wang et al., 2015). Such efficacy of different color models than RGB and gray for estimating levels of exertion along physical exercise is still unknown. The literatures have shown that facial color variation is related to the increase of exercise intensity and tiredness from physical exertion subject to redder the skin (Hasan, Hossain, Thakur, & Podder, 2014). Dynamic appearance model of skin color built from in vivo and real time

measurement of melanin and hemoglobin concentration has also been described to change according high-intensity exercise (Jimenez et al., 2010). Resting heart rate measurements using R, G, and B color channels in facial color analysis have also been related (More et al., 2019; Poh et al., 2010; Poh et al., 2011). These evidences point out towards the potential of using facial color analysis to estimate physical exertion levels, however, there are no studies advancing towards the dynamics of gold standard measures as heart rate and perceived exertion, across physical exercise. Thus, in this chapter, we aim to evaluate the efficacy of various color models to facial color tracking in relating to heart rate dynamics during fatiguing exercise using regression models.

5.2. Proposed Method

5.2.1. Dataset Preparation

The video data collection was the same as described in chapter four, only the preparation of data for the experiment is different. In this section, the preparation of data is presented. The summary of all the subjects' details who participated in data collection are presented in Table 11.

TABLE 11 - PHYSIOLOGICAL DATA ABOUT THE SUBJECTS PARTICIPATING IN THE DATA ACQUISITION PROTOCOL. THE INITIAL HEART RATE WAS RECORDED IN THE BEGINNING OF THE EXERCISE (AFTER 5 MINUTES OF WARM UP EXERCISE) AND THE FINAL HR WAS RECORDED AT THE END OF THE EXERCISE.

Subject ID	Gender	Age (years)	Weight (KG)	Height (cm)	Initial HR (bpm)	Final HR (bpm)	Duration (mm:ss)
Subject1	Male	22	64.2	172	93	191	9:30
Subject2	Male	33	66.9	177	70	180	16:00
Subject3	Female	19	64.2	177	87	191	9:05
Subject4	Male	36	83.2	182	91	191	15:00
Subject5	Female	24	66.6	170	97	184	9:20
Subject6	Female	29	47	157	114	193	8:00
Subject7	Male	33	83	186	101	180	16:00

Subject8	Male	22	90.7	195	101	201	12:00
Subject9	Male	24	87.3	194	115	188	11:00

5.3. Data Processing

The recorded video was processed off-line to track the average color intensity of a small patch in the subject's forehead during the exercise. The block diagram of the proposed system is shown in Figure 34. Although the frame rate of recorded video was 25 Hz, it was resampled at 1 Hz, extracting one frame for each second.

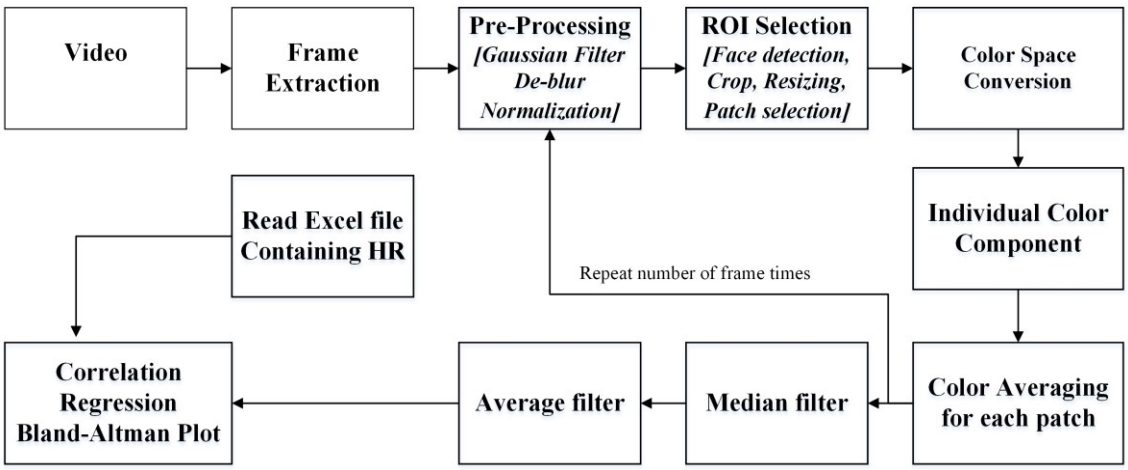


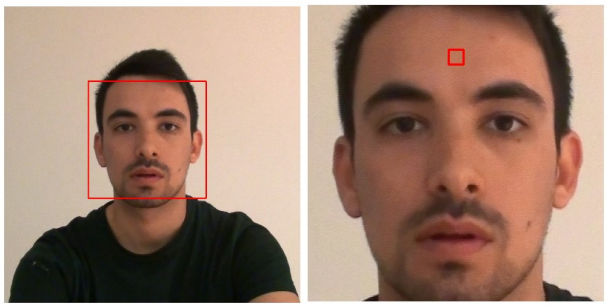
FIGURE 34 - SYSTEM BLOCK DIAGRAM TO ANALYSES CONTINUES EXERCISE INTENSITY USING CORRELATION, REGRESSION, AND BLAN-ALTMAN PLOT.

5.3.1. Pre-processing

Due to the head movement during cycling, some frames extracted from the video recording are blurred. To overcome this problem Wiener–Hunt deconvolution algorithm was applied in each frame (Orieux, Giovannelli, & Rodet, 2010). After the de-blurring operation, a Gaussian filter was applied for noise reduction.

5.3.2. Face Detection and ROI Selection

The subject's face is detected in the image by Viola and Jones face detection algorithm (Viola & Jones, 2001). The facial region is cropped and resized to 400 ×400 pixels (Figure 35a). Based on the results reported in the literature, forehead patch provides the highest accuracy of heart rate measurement among other patch locations, even better than whole face (Hassan et al., 2016; Ramirez, Fuentes, Crites, Jimenez, & Ordonez, 2014). Therefore, an image patch of 16 x 16 pixels is cropped from the whole face, in the lower part of forehead patch which is considered as a color tracking patch. This is a region in the face with less source of noise such as e.g. hair (Figure 35b).



a. b.

FIGURE 35 - FACE DETECTION AND PATCH LOCATION.

5.3.3. Normalization

The images were recorded in a room with lighting background from the AC florescent bulb, where the frequency of supplying electricity was 50 Hz, therefore, affecting the image brightness. The solution to this issue was found in the use of color normalization to eliminate brightness fluctuation. Consider an RGB image I (M x N x c), where M = width, N = height and c are the color component. The normalized value of each color component for each pixel is calculated by expression 5.1.

$$Inorm\ (i,j,c)\ = \frac{I[i,j,c]}{\sum_{c=1}^{3}(I[i,j,c])}, \qquad i = 1..M, and\ j = 1..N \tag{5.1}$$

Where c = 1, 2, 3 corresponding to red (R), green (G), and blue (B) component of image I.

From the above expression, it is common that

$$\sum_{c=1}^{3} I_{norm}[i, j, c] = 1 \quad (5.2)$$

5.3.4. Color Space Conversion

The RGB color model is an additive color system. The combination of the three colors results in visible colors sensed by the human eye. Although the RGB color model is the most common for image processing, we added in the study, other four different models for tracking color components to compare its efficiency: Hue-Saturation-Value (HSV), Luminance -Chroma: Blue-Chroma: Red (YCbCr), Lightness-Blue: Red, Blue: Yellow (Lab), and YUV. After the normalization of the original RGB image, it was converted to each of the color models, described as follows.

HSV color space

This color space is more closely the way human vision perceives color attributes (Paschos, 2001). The Hue, Saturation, and Value are the three components of the HSV model, where Hue and Saturation are chrominance components expected to be better for human color perception than RGB color space and Value is the component of brightness and luminance. The measured color differences are proportional to the human perception. This color space is specially used to reduce the influence of brightness of the images. The conversion of RGB images to HSV color space is implemented using expressions 5.3-5.11.

$$R' = \frac{R}{255} \quad (5.3)$$

$$G' = \frac{G}{255} \tag{5.4}$$

$$B' = \frac{B}{255} \tag{5.5}$$

$$Cmax = max\,(R', G', B') \tag{5.6}$$

$$Cmin = min(R', G', B') \tag{5.7}$$

$$\Delta = Cmax - Cmin \tag{5.8}$$

Hue (H) calculation:

$$H = \begin{cases} 0° & \Delta = 0 \\ 60° \times \left(\frac{G' - B'}{\Delta} mod 6\right), C_{max} = R' \\ 60° \times \left(\frac{B' - R'}{\Delta} + 2\right), C_{max} = G' \\ 60° \times \left(\frac{R' - G'}{\Delta} + 4\right), C_{max} = B' \end{cases} \tag{5.9}$$

Saturation (S) calculation:

$$S = \begin{cases} 0°, & C_{max} = 0 \\ \frac{\Delta}{C_{max}}, & C_{max} \neq R' \end{cases} \tag{5.10}$$

Value (V) calculation:

$$V = Cmax \tag{5.11}$$

Hue is defined as the pure spectrum colors measured in degree between 0° to 360° and it follows non-linear transformation of RGB. The variation in hue (H) represents only changes in color not the brightness. This may be more informative to detect the differences between the absorptions of red, green, and blue ambient light in the blood, since time-variation in the relative reflections will be translated to rotation by H. Given this, it is expected that this component could provide an accurate estimation or a best correlation coefficient with the heart rate.

YCbCr color space

It is a common color space used in digital photography which is encoded from the non-linear RGB signal. It is a simplicity and explicit separation of luminance and chrominance components. Y represents luminance, Cb is blue difference, and Cr is red difference chroma component. Luminance represents light intensity and chrominance represents color values. The conversion of RGB images to YCbCr color space is implemented using expressions 5.12-5.14.

$$Y = 0.299R + 0.587G + 0.114B \tag{5.12}$$

$$Cb = -0.169R - 0.331G + 0.5B \tag{5.13}$$

$$Cr = 0.5R - 0.419G - 0.081B \tag{5.14}$$

Lab color space

It is a perceptually uniform color space in which measured color differences are proportional to the human perception. The Lab color space is derived from XYZ color space. First, the XYZ components are calculated to then find its corresponding Lab values. L represents whiteness or brightness values (ranging from 0 to 100), whereas a

and b represent a red and blue component's perceived difference, respectively. The conversion of RGB images to Lab color space is implemented using expressions 5.15-5.17.

$$L^* = \begin{cases} 116 * \left(\frac{Y}{3Yn}\right) - 16, & \text{for } \left(\frac{Y}{Yn}\right) > 0.008856 \\ 903.3 * \left(\frac{Y}{Yn}\right), & \text{Otherwise} \end{cases} \quad (5.2)$$

$$a* = 500 * (f(-)) - f(\frac{Y}{Yn}) \quad (5.3)$$

$$b* = 200 * (f(\frac{Y}{Yn}) - f(-) \quad (5.4)$$

Where, $f(t) = (\frac{t1}{3}), \quad for\ t > 0.008856$

$$f(t) = 7.787 * t + 0.1379, \text{ otherwise}$$

Here, Xn, Yn and Zn are the tristimulus values of the reference white.

YUV color space

The YUV color space is defined in terms of one Luminance and two Chrominance components. The Luminance (Y) represents the brightness whereas the Chrominance represents the color values. This color space is quite similar to YCbCr. Historically, the YUV was used for analog encoding, whereas YCbCr was used for digital encoding. The conversion of RGB images to YUV color space is implemented using expressions 5.18-5.20.

$$Y = R * 0.2990 + G * 0.5870 + B * 0.1140 \quad (5.5)$$

$$U = R * -0.1687 + G * -0.3312 + B * 0.5000 + 128 \quad (5.6)$$

$$V = R * 0.5000 + G * -0.4186 + B * -9.0812 + 128 \quad (5.7)$$

The different color components for each model are shown in Figure 36.

5.3.5. Patch Averaging

The average patch ($\overline{y}_c$) value for each channel of the different color spaces is calculated in each frame, using expression 5.21.

$$\overline{y}_c(i) = \frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} y_c(m, n) \tag{5.28}$$

Where, i is color channel such that R, G, or B for RGB, H, S, or V for HSV and so on. M and N are the patch dimensions (16x16) and y_c (m, n) represents the pixel value in the normalized color channel i.

Applying expression 5.21 for each frame, results in a vector for each component with the length equal to the number of seconds in the video; therefore, for each color model, there are three vectors each one representing a color component.

5.3.6. Median Filter

The vectors of each component usually are noisy, thus, 1D median filter and moving average was applied to smooth the signal. The signal padding and signal reflection were applied before filtering operations to overcome the end-point effects in the filtered signal.

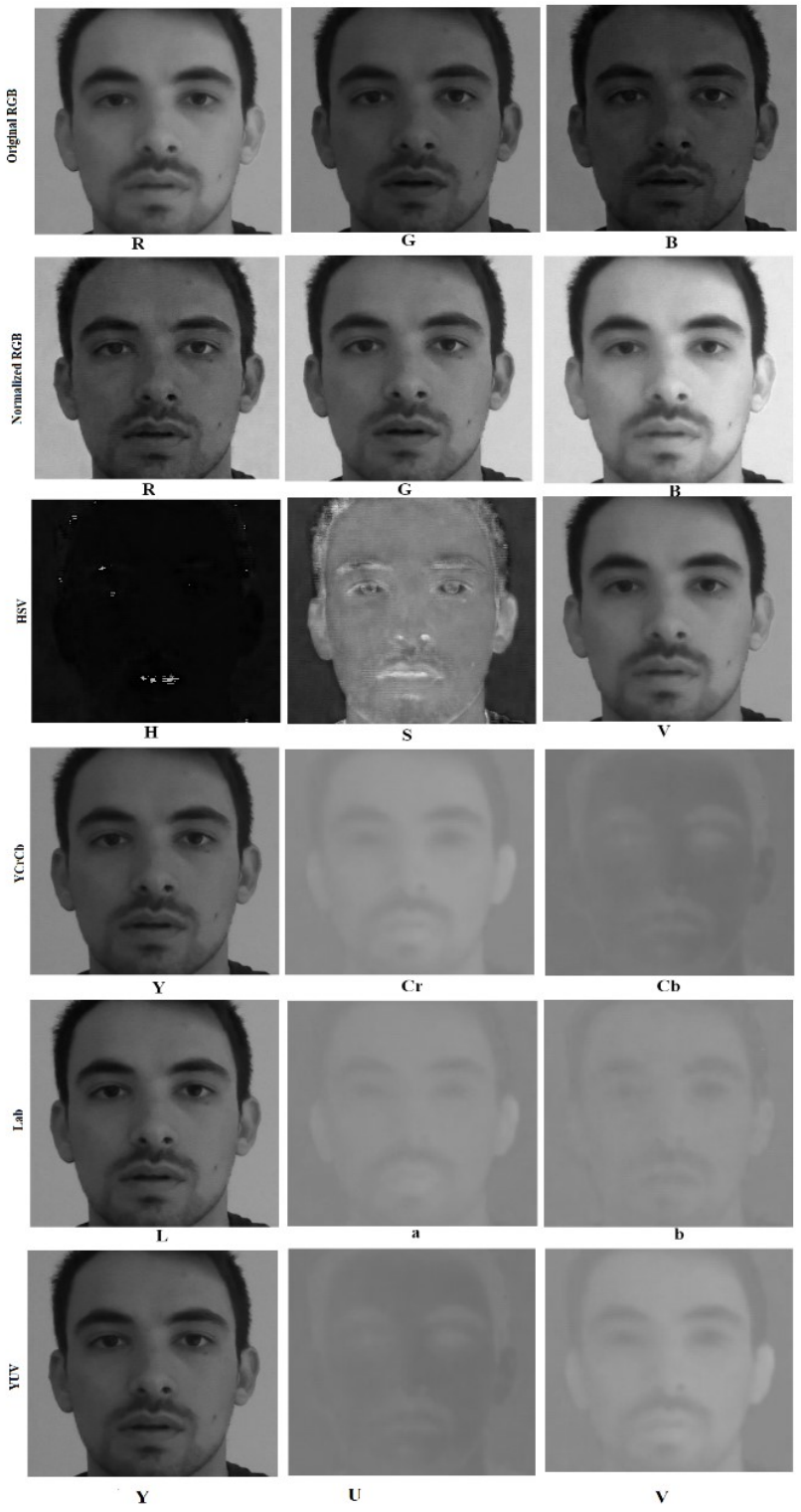


FIGURE 36 - THE FACIAL IMAGE REPRESENTED IN ORIGINAL RGB, NORMALIZED RGB, HSV, YCBCR, LAB, YUV COLOR SPACE.

5.3.7. Average Filter

To smooth the small peaks resultant from the median filter a moving average filter with a small *WindowSize* was applied. The *WindowSize* is the number of values considered to calculate the average value. The average value within a sliding window was calculated as expression 5.22.

$$y(n) = \frac{1}{WS}\big(x(n) + x(n-1) + .. + x\big(n-(WS-1)\big)\big) \tag{5.22}$$

Where *y(n)* is an average value, *x(n)* is a given value and WS is the window size considered for smoothing operation.

The result of median and the moving average filter is shown in Figure 37 for Red, Green and Blue components of RGB color space. This operation was repeated for each channel of HSV, YCbCr), Lab and YUV color spaces.

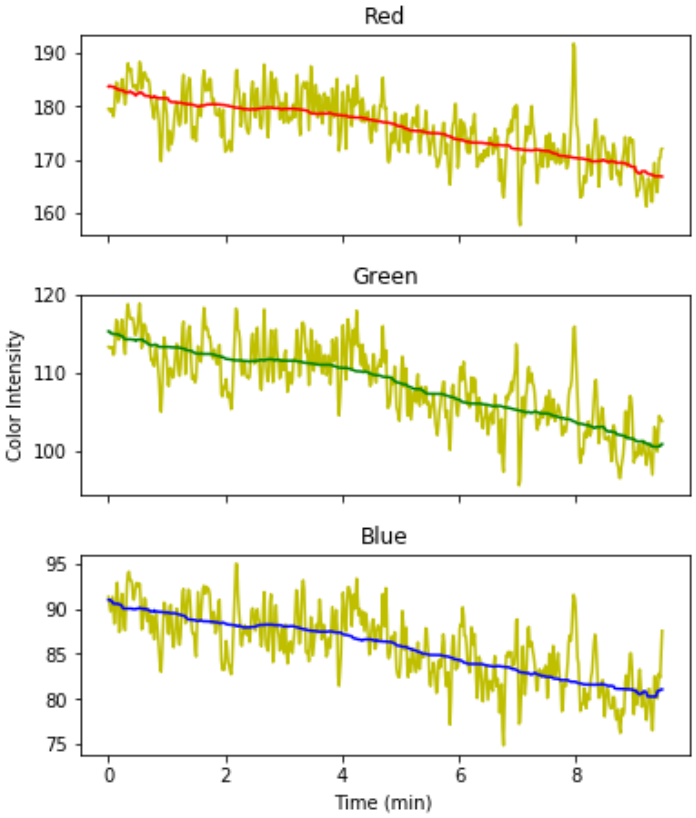


FIGURE 37 - FILTERED SIGNALS AFTER SMOOTHING OPERATION (MEDIAN FILTER AND AVERAGE FILTER) IN NORMALIZED RGB IMAGES.

5.4. Data Analysis

For all data analysis, both the heart rate and video recordings were re-sampled to 1Hz. The individual color component of each color space was analyzed separately, such as *RGB = [R, G, B], HSV = [H, S, V], YCbCr = [Y, Cb, Cr], Lab = [L, a, b1] and YUV = [Y1, U, V1].* At first, the heart rate and the color component were plotted against time to visualize the pattern of color intensity changes in accordance to heart rate changes for individual subjects. The Pearson correlation of heart rate and color components were calculated individually as well as for the whole sample.

Following, the linear regression methods were adopted to estimate the HR from the color intensity of the individual color channel (**univariate linear regression**) and combination of the individual color channels (**multivariate linear regression**) of all the five color spaces. These methods were used for individual subject (**individual regression model**) as well as all subjects at once (**global regression model**). The root mean square error (RMSE) was also calculated for both models to find the best predictor of the heart rate through the color components of each of the models.

Bland-Altman plot (level of confidence = 95%) was applied to test the degree of agreement between the different color models in the prediction of heart rate and the actual measure of heart rate (Bland & Altman, 1986; Giavarina, 2015) where the HR was predicted using **multivariate regression model** individually for each subject. All statistics were performed using dedicated codes written in Python programming V3.5.

5.5. Results

The main objective of this study was to analyze the relationship of heart rate and facial color changes during fatiguing exercise using regression model. The results of images pre and post-processing are shown in respective sections.

5.5.1. Heart Rate and Color Component vs Time

The heart rate and individual color component of all five color models of one random participant are plotted in Figure 38. The result indicates that facial color changes according to the effort performed during fatiguing exercise, and this variation correspond to changes in heart rate. The behavior of heart rate as a function of the color component is either linearly positive (*S, Cb, U*) or negative (*R, G, B, H, V, Y, Cr, L, a, b, Y, and V*), as it can be verified in Figure 38. Increase in heart rate is related to the increase in blood pump to face's vessels, thus, turning face skin redder. Therefore, increase in heart rate indicates heart pumps blood to vessels more often. However, we observe a decrease in the Red component as the heart rate increases. This result can be explained by the color normalization operation performed image pre-processing.

5.5.2. Correlation Between Heart Rate and Color Component

Individual's Pearson correlation coefficients between heart rate and color models are presented in Table 14 for all the participants in the study. Most of the predictors have negative linear correlation coefficients, although some are positively correlated. The positive value of correlation coefficient indicates that an increase in one variable increases other, whereas negative value of correlation coefficient indicates that an increase in one variable decreases other. Only the S component of HSV and U component of YUV has a positive correlation coefficient whereas all the others are negative correlation coefficient values. The strongest predictors of heart rates are R (-0.79 ±0.09) of RGB, V (-0.79±0.08)

of HSV, Y (-0.79 ±0.09) of YCbCr, and L (-0.79 ±0.08) of Lab, while the weakest predictors are H (-0.67±0.17) of HSV, b (-0.68±0.22) of Lab, and V (-0.68±0.20) of YUV. Most consistent color components are G of RGB, V of HSV, Y of YCbCr and L of Lab and Y of YUV with standard deviation (SD) value of 0.08 whereas least consistent color component are B or Lab with SD of 0.22 and V of YUV with SD of 0.20. Subject1 and Subject2 showed best values of negative correlation coefficient (-0.98) between heart rate and Gray and positive correlation coefficient (-0.98) between heart rate and S of HSV, whereas Subject2, Subject3, and Subject7 showed the worst values of correlation coefficient between heart rate and V of YUV (-0.53), U of YUV (0.49), and H of HSV (-0.50).

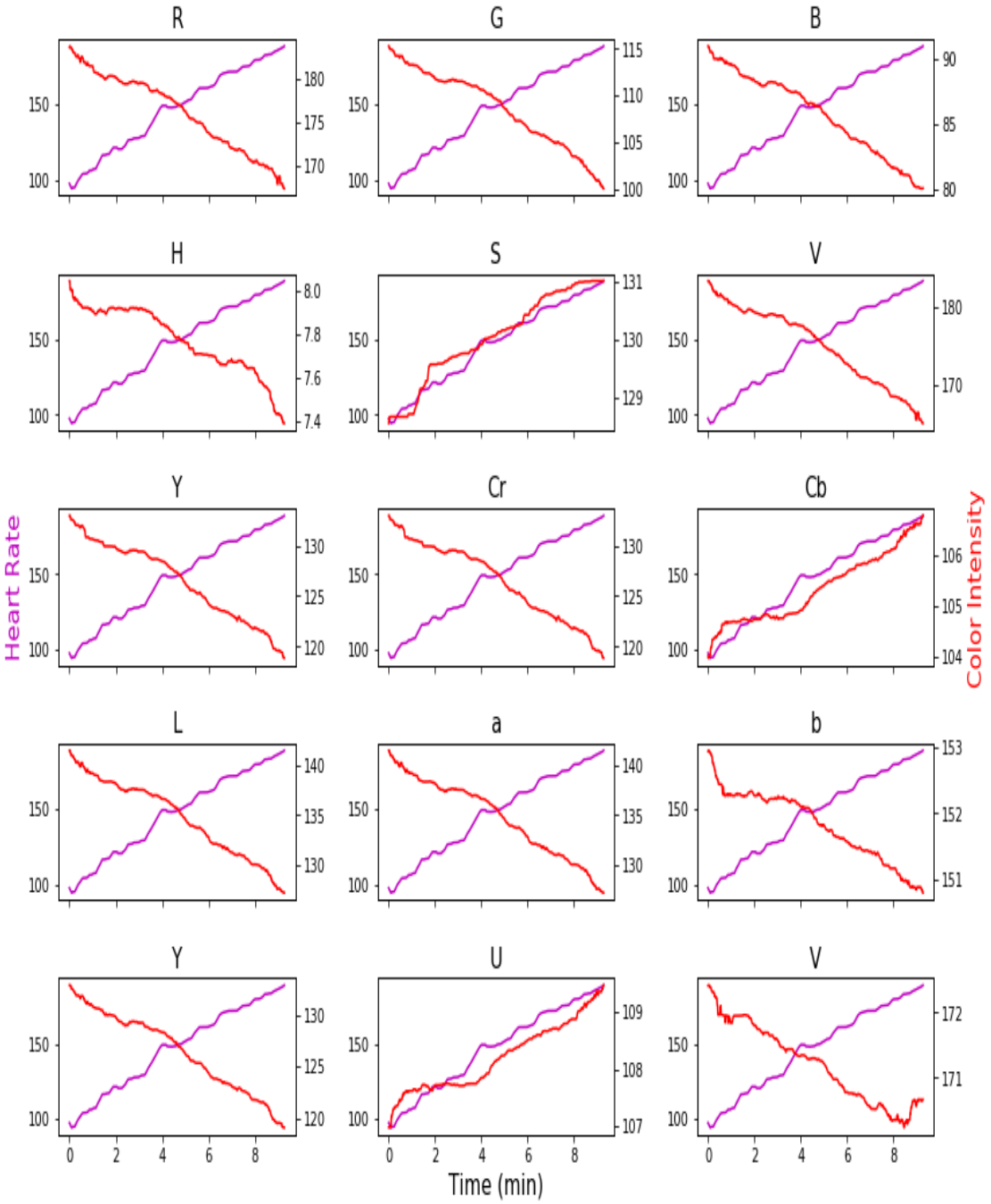


FIGURE 38 - PLOT OF AVERAGE PATCH VALUE FOR EACH COMPONENT OF RGB, HSV, YCBCR, LAB, AND YUV VS. TIME. THE LEFT Y-AXIS INDICATES THE HEART RATE IN BPM (BEATS PER MINUTE) AND RIGHT Y-AXIS INDICATES THE AVERAGE PATCH VALUE IN THE RESPECTIVE COLOR COMPONENT.

The global Pearson correlation coefficients of color and heart rate were calculated using the data of all participants at once. Results from the correlation of heart rate with

individual color components of each color models were: RGB (R:-0.13, G:-0.11, B:0.003), HSV (H:-0.07, S:-0.09, V:-0.12), YCbCr (Y:-0.13, U:-0.13, V:0.31), Lab (L:-0.15, a:-0.15, b:-0.26), and YUV (Y:-0.13, U:0.31, V:-0.03), suggesting low associations. The correlation coefficient of heart rate and color component for all the samples were very poor, suggesting that individual modelling might be a better approach to achieve higher precision.

TABLE 12 - CORRELATIONS BETWEEN GRAYSCALE IMAGE, EACH COLOR COMPONENT OF RGB, HSV, YCBCR, LAB, AND YUV AND HEART RATE FOR EACH SUBJECT (SUB1, SUB2 ETC. IN THE HEADING INDICATE SUBJECTS AND SD INDICATES STANDARD DEVIATION.

	Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Average	SD
Gray	-0.98	-0.98	-0.98	-0.75	-0.67	-0.78	-0.82	-0.78	-0.90	-0.75	0.12
R	-0.97	-0.98	-0.96	-0.97	-0.78	-0.94	-0.80	-0.75	-0.88	-0.79	0.09
G	-0.96	-0.84	-0.95	-0.86	-0.71	-0.89	-0.77	-0.83	-0.90	-0.76	0.08
B	-0.97	-0.73	-0.96	-0.68	-0.64	-0.84	-0.81	-0.91	-0.91	-0.73	0.12
H	-0.93	-0.86	-0.57	-0.82	-0.62	-0.95	-0.50	-0.75	-0.91	-0.67	0.17
S	0.98	0.63	0.82	0.78	0.94	0.94	0.73	0.91	0.92	0.78	0.12
V	-0.96	-0.95	-0.96	-0.97	-0.80	-0.93	-0.78	-0.79	-0.88	-0.79	0.08
Y	-0.97	-0.93	-0.97	-0.93	-0.75	-0.93	-0.80	-0.81	-0.90	-0.79	0.08
Cr	-0.97	-0.93	-0.96	-0.58	-0.75	-0.93	-0.79	-0.81	-0.84	-0.74	0.13
Cb	0.95	0.94	0.39	0.95	0.82	0.91	0.63	0.75	0.85	0.74	0.19
L	-0.97	-0.93	-0.96	-0.94	-0.74	-0.92	-0.79	-0.80	-0.89	-0.79	0.08
a	-0.97	-0.93	-0.95	-0.75	-0.74	-0.92	-0.79	-0.80	-0.89	-0.77	0.09
b	-0.95	-0.95	-0.27	-0.96	-0.84	-0.91	-0.61	-0.73	-0.82	-0.68	0.22
Y	-0.97	-0.93	-0.97	-0.94	-0.75	-0.92	-0.79	-0.81	-0.89	-0.79	0.08
U	0.94	0.93	0.40	0.95	0.82	0.90	0.63	0.75	0.84	0.73	0.18
V	0.97	-0.53	-0.81	-0.59	-0.88	-0.93	-0.77	-0.87	-0.82	-0.68	0.20

5.5.3. Individualized Linear Regression Model

Two types of individual linear regression models were implemented to estimate the heart rate from facial color intensity. The individual components of each color space were the input variables for the univariate linear regression (including grayscale). For

multiple linear regression the three independent components of each color space were considered as input. The linear regression equation of univariate variable is shown in Table 13 where a subject was randomly selected from nine subjects to present the results. The slope (m), y-intercept (b), linear regression line, root mean square error for all the color components which presents a high individual variation.

TABLE 13 - LINEAR REGRESSION OF HEART RATE WITH INDIVIDUAL COLOR COMPONENT OF A SUBJECT WITH ALL FIFTEEN COLOR COMPONENTS OF FIVE COLOR SPACES.

Color Component	Linear Regression Equation	Root mean square error
Gray	-14.61Gray-53.61	7.8
R	-3.51R+656.3	12.03
G	-3.95G+563.02	10.91
B	-4.47B+538.8	10.48
H	26.56H-160.2	13.16
S	15.54S-1511.79	5.03
V	-3.56V+664.42	12.03
Y	-3.87Y+592.66	11.21
Cb	54.88Cb-5991.83	15.99
Cr	-48.9Cr+7425.47	17.80
L	-3.76L+601.77	11.34
a	-36.82a+5346.4	24.47
b	-55.85B1+8187.26	15.34
Y1	-3.87Y1+592.61	11.21
U	61.92U-6905.22	16.40
V1	-41.22V1+6482.41	17.59

As we can see in Table 14, the room mean square error value of multiple linear regression using all the three individual components gives better result than univariate linear regression. The result also shows that the HSV (5.12±2.78), YCbCr (5.76±2.65), Lab (5.93±2.84), and YUV (5.77±2.80) has lower RMSE value than RGB (9.26±3.50). The F-values and R2 values from the regression analysis of the participants individually are presented in Table 15. In all the cases, p-values are statistically significant ($p<0.05$).

The results indicate that heart rate is strongly related to color components, although there are tenue individual differences.

TABLE 14 - THE ROOT MEAN SQUARE ERROR (RMSE) VALUES USING LINEAR REGRESSION BETWEEN EACH COLOR COMPONENT AND HR AND ALSO MULTIPLE LINEAR REGRESSION OF RGB, HSV, YCBCR, LAB, AND YUV COLOR MODEL. (SUB1, SUB2 ETC. IN THE HEADING INDICATE SUBJECTS AND SD INDICATE.

		Sub1	Sub2	Sub3	Sub4	Sub5	Sub6	Sub7	Sub8	Sub9	Average	SD
Univariate Regression	**Gray**	10.73	11.02	5.90	3.92	8.00	14.92	13.31	14.16	7.80	9.97	3.82
	R	12.03	5.24	5.50	5.80	11.07	14.31	11.94	9.09	8.32	9.26	3.50
	G	10.91	14.48	6.87	6.28	11.99	10.68	9.77	13.03	8.56	10.29	2.83
	B	10.48	19.09	6.18	16.84	10.82	8.29	7.07	5.62	8.69	10.34	4.97
	H	13.16	13.26	19.43	5.01	20.82	19.90	17.89	7.01	6.17	13.63	6.03
	S	5.03	16.65	15.17	7.09	15.90	11.46	8.39	3.74	9.13	10.29	5.09
	V	12.03	6.95	5.90	6.03	11.41	12.90	12.59	10.79	8.32	9.66	3.01
	Y	11.21	9.59	5.77	6.47	11.10	11.83	8.53	12.60	8.50	9.51	2.52
	Cb	15.99	8.98	22.21	5.48	15.86	15.08	20.24	6.00	7.23	13.01	6.28
	Cr	17.80	23.71	8.43	9.82	12.48	11.85	21.09	9.81	6.62	13.51	5.74
	L	11.34	9.35	6.28	6.58	11.29	12.13	9.16	10.69	8.40	9.47	2.20
	a	24.47	24.23	10.39	14.53	12.94	11.07	18.29	16.89	5.08	15.32	5.48
	b	15.34	6.88	23.37	5.34	16.19	16.33	22.90	5.13	6.87	13.15	7.40
	Y1	11.21	9.51	5.56	6.55	11.25	11.73	8.53	12.70	8.50	9.50	2.57
	U	16.40	8.42	22.01	5.29	16.00	15.43	19.18	5.97	7.28	12.89	6.24
	V1	17.59	23.14	7.59	9.63	12.52	11.83	19.38	9.75	6.57	13.11	5.48
Multiple regression	**RGB**	12.03	5.24	5.50	5.80	11.07	14.31	11.94	9.09	8.32	9.26	3.50
	HSV	3.68	4.23	3.37	2.79	10.85	7.79	6.02	3.69	3.64	5.12	2.78
	YCbCr	4.86	5.24	3.49	3.70	9.98	9.60	5.91	3.70	5.37	5.76	2.60
	Lab	7.80	4.85	3.24	3.27	10.03	9.97	5.19	3.98	5.02	5.93	2.84
	YUV	4.67	5.14	3.24	3.18	9.99	10.07	6.33	3.83	5.47	5.77	2.80

A scattered plot with the predicted heart rate vs recorded heart rate is shown in Figure 39. We found that the best heart rate predictor is the S of HSV color space (RMSE of 5.03), whereas the worst predictor was the L color space (RMSE of 24.47). These results indicate that, to increase the accuracy of regression model for the heart rate prediction, computer vision techniques might perform an online recognition of individual's face color and adjust the best color component model to him.

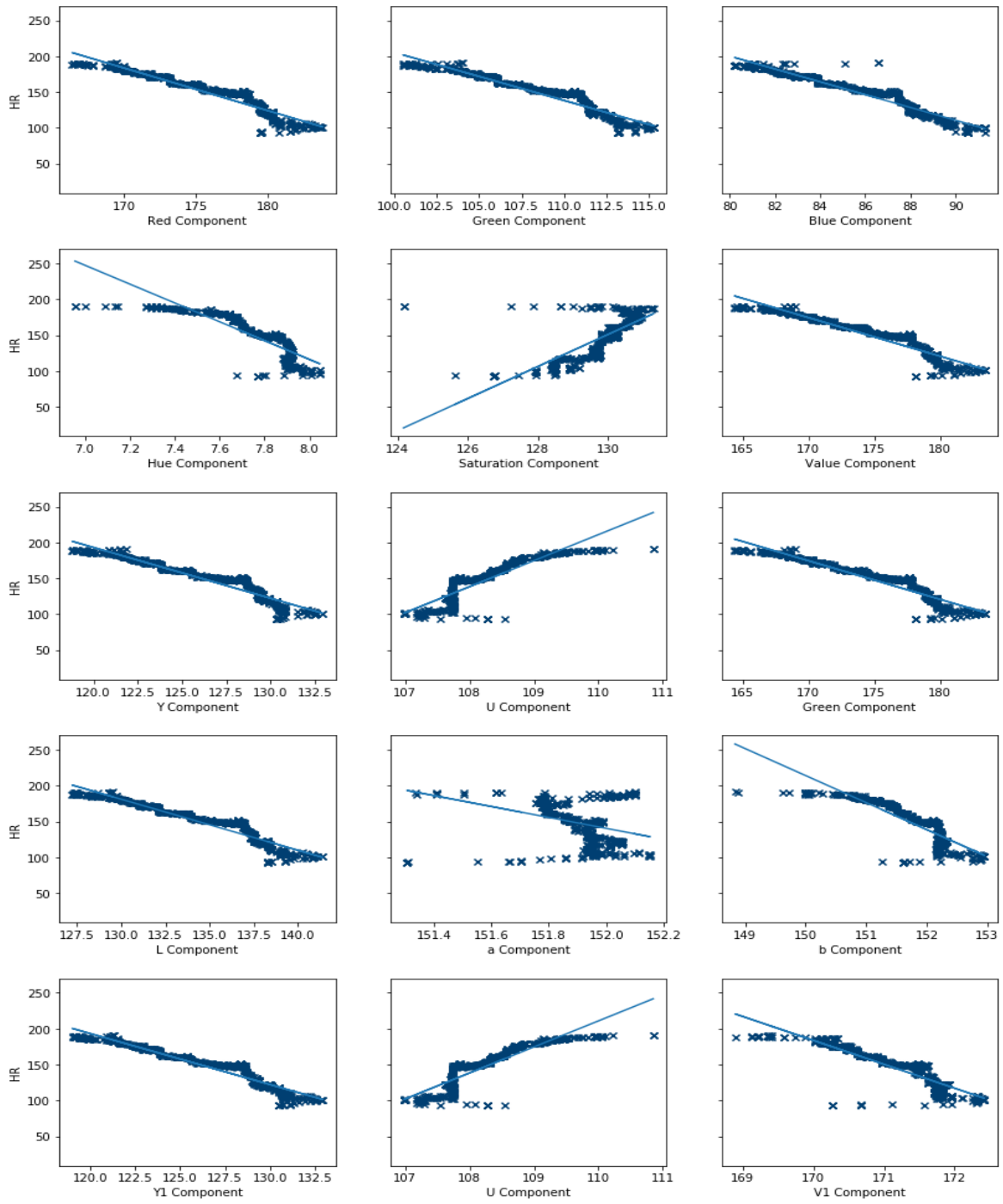


FIGURE 39 - THE SCATTERED PLOT AND LINEAR REGRESSION LINE USING BEST FIT METHOD. THE Y-AXIS IS HEART RATE AND INDIVIDUAL COLOR COMPONENT IS IN X-AXIS. A RANDOMLY SELECTED SUBJECT IS CONSIDERED FOR PLOTTING THE SCATTERED PLOT.

The F-value and R-square values were also calculated individually for each subject where HR was predicted using the multiple regression model as presented in Table

15. The multiple regression in each color was also performed using data from all participants. The overall model was significant in most of the color models, however, the standard error was substantially high. *RGB: [F(3,5787) = 322.33, p<0.005, R2 = 0.14], HSV: [F(3, 5787) = 202.49, p<0.005, R2 = 0.09], YCbCr: [F(3, 5787) = 357.24, p<0.005, R2 = 0.11], Lab: [F(3,5787) = 347.47, p<0.005, R2 = 0.11], YUV: [F(3, 5787) = 284.53, p<0.005, R2 = 0.13], and Gray: [F(3, 5787) = 41.13, p<0.001, R2 = 0.007].*

TABLE 15 - THE F-VALUES AND R-SQUARED VALUES CALCULATED INDIVIDUALLY USING MULTIPLE REGRESSION FOR ALL FIVE COLOR MODELS (RESULTS WERE ALWAYS P<0.05).

	Subject 1	**Subject 2**	**Subject 3**
RGB	F(3,570) = 2296.60, R^2 = 0.92	F(3,960) = 15829.2, R^2 = 0.98	F(3,540) = 9342.20, R^2 = 0.98
HSV	F(3,570) = 6201.03, R^2 = 0.97	F(3,960) = 14590.4, R^2 = 0.97	F(3,540) = 6817.96, R^2 = 0.97
YCbCr	F(3,570) = 4148.99 , R^2 = 0.93	F(3,960) = 3931.68, R^2 = 0.89	F(3,540) = 13259.7, R^2 = 0.98
Lab	F(3,570) = 1193.66, R^2 = 0.80	F(3,960) = 8934.8, R^2 = 0.95	F(3,540) = 17834.8, R^2 = 0.99
YUV	F(3,570) = 4916.11, R^2 = 0.96	F(3,960) = 10187.6, R^2 = 0.97	F(3,540) = 12183.7, R^2 = 0.99

	Subject 4	**Subject 5**	**Subject 6**
RGB	F(3,485) = 1118.91, R^2 = 0.87	F(3,960) = 849.42, R^2 = 0.72	F(3,720) = 3372.67, R^2 = 0.93
HSV	F(3,485) = 1118.00, R^2 = 0.87	F(3,960) = 704.14, R^2 = 0.68	F(3,720) = 2515.3 , R^2 = 0.91
YCbCr	F(3,485) = 1459.74, R^2 = 0.85	F(3,960) = 1358.29, R^2 = 0.74	F(3,720) = 1038.5, R^2 = 0.74
Lab	F(3,485) = 1341.29, R^2 = 0.84	F(3,960) =1252.65, R^2 = 0.72	F(3,720) = 1123.67, R^2 = 0.75
YUV	F(3,485) = 865.730, R^2 = 0.84	F(3,960) = 870.23, R^2 = 0.72	F(3,720) = 1146.83, R^2 = 0.82

	Subject 7	**Subject 8**	**Subject 9**
RGB	F(3,660) = 549.86, R^2 = 0.71	F(3,650) = 4816.16, R^2 = 0.95	F(3,510) = 853.02, R^2 = 0.83
HSV	F(3,660) = 606.60, R^2 = 0.73	F(3,650) = 13411.9, R^2 = 0.98	F(3,510) = 1094.56, R^2 = 0.87
YCbCr	F(3,660) = 624.50, R^2 = 0.77	F(3,650) = 10551.6, R^2 = 0.97	F(3,510) = 2245.70, R^2 = 0.89
Lab	F(3,660) = 450.50, R^2 = 0.75	F(3,650) = 11670.2, R^2 = 0.97	F(3,510) = 2261.82, R^2 = 0.90
YUV	F(3,660) = 533.50, R^2 = 0.71	F(3,650) = 7368.72, R^2 = 0.97	F(3,510) = 1577.04, R^2 = 0.90

5.5.4. Bland-Altman Plot

The agreement between HR measured from contact sensor technology and HR estimated from linear regression methods are calculated using difference plot as shown in Figure 40. From the result obtained in linear regression (Table 13), it is shown that multiple linear regression has consistent value of RMSE therefore only the HR estimated

from multivariate linear regression are considered to plot Bland-Altman plot. Figure 40 shows a Bland-Altman plot of a randomly selected subject. Table 16 shows the Bland-Altman analysis of predicted heart rate using multiple linear regression and the heart rate collected using a contact sensor (considered as true value). As an example, Figure 39 shows the Bland-Altman plot for a randomly selected one subject.

From the Table 16, the mean differences for GRAY, RGB, HSV, YCbCr, Lab and YUV are -0.49 ±0.33, -0.21 ±0.15, -0.20 ±0.17, -0.20 ±0.18, -0.24 ±0.18, -0.22 ±0.17 respectively. From the result it is found that HSV and YCbCr color models obtains the best values of mean differences, where the upper limit of agreements is 10.07 ± 4.71 and 11.01±4.63 and lower limit of agreements are -10.48 ± 5.05 and -11.93±4.98 for HSV and YCbCr respectively. From the overall result, it was found that heart rate can be predicted using facial color in real time. Among the five color spaces, HSV and YCbCr were the better color space for the heart rate predictor. As an individual color component, saturation (S) is the best predictor of heart rate among all the components of five color spaces. All the mean differences are negative which indicates that predicted values are underestimated comparing with the real values.

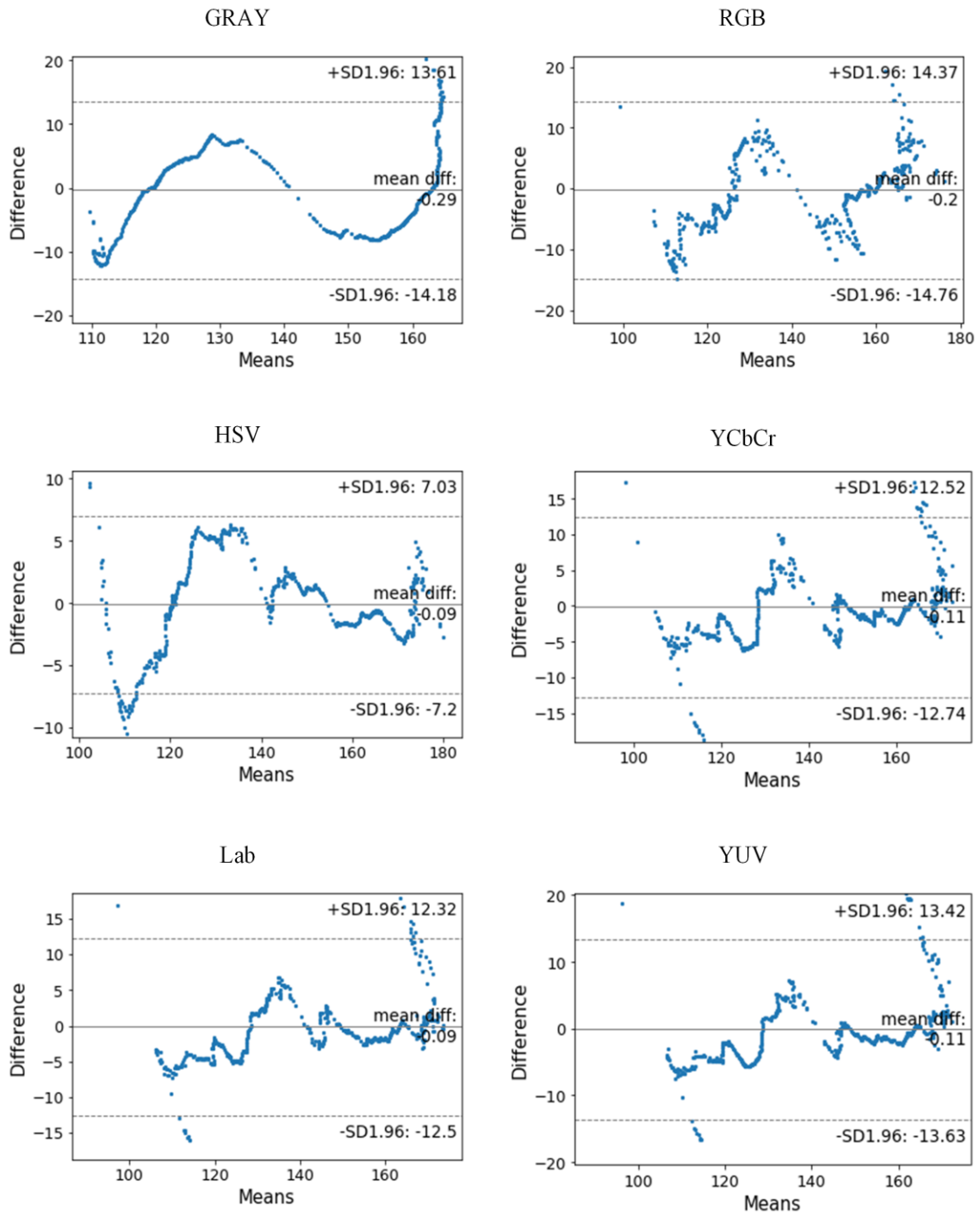


FIGURE 40 - BLAND-ALTMAN PLOT OF HEART RATE RECORDED USING CONTACT SENSOR AND PREDICTED HEART RATE USING MULTIVARIATE LINEAR REGRESSION METHODS.

TABLE 16 - THE BLAND-ALTMAN ANALYSIS OF HR MEASURED FROM CONTACT SENSOR TECHNOLOGY AND HR ESTIMATED USING MULTIVARIATE LINEAR REGRESSION WHERE THE INPUT PARAMETERS ARE INDIVIDUAL COMPONENT OF EACH COLOR MODEL.

		Sub1	**Sub2**	**Sub3**	**Sub4**	**Sub5**	**Sub6**	**Sub7**	**Sub8**	**Sub9**	**Avg**	**SD**
Gray	Diff	-0.57	-0.78	-0.25	-0.06	-0.22	-0.88	-0.68	-0.83	-0.11	-0.49	0.33
	-1.96SD	-18.08	-25.23	-10.85	-6.19	-11.85	-26.18	-23.98	-26.05	-14.61	-18.11	7.58
	1.96SD	16.94	23.67	10.36	6.07	11.40	24.42	22.62	24.38	14.38	17.14	6.96
RGB	Diff	-0.25	-0.16	-0.06	-0.35	-0.03	-0.48	-0.22	-0.29	-0.06	-0.21	0.15
	-1.96SD	-11.14	-10.72	-5.54	-16.66	-4.34	-18.73	-14.72	-15.03	-7.09	-11.55	5.11
	1.96SD	10.64	10.40	5.42	15.97	4.28	17.77	14.28	14.46	6.98	11.13	4.82
HSV	Diff	-0.12	-0.22	-0.07	-0.12	-0.04	-0.56	-0.28	-0.37	-0.05	-0.20	0.17
	-1.96SD	-7.35	-9.10	-6.52	-7.95	-5.25	-19.81	-14.70	-15.93	-7.68	-10.48	5.05
	1.96SD	7.10	8.67	6.37	7.71	5.17	18.70	14.13	15.19	7.58	10.07	4.71
YCbCr	Diff	-0.14	-0.22	-0.06	-0.10	-0.02	-0.48	-0.48	-0.27	-0.04	-0.20	0.18
	-1.96SD	-8.74	-10.57	-7.05	-12.73	-6.00	-18.22	-18.50	-14.80	-5.97	-11.40	4.94
	1.96SD	8.43	10.13	6.92	12.52	6.00	17.25	17.68	14.25	5.88	11.01	4.63
Lab	Diff	-0.40	-0.20	-0.05	-0.22	-0.02	-0.52	-0.42	-0.25	-0.06	-0.24	0.18
	-1.96SD	-15.97	-9.89	-6.57	-11.49	-5.62	-18.93	-18.21	-14.05	-6.66	-11.93	5.12
	1.96SD	15.17	9.49	6.46	11.05	5.57	17.89	17.36	13.56	6.54	11.46	4.77
YUV	Diff	-0.13	-0.23	-0.06	-0.22	-0.03	-0.50	-0.41	-0.35	-0.05	-0.22	0.17
	-1.96SD	-8.18	-10.36	-6.26	-11.22	-5.62	-18.70	-17.78	-15.65	-6.21	-11.11	5.12
	1.96SD	7.92	9.89	6.14	10.79	5.56	17.69	16.95	14.96	6.10	10.67	4.78

5.6. Discussion

This study aimed to quantify the relationship between heart rate measured using contact sensor and heart rate estimation through facial skin color modelling in healthy young people during incremental fatiguing test. The main findings suggested that the accuracy of estimated heart rate was higher when facial skin color was modeled by **multiple linear regression** using three independent variables using combined color space channels (RGB, HSV, YCbCr, Lab, and YUV) as compared to the model (single variable linear regression) that accounted for the individual color channel (R, G, B, H, S, V, etc.).

It has been a while since the change in the use of alternative color spaces in image processing, including in skin color and facial expression analysis (Chandrappa, Ravishankar, & Rameshbabu, 2011; Wang et al., 2015). Thus, it was expected that the

models would show better performance for face skin using HSV and YCbCr rather than RGB, as already reported previously (Sanchez-Cuevas, Aguilar-Ponce, & Tecpanecatl-Xihuitl, 2013).

It has been found that the best correlated color channels for heart rate estimation are R (-0.79 ±0.09) of RGB, Y (-0.79 ±0.08) of YCbCr, V (-0.79 ±0.08) of HSV, and L (-0.79 ±0.08) of Lab. The green channel of the addictive color space showed the highest accuracy and consistency of the correlation model for heart rate estimation. This is justified by the fact that green light component is better absorbed by hemoglobin than red and blue light component (Seidman et al., 2003). It is important to notice that the average value of correlation coefficients obtained from R and G component was almost similar, which may be due to the fact that skin color becomes redder when people reach fatigue. In contrast, the weakest predictors are H of HSV (-0.67±0.17), b of Lab (-0.68±0.22), and V of YUV (-0.68±0.20), which represents brightness. However, although not the most accurate, it is common to apply color channels (e.g. R, G, B, instead of combination of RGB) in facial color analysis, as well as remote physiological features measurements (Qerem, 2016; Wang et al., 2014; Wang et al., 2015).

Based on the results obtained from individual statistical analysis and generalized data analysis, the results for the generalized analysis seem to offer a poor solution to the problem. Therefore, these models, especially the multiple regression, should be fitted according to each individual feature. Previous studies have already reported a strong relationship between heart rate and color space components, but mostly for people in resting state, not continuously during exercise (Hassan et al., 2016). However, because measurement of intensity is essential to be assessed during physical performance, the determination of the behavior of the color components to predict heart rate in the timeframe enhances the importance of this study. Another aspect worth of notice is on the individuality observed between heart rate and color intensity values. We believe that the current study reaches an important next step by estimating physical effort using image recognition effectively during exercise, nevertheless, it seems that individual variability is still an issue that might be addressed to achieve higher accuracy in further studies.

5.7. Limitations

There are some limitations to this study: (1) Only the healthy subjects between the age of 18-40 years were recruited for participation; therefore, the findings of this study may not generalize for other populations such as the elderly. (2) The recruited participants were all white Caucasian; therefore, the findings may not generalize for other than Caucasian peoples. (3) This study was conducted in a laboratory setting in the stationary cycles; therefore, the findings for other types of exercise protocol may not fit into this result. (4) The experiments were performed for a participant once. (5) The video was recorded in fluorescent light; therefore, the results may not generalize for other light sources such as ambient and other types of light sources. (6) The participants were instructed to watch to the camera during the experiment which might effect on natural facial expression. (7) The number of participants was limited (only nine: six males and three females), and not equally gender-distributed, therefore, it is suggested for further investigation with a greater number of participants with diverse ethnicity and gender. (8) Only four alternative color spaces were implemented but many color models are already used for similar purposes in literature, other color space might result in more accurately. (9) The experiments and data analysis were carried out individually, therefore, the results cannot be generalized universally.

5.8. Conclusion and Future Work

The proposed methodology in this study consisted in describing the efficiency of color analysis algorithms to estimate heart rate, thus, internal load and exercise intensity. The experimental results showed that the facial color changes with the increase in exercise intensity. The main findings suggested that the accuracy of estimated heart rate was higher when facial skin color was modeled (multiple linear regression using three independent variables) using combined color space channels (RGB, HSV, YCbCr, Lab, and YUV), compared to the model (univariate linear regression) that accounted for the individual color channel (R, G, B, H, S, V, etc.). As an individual color component, the L color

channel of Lab color space almost reflects the heart responses and showed the highest average correlation (-0.79±0.08), lower error (RMSE of 9.47±2.20). As an alternative color spaces, HSV color space results best accuracy (average RMSE of 5.12 and mean difference in Bland-Altman plot -0.09) in comparison with others color spaces. The experiments were also carried out to fit a generalized regression models using data from all subjects. The results were statistically significant, but the error rate was very high, so, it might be concluded that, the proposed methods are more appropriate for individual fingerprinting.

This work can be extended to relate heart rate and the color intensity with more than five-color models, as well as to the investigation of different skin-colors. From the results, it was reported that this method can be appropriate for individual analysis, therefore, this work can be extended by analyzing the data from same subject in various days and various situations.

6. CONCLUSIONS AND FUTURE WORK

This chapter presents the overall conclusion drawn from the thesis. The future directions of research work are also described.

6.1. Thesis Contributions

In this thesis, it was developed two approaches for the quantification of exercise intensity using discrete and continuous predictions. The primary aim was to predict the physical exercise intensity based on facial expression and color using computer vision techniques. First, we have presented the mentioned approach to quantify physical exercise intensity based on incremental heart rate. The level of exercise intensity was classified into four classes and detected the particular level of exercise intensity using image processing and machine learning techniques. The level of intensity was detected from the face by the assessment using deep learning architecture. The results showed significant differences between fatigued and non-fatigued faces. The four levels of exercise intensities were correctly classified with accuracy values up to 99%, using convolutional neural network. It is a novelty, once studies published in the area present two classes successfully identified, thus, the present results might bring a great contribution in physical exercise monitoring (Almeida, Bottino, Ramos, & Araujo, 2019; Battaglini, Bottaro, Campbell, Novaes, & Simão, 2004; Xu et al., 2018). Second, the quantification of exercise intensity was proposed and applied using the continuous tracking of facial color changes during physical exercise. The changes in facial color are better identified using the HSV color model when compared to the original RGB model. The results also showed that the changes in facial expression and color during exercise are more significant than the direct heart rate measurement.

6.2. Limitations and Future Work

The primary limitation on the work developed throughout this thesis is related to the classification problem and the number of subjects available to participate in the experiment. The dataset included fewer subjects than planned, which demanded it to be split before performing the training and testing procedures; therefore, the results cannot be generalized. However, the methodology proposed in this thesis gives to further

research all the potential to be tested in a greater number of subjects. Secondly, from the evidences of four intensity levels identified with high accuracy and the identification of facial expressions and emotions, the advance to extend the classification into more classes and embracing facial features might give additional details to the study of physical exertion.

REFERENCES

Aboagye, E., Hagberg, J., Axén, I., Kwak, L., Lohela-Karlsson, M., Skillgate, E., . . . Jensen, I. (2017). Individual preferences for physical exercise as secondary prevention for non-specific low back pain: A discrete choice experiment. *PloS one, 12*(12), e0187709-e0187709. doi:10.1371/journal.pone.0187709

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-Based Learning Algorithms. *Mach. Learn., 6*(1), 37-66. doi:10.1023/a:1022689900470

Ahmadi Ak Fau - Moradi, P., Moradi P Fau - Malihi, M., Malihi M Fau - Karimi, S., Karimi S Fau - Shamsollahi, M. B., & Shamsollahi, M. B. (2015). Heart Rate monitoring during physical exercise using wrist-type photoplethysmographic (PPG) signals. (1557-170X (Print)).

Aliverti, A. (2016). The respiratory muscles during exercise. *Breathe (Sheffield, England), 12*(2), 165-168. doi:10.1183/20734735.008116

Allen, J. K., Stephens, J., Dennison Himmelfarb, C. R., Stewart, K. J., & Hauck, S. (2013). Randomized controlled pilot study testing use of smartphone technology for obesity treatment. *Journal of obesity, 2013*.

Almeida, M., Bottino, A., Ramos, P., & Araujo, C. G. (2019). Measuring Heart Rate During Exercise: From Artery Palpation to Monitors and Apps. *International Journal of Cardiovascular Sciences, 32*, 396-407.

Alugupally, N., Samal, A., Marx, D., & Bhatia, S. (2011). Analysis of landmarks in recognition of face expressions. *Pattern Recognition and Image Analysis, 21*(4), 681-693. doi:10.1134/S105466181104002X

Aoki, H., Ichimura, S., Kiyooka, S., & Koshiji, K. (2007). *Non-contact measurement method of respiratory movement under pedal stroke motion.* Paper presented at the 29th Annual International Conference of IEEE-EMBS, Engineering in Medicine and Biology Society, EMBC'07, Lyon.

Aoki, H., & Nakamura, H. (2018). Non-Contact Respiration Measurement during Exercise Tolerance Test by Using Kinect Sensor. *Sports, 6*(1). doi:ARTN 2310.3390/sports6010023

Aoki, H., Nakamura, H., Fumoto, K., Nakahara, K., & Teraoka, M. (2015). Basic Study on Non-contact Respiration Measurement during Exercise Tolerance Test by Using Kinect Sensor. *2015 Ieee/Sice International Symposium on System Integration (Sii)*, 217-222.

Balakrishnan, G., Durand, F., & Guttag, J. (2013). *Detecting Pulse from Head Motions in Video.* Paper presented at the 2013 IEEE Conference on Computer Vision and Pattern Recognition.

Battaglini, C. L., Bottaro, M., Campbell, J. S., Novaes, J., & Simão, R. (2004). Atividade física e níveis de fadiga em pacientes portadores de câncer. *Revista Brasileira de Medicina do Esporte, 10*, 98-104.

Bayles, M. P., & Swank, A. M. (2018). *ACSM's exercise testing and prescription*: American College of Sports, Medicine.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*: Oxford University Press, Inc.

Bishop, C. M. (2006). *Pattern recognition and machine learning*: springer.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurements. *The Lancet, 327*(8476), 307-310.

Borg, G. (1962). *Physical Performance and Perceived Exertion*. Copenhagan: Lund.

Borg, G. (1998). *Borg's Perceived exertion and pain scales*. Rimbi, Sweden: Human Kinetics.

Bosi, I., Cogerino, C., & Bazzani, M. (2016). *Real-time monitoring of heart rate by processing of Microsoft Kinect™ 2.0 generated streams.* Paper presented at the 1st International Multidisciplinary Conference on Computer and Energy Science, SpliTech 2016.

Breiman, L. (1996). Bagging Predictors. *Machine Learning, 24*(123). doi:https://doi.org/10.1023/A:1018054314350

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. doi:10.1023/A:1010933404324

Capraro, G., Etebari, C., Luchette, K., Mercurio, L., Merck, D., Kirenko, I., . . . Kobayashi, L. (2018, 17-19 Oct. 2018). *'No Touch' Vitals: A Pilot Study of Non-contact Vital Signs Acquisition in Exercising Volunteers.* Paper presented at the 2018 IEEE Biomedical Circuits and Systems Conference (BioCAS).

Caudill, M. (1989). Neural Network Primer, reprinted from issues of. *AI Expert*.

Cerezuela-Espejo, V., Courel-Ibáñez, J., Morán-Navarro, R., Martínez-Cava, A., & Pallarés, J. G. (2018). The Relationship Between Lactate and Ventilatory Thresholds in Runners: Validity and Reliability of Exercise Test Performance Parameters. *Frontiers in Physiology, 9*(1320). doi:10.3389/fphys.2018.01320

Chan, C., Inskip, J., Kirkham, A., Ansermino, J., Dumont, G., Li, L., . . . Hoens, A. (2018). A smartphone oximeter with a fingertip probe for use during exercise training: usability, validity and reliability in individuals with chronic lung disease and healthy controls. *Physiotherapy*.

Chandrappa, D. N., Ravishankar, M., & Rameshbabu, D. R. (2011). *Face detection in color images using skin color model algorithm based on skin color information.* Paper presented at the International Conference on Electronics Computer Technology, Kanyakumari, India.

Chaves-González, J. M., Vega-Rodríguez, M. A., Gómez-Pulido, J. A., & Sánchez-Pérez, J. M. (2010). Detecting skin in face recognition systems: A colour spaces study. *Digital Signal Processing, 20*(3), 806-823. doi:https://doi.org/10.1016/j.dsp.2009.10.008

Chen, D., Liu, S., Kingsbury, P., Sohn, S., Storlie, C. B., Habermann, E. B., . . . Liu, H. (2019). Deep learning and alternative learning strategies for retrospective real-world clinical data. *npj Digital Medicine, 2*(1), 43. doi:10.1038/s41746-019-0122-0

Chen, Y.-L., Chiou, W.-K., Tzeng, Y.-T., Lu, C.-Y., & Chen, S.-C. (2017). A rating of perceived exertion scale using facial expressions for conveying exercise intensity for children and young adults. *Journal of Science and Medicine in Sport, 20*(1), 66-69. doi:https://doi.org/10.1016/j.jsams.2016.05.009

Cheng, W.-C., Wu, T.-Y., & Li, D.-W. (2018). *Ensemble Convolutional Neural Networks for Face Recognition*. Paper presented at the Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence, Sanya, China.

Comon, P. (1994). Independent component analysis, A new concept? *Signal Processing, 36*(3), 287-314. doi:https://doi.org/10.1016/0165-1684(94)90029-9

Cui, Y., Fu, C. H., Hong, H., Zhang, Y., & Shu, F. (2015). *Non-contact time varying heart rate monitoring in exercise by video camera.* Paper presented at the International Conference on Wireless Communications and Signal Processing, WCSP 2015.

Das P Fau - Horton, R., & Horton, R. (2012). Rethinking our approach to physical activity. (1474-547X (Electronic)).

Dias, D., & Cunha, J. (2018). Wearable Health Devices—Vital Sign Monitoring, Systems and Technologies. *Sensors*, 1-28.

Docampo, G. N. (2012). *Heart rate estimation using facial video information*. Retrieved from Pontevedra:

Dooley, E. E., Golaszewski, N. M., & Bartholomew, J. B. (2017). Estimating Accuracy at Exercise Intensities: A Comparative Study of Self-Monitoring Heart Rate and Physical Activity Wearable Devices. *JMIR mHealth and uHealth, 5*(3).

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern Classification (2nd Edition)*: Wiley-Interscience.

Duking, P., Achtzehn, S., Holmberg, H.-C., & Sperlich, B. (2018). Integrated Framework of Load Monitoring by a Combination of Smartphone Applications, Wearables and Point-of-Care Testing Provides Feedback that Allows Individual Responsive Adjustments to Activities of Daily Living. *sensor, 18*(5), 1-11.

Eka Adi Prasetyo Joko, P., Chia-Chun, H., Yi-Sheng, C., Cheng-Hung, C., & Yuan-Hsiang, L. (2014, 2014). *A heart rate detection method for low power exercise intensity monitoring device.* Paper presented at the 2014 IEEE International Symposium on Bioelectronics and Bioinformatics (IEEE ISBB 2014).

Ekman, P., Wallace, V., & Ancoli, F. S. (1980). Facail Signs of emotion experiences. *Journal of Personality and socail psychology*, 1125-1134.

Ellwein, L., Samyn, M. M., Danduran, M., Schindler-Ivens, S., Liebham, S., & LaDisa, J. F. (2017). Toward translating near-infrared spectroscopy oxygen saturation data for the non-invasive prediction of spatial and temporal hemodynamics during exercise. *Biomechanics and Modeling in Mechanobiology, 16*(1), 75-96. doi:10.1007/s10237-016-0803-4

Fallet, S., & Vesin, J.-M. (2017). Robust heart rate estimation using wrist-type photoplethysmographic signals during physical exercise: an approach based on adaptive filtering. *Physiological Measurement, 38*(2), 155-170. doi:10.1088/1361-6579/aa506e

Fatemeh Shirbani, C. B., Christina Kazzi, Isabella Tan, Mark Butlin. (2018). *Sensitivity of Video-Based Pulse Arrival Time to Dynamic Blood Pressure Changes.* Paper presented at the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA.

Faulkner, J., & Eston, R. G. (2008). Percieved exertion research in the 21st century: Developments, reflections and questions for the future. *Journal of Exercise Science & Fitness, 6*(1).

Fernandes, A. A., Amorim, P. R. S., Brito, C. J., Sillero-Quintana, M., & Marins, J. C. B. (2016). Regional skin temperature response to moderate aerobic exercise measured by infrared thermography. *Asian Journal of Sports Medicine, 7*(1). doi:10.5812/asjsm.29243

Flury, B. (1988). *Common principal components \& related multivariate models*: John Wiley \& Sons, Inc.

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences, 55*(1), 119-139. doi:https://doi.org/10.1006/jcss.1997.1504

Gade, R., Larsen, R. G., & Moeslund, T. B. (2017). Measuring energy expenditure in sports by thermal video analysis. *2017 Ieee Conference on Computer Vision and Pattern Recognition Workshops (Cvprw)*, 187-194. doi:10.1109/Cvprw.2017.29

Gambi, E., Agostinelli, A., Belli, A., Burattini, L., Cippitelli, E., Fioretti, S., . . . Spinsante, S. (2017). Heart Rate Detection Using Microsoft Kinect: Validation and Comparison to Wearable Devices. *Sensors (Basel, Switzerland), 17*(8), 1776. doi:10.3390/s17081776

García-García, F., Benito, P. J., & Hernando, M. E. (2016). Automatic Identification of Physical Activity Intensity and Modality from the Fusion of Accelerometry and Heart Rate Data. *Methods Inf Med, 55*(06), 533-544. doi:10.3414/ME15-01-0130

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning, 63*(1), 3-42. doi:10.1007/s10994-006-6226-1

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochem Med (Zabreb), 25*(2), 141-151.

Glynn, L. G., Hayes, P. S., Casey, M., Glynn, F., Alvarez-Iglesias, A., Newell, J., . . . Murphy, A. W. (2014). Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. *Br J Gen Pract, 64*(624), e384-e391.

Gordienko, Y., Stirenko, S., Kochura, Y., Alienin, O., Novotarskiy, M., & Gordienko, N. (2017). Deep Learning for Fatigue Estimation on the Basis of Multimodal Human-Machine Interactions. *CoRR, abs/1801.06048*, 1-12.

Gredell, D. A., Schroeder, A. R., Belk, K. E., Broeckling, C. D., Heuberger, A. L., Kim, S.-Y., . . . Prenni, J. E. (2019). Comparison of Machine Learning Algorithms for Predictive Modeling of Beef Attributes Using Rapid Evaporative Ionization Mass Spectrometry (REIMS) Data. *Scientific Reports, 9*(1), 5721. doi:10.1038/s41598-019-40927-6

Hannan, A. L., Harders, M. P., Hing, W., Climstein, M., Coombes, J. S., & Furness, J. (2019). Impact of wearable physical activity monitoring devices with exercise prescription or advice in the maintenance phase of cardiac rehabilitation: systematic review and meta-analysis. *BMC sports science, medicine & rehabilitation, 11*, 14-14. doi:10.1186/s13102-019-0126-8

Haque, M. A., Irani, R., Nasrollahi, K., & Moeslund, T. B. (2016). Facial video based detection of physical fatigue for maximal muscle activity. *I E T Computer Vision*.

Hasan, M. M., Hossain, M. F., Thakur, J. M., & Podder. (2014). Driver fatigue recognition using skin color modeling. *Internataional Journal of Computer Application, 97*(16), 34-41.

Hassan, M. A., Malik, G. S., Saad, N., Karasfi, B., Ali, Y. S., & Fofi, D. (2016). *Optimal source selection for image photoplethysmography.* Paper presented at the IEEE International Instrumentation and Measurement Technology Conference Proceedings, Taipei, Taiwan.

Hasselmann, V., Oesch, P., Fernandez-Luque, L., & Bachmann, S. (2015). Are exergames promoting mobility an attractive alternative to conventional self-regulated exercises for elderly people in a rehabilitation setting? Study protocol of a randomized controlled trial. *BMC geriatrics, 15*(1), 108.

Henriksen, A., Haugen Mikalsen, M., Woldaregay, A. Z., Muzny, M., Hartvigsen, G., Hopstock, L. A., & Grimsgaard, S. (2018). Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables. *Journal of medical Internet research, 20*(3), e110-e110. doi:10.2196/jmir.9157

Hensen, S. J. (2017). Measuring physical activity with heart rate monitors. *American Journal of Public Health, 107*(12), e24-e24.

Hernndez-Blanco, A., Herrera-Flores, B., Toms, D., & Navarro-Colorado, B. (2019). A Systematic Review of Deep Learning Approaches to Educational Data Mining. *Complexity, 2019*, 22. doi:10.1155/2019/1306039

Huanga, D. H., Chioua, W. K., & Chenb, B. H. (2015). *Judgment of perceived exertion by static and dynamic facial expression.* Paper presented at the Triennial Congress of the IEA, Melbourne.

Hulme, K., Safari, R., Thomas, S., Mercer, T., White, C., Van der Linden, M., & Moss-Morris, R. (2018). Fatigue interventions in long term, physical health conditions: A scoping review of systematic reviews. *PloS one, 13*(10), e0203367. doi:10.1371/journal.pone.0203367

Hung, K., & Zhang, Y. (2006). *Preliminary investigation of pupil size variability: toward non-contact assessment of cardiovascular variability.* Paper presented

at the 2006 3rd IEEE/EMBS International Summer School on Medical Devices and Biosensors.

Hunt, K. J., & Fankhauser, S. E. (2016). Heart rate control during treadmill exercise using input-sensitivity shaping for disturbance rejection of very-low-frequency heart rate variability. *Biomedical Signal Processing and Control, 30*, 31-42. doi:https://doi.org/10.1016/j.bspc.2016.06.005

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks, 13*(4), 411-430. doi:https://doi.org/10.1016/S0893-6080(00)00026-5

Inder, J. D., Carlson, D. J., Dieberg, G., McFarlane, J. R., Hess, N. C., & Smart, N. A. (2016). Isometric exercise training for blood pressure management: a systematic review and meta-analysis to optimize benefit. *Hypertension Research, 39*(2), 88.

Irani, R., Nasrollahi, K., & Moeslund, T. B. (2014). Contactless measurement of muscle fatigue by tracking facial feature points in video. 4181-5186.

Jackson, J. E. (1991). *A User's Guide to Principal Components*: John Wiley & Sons, Inc. .

James, C. A., Richardson, A. J., Watt, P. W., & Maxwell, N. S. (2014). Reliability and validity of skin temperature measurement by telemetry thermistors and a thermal camera during exercise in the heat. *Journal of Thermal Biology, 45*, 141-149. doi:10.1016/j.jtherbio.2014.08.010

Jensen, M. M., Poulsen, M. K., Alldieck, T., Larsen, R. G., Gade, R., Moeslund, T. B., & Franch, J. (2016). Estimation of Energy Expenditure during Treadmill Exercise via Thermal Imaging. *Med Sci Sports Exerc, 48*(12), 2571-2579.

Jeong, I. C., & Finkelstein, J. (2016). Introducing Contactless Blood Pressure Assessment Using Speed Video Camera. *J Med Syst*, 1-10.

Jeong, M., & Ko, B. C. (2018). Driver's Facial Expression Recognition in Real-Time for Safe Driving. *Sensors (Basel, Switzerland), 18*(12), 4270. doi:10.3390/s18124270

Jimenez, J., Scully, T., Barbosa, N., Donner, C., Alvarez, X., Vierira, T., . . . Weyrich, T. (2010). A practical appearance model for dynamic facial color. *ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH Asia, 29*(6), 1-8.

Karlsen, K., Larsen, J. P., Tandberg, E., & Jørgensen, K. (1999). Fatigue in patients with Parkinson's disease. *Movement Disorders, 14*(2), 237-241.

Khan, S., & Yairi, T. (2018). A review on the application of deep learning in system health management. *Mechanical Systems and Signal Processing, 107*, 241-265. doi:https://doi.org/10.1016/j.ymssp.2017.11.024

Khanal, S., Reis, A., Barroso, J., & Filipe, V. (2018). *Using Emotion Recognition in Intelligent Interface Design for Elderly Care.* Paper presented at the Trends and Advances in Information Systems and Technologies, Cham.

Khanal, S. R., Barroso, J., Sampaio, J., & Filipe, V. (2018). *Classification of physical exercise intensity by using facial expression analysis.* Paper presented at the 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), Erode, India.

Khanal, S. R., Fonseca, A., Marques, A., Barroso, J., & Filipe, V. (2018). *Physical exercise intensity monitoring through eye-blink and mouth's shape analysis.* Paper presented at the 2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW).

Khanal, S. R., Sampaio, J., Barroso, J., & Filipe, V. (2019). *Classification of Physical Exercise Intensity Based on Facial Expression Using Deep Neural Network.* Paper presented at the Universal Access in Human-Computer Interaction. Multimodality and Assistive Environments, Cham.

Kim, J., Son, J., Ko, N., & Yoon, B. (2013). Unsupervised virtual reality-based exercise program improves hip muscle strength and balance control in older adults: a pilot study. *Archives of physical medicine and rehabilitation, 94*(5), 937-943.

King, D. E. (2009). Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research, 10*, 1755-1758.

Kiviniemi, A. M., Hautala, A. J., Kinnunen, H., Nissila, J., Virtanen, P., Karjalainen, J., & Tulppo, M. P. (2015). Daily Exercise Prescription on the Basis of HR Variability among Men and Women. *Medicine & Science in Sports & Exercise, 42*(7), 1355-1363.

Ko, B. C. (2018). A Brief Review of Facial Emotion Recognition Based on Visual Information. *Sensors (Basel, Switzerland), 18*(2), 401. doi:10.3390/s18020401

Koehler, K., & Drenowatz, C. (2017). Monitoring Energy Expenditure Using a Multi-Sensor Device—Applications and Limitations of the SenseWear Armband in Athletic Populations. *Frontiers in Physiology, 8*(983). doi:10.3389/fphys.2017.00983

Koporec, G., Vučković, G., Milić, R., & Perš, J. (2018). Quantitative Contact-Less Estimation of Energy Expenditure from Video and 3D Imagery. *Sensors (Basel, Switzerland), 18*(8). doi:10.3390/s18082435

Korman, P., Straburzynska-Lupa, A., Kusy, K., Kantanista, A., & Zielinski, J. (2016). Changes in body surface temperature during speed endurance work-out in highly-trained male sprinters. *Infrared Physics & Technology, 78*, 209-213. doi:10.1016/j.infrared.2016.08.003

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet classification with deep convolutional neural networks*. Paper presented at the Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Lake Tahoe, Nevada.

Krupp, L. B. Fatigue in multiple sclerosis: definition, pathophysiology and treatment. (1172-7047 (Print)).

Kumar, M., Veeraraghavan, A., & Sabharwal, A. (2015). DistancePPG: Robust non-contact vital signs monitoring using a camera. *Biomedical optics express, 6*(5), 1565-1588. doi:10.1364/BOE.6.001565

Lajevardi, S. M., & Wu, H. R. (2012). Facial Expression Recognition in Perceptual Color Space. *IEEE Transactions on Image Processing, 21*(8), 3721-3733. doi:10.1109/TIP.2012.2197628

Leat, Mei, M., & Susan, J. (2008). Quantitative Assessment of Perceived Visibility Enhancement with Image Processing for Single Face Images: A Preliminary Study. *Investigative Ophthalmology & Visual Science*, 4502-4508.

Li, K. H. C., White, F. A., Tipoe, T., Liu, T., Wong, M., Jesuthasan, A., . . . Yan, B. P. (2019). The Current State of Mobile Phone Apps for Monitoring Heart Rate, Heart Rate Variability, and Atrial Fibrillation: Narrative Review. *JMIR mHealth and uHealth, 7*(2), e11606-e11606. doi:10.2196/11606

Li, S., Li, X., Lv, Q., & Zhang, D. (2018). *WiFit: A Bodyweight Exercise Monitoring System with Commodity Wi-Fi*. Paper presented at the Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, Singapore, Singapore.

Liao, Y., Vakanski, A., & Xian, M. (2019). A Deep Learning Framework for Assessing Physical Rehabilitation Exercises. *CoRR, abs/1901.10435*.

Liliana, D. Y. (2019). Emotion recognition from facial expression using deep convolutional neural network. *Journal of Physics: Conference Series, 1193*, 012004. doi:10.1088/1742-6596/1193/1/012004

Lin, Y.-C., & Lin, Y.-H. (2018). Step Count and Pulse Rate Detection Based on the Contactless Image Measurement Method. *IEEE Transactions on Multimedia, 20*(8), 2223-2231.

Lin, Y. (2018). Step Count and Pulse Rate Detection Based on the Contactless Image Measurement Method. *IEEE Transactions on Multimedia, 20*(8), 2223-2231. doi:10.1109/TMM.2018.2790172

Lopes, A. T., de Aguiar, E., De Souza, A. F., & Oliveira-Santos, T. (2017). Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order. *Pattern Recognition, 61*, 610-628. doi:https://doi.org/10.1016/j.patcog.2016.07.026

Lopez, M. B., del-Blanco, C. R., & Garcia, N. (2017). *Detecting exercise-induced fatigue using thermal imaging and deep learning.* Paper presented at the 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), Montreal, QC, Canada.

Ludwig, N., Formenti, D., Gargano, M., & Alberti, G. (2014). Skin temperature evaluation by infrared thermography: Comparison of image analysis methods. *Infrared Physics & Technology, 62*, 1-6. doi:10.1016/j.infrared.2013.09.011

Mackingnon, S. N. (1999). Relating heart rate and rate of perceived exertion in two simulated occupational tasks. *Ergonomics, 42*(5), 761-766.

Maclin, R., & Opitz, D. W. (2011). Popular Ensemble Methods: An Empirical Study. *CoRR, abs/1106.0257*.

Marco C. Uchida, R. C., Vitor Daniel Tessutti, Reury Frank, Pereira Bacurau, He´ lio Jose´ Coelho-Ju´ nior, Luciane Portas Capelo,Heloiza Prando Ramos, Marcia Calixto dos Santos, Luı´s Felipe Milano Teixeira, Paulo Henruque marchetti. (2018). Identification of muscle fatigue by tracking. *PloS one, 13*(12), e0208834.

McFarlin, B., Venable, A., Williams, R., & Jackson, A. (2015). Comparison of techniques for the measurement of skin temperature during exercise in a hot, humid environment. *Biology of Sport, 32*(1), 11-14. doi:10.5604/20831862.1124569

Meyer, T., Gabriel Hh Fau - Kindermann, W., & Kindermann, W. (1999). Is determination of exercise intensities as percentages of VO2max or HRmax adequate? (0195-9131 (Print)).

Miles, K. H., Clark, B., Periard, J. D., Goecke, R., & Thompson, K. G. Facial feature tracking: a psychophysiological measure to assess exercise intensity? (1466-447X (Electronic)).

Miles, K. H., Clark, B., Périard, J. D., Goecke, R., & Thompson, K. G. (2017). Facial feature tracking: a psychophysiological measure to assess exercise intensity? *Journal of Sports Sciences*, 1-9.

Mitra, S. K., Li, H., Lin, I., & Yu, T. (1991). *A new class of nonlinear filters for image enhancement.* Paper presented at the [Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing.

Monkaresi, H., Calvo, R. A., & Yan, H. (2014). A Machine Learning Approach to Improve Contactless Heart Rate Monitoring Using a Webcam. *IEEE Journal of Biomedical and Health Informatics, 18*(4), 1153-1160. doi:10.1109/JBHI.2013.2291900

Montoye, A. H. K., Westgate, B. S., Fonley, M. R., & Pfeiffer, K. A. (2018). Cross-validation and out-of-sample testing of physical activity intensity predictions with a wrist-worn accelerometer. *Journal of Applied Physiology, 124*(5), 1284-1293.

More, A. V., Wakankar, A., & Gawande, J. P. (2019). Automated heart rate measurement using wavelet analysis of face video sequences. In H. S. Saini, R. K. Sing, V. M. Patel, K. Santhi, & S. V. Ranganayakulu (Eds.), *Innovations in*

Electronics and Communication Engineering (pp. 113-120). Singapore: Springer, Singapore.

Nakamura, F. Y., Flatt, A. A., Pereira, L. A., Campillo, R. R., Loturco, I., & Esco, M. R. (2015). Ultra-Short-Term Heart Rate Variability is Sensitive to Training Effects in Team Sports Players. *Journal of Sports Science and Medicine, 14*(3), 602-605.

Nelson, M. W., Rejeski, W. J., Blair, S. N., Duncan, P. W., Judge, J. O., King, C. A., & -Sceppa, M. C. (2007). Physical Activity and Public Health in Older Adults: Recommendation from the American College of Sports Medicine and the American Heart Association. *Medicine & Science in Sports & Exercise, 39*(8), 1435-1445.

Nikolaidis, P. T., Rosemann, T., & Knechtle, B. (2018). Age-Predicted Maximal Heart Rate in Recreational Marathon Runners: A Cross-Sectional Study on Fox's and Tanaka's Equations. (1664-042X (Print)).

Orieux, F., Giovannelli, J., & Rodet, T. (2010). Bayesian estimation of regularization and point spread function parameters for Wiener–Hunt deconvolution. *J. Opt. Soc. Am. A, 27*(7), 1593-1607.

Pal, M., Roy, R., Basu, J., & Bepari, M. S. (2013). *Blind source separation: A review and analysis.* Paper presented at the 2013 International Conference Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE).

Palatini, P. (1988). Blood Pressure Behaviour During Physical Activity. *Sports Medicine, 5*(6), 353-374. doi:10.2165/00007256-198805060-00002

Paschos, G. (2001). Perceptually Uniform Color Spaces for Color Texture Analysis: An Emperical Evaluation. *IEEE Transactions on Image Processing, 10*(6), 932-937.

Paterson, D. H., & Warburton, D. E. (2010). *Physical activity and functional limitations in older adults: a systematic review related to Canada's Physical Activity Guidelines.* Paper presented at the International Journal of Behavioral Nutrition and Physical Activity.

Pernek, I., Kurillo, G., Stiglic, G., & Bajcsy, R. (2015). Recognizing the intensity of strength training exercises with wearable sensors. *Journal of Biomedical Informatics, 58*, 145-155. doi:https://doi.org/10.1016/j.jbi.2015.09.020

Persinger, R., Foster, C., Gibson, M., Fater, D. C. W., & Porcari, J. P. (2004). Consistency of the Talk Test for exercise prescription. *Medicine and Science in Sports and Exercise, 36*(9), 1632-1636. doi:10.1249/01.MSS.0000074670.03001.98

Pichierri, G., Coppe, A., Lorenzetti, S., Murer, K., & de Bruin, E. D. (2012). The effect of a cognitive-motor intervention on voluntary step execution under single and dual task conditions in older adults: a randomized controlled pilot study. *Clinical interventions in aging, 7*, 175.

Poh, M.-Z., McDuff, D. J., & Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express, 18*(10), 10762-10774.

Poh, M. Z., McDuff, D. J., & Picard, R. W. (2011). *Advancements in noncontact, multiparameter physiological measurements using webcam.* Paper presented at the IEEE Transactions on Biomedical Engineering

Pollock, M. L., & Wilmore, J. H. (1990). *Exercise in Health and Disease: Evaluation and Prescription for Prevention and Rehabilitation* (2nd Edition ed.): Philadelphia.

Qerem, A. (2016). Face detection and recognition using fusion of color space. *International journal of Computer Science and Electronics Engineering, 4*(1), 12-16.

Ramirez, G. A., Fuentes, O., Crites, S. L., Jimenez, M., & Ordonez, J. (2014). *Color Analysis of Facial Skin: Detection of Emotional State.* Paper presented at the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA.

Reed, J. L., & Pipe, A. L. (2016). Practical Approaches to Prescribing Physical Activity and Monitoring Exercise Intensity. *Canadian Journal of Cardiology, 32*(4), 514-522. doi:https://doi.org/10.1016/j.cjca.2015.12.024

Reynolds, P. (2014). Exercise considerations for aging adults. In T. L. Kauffman, R. Scott, J. O. Barr, & M. L. Moran (Eds.), *A Comprehensive Guide to Geriatric Rehabilitation (Third Edition)* (pp. 267-274). Oxford: Churchill Livingstone.

Riou, M.-È., Rioux, F., Lamothe, G., & Doucet, É. (2015). Validation and Reliability of a Classification Method to Measure the Time Spent Performing Different Activities. *PloS one, 10*(6), e0128299. doi:10.1371/journal.pone.0128299

Robertson, R. J., & Noble, B. J. (1997). Perception of physical exertion: methods, mediators, and applications. (0091-6331 (Print)).

Rokach, L. (2005). Ensemble Methods for Classifiers. In O. Maimon & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 957-980). Boston, MA: Springer US.

Samadiani, N., Huang, G., Cai, B., Luo, W., Chi, C.-H., Xiang, Y., & He, J. (2019). A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data. *Sensors (Basel, Switzerland), 19*(8), 1863. doi:10.3390/s19081863

Sanchez-Cuevas, M. C., Aguilar-Ponce, R. M., & Tecpanecatl-Xihuitl, J. L. (2013). A Comparison of Color Models for Color Face Segmentation. *Procedia Technology, 7*, 134-141.

Sarzynski, M. A., Rankinen, T., Earnest, C. P., Leon, A. S., Rao, D. C., Skinner, J. S., & Bouchard, C. (2013). Measured maximal heart rates compared to commonly

used age-based prediction equations in the Heritage Family Study. *American journal of human biology : the official journal of the Human Biology Council, 25*(5), 695-701. doi:10.1002/ajhb.22431

Sawka, M. N., Wenger, C. B., Young, A. J., & Pandolf, K. B. (1993). Physiological Responses to Exercise in the Heat. In M. BM (Ed.), *Nutritional Needs in Hot Environments: Applications for Military Personnel in Field Operations* (Vol. 3). Washington (DC),National Academies Press (US), Institute of Medicine (US) Committee on Military Nutrition Research.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85-117. doi:https://doi.org/10.1016/j.neunet.2014.09.003

Schoene, D., Lord, S. R., Delbaere, K., Severino, C., Davies, T. A., & Smith, S. T. (2013). A randomized controlled pilot study of home-based step training in older people using videogame technology. *PloS one, 8*(3), e57734.

Schuch, F. B., Vancampfort, D., Richards, J., Rosenbaum, S., Ward, P. B., & Stubbs, B. (2016). Exercise as a treatment for depression: A meta-analysis adjusting for publication bias. (1879-1379 (Electronic)).

Scribbans, T. D., Vecsey, S., Hankinson, P. B., Foster, W. S., & Gurd, B. J. (2016). The Effect of Training Intensity on VO(2)max in Young Healthy Adults: A Meta-Regression and Meta-Analysis. *International journal of exercise science, 9*(2), 230-247.

Seidman, D. S., Moise, J., Ergaz, Z., Laor, A., Vreman, H. J., & al., D. K. S. e. (2003). A prospective randomized controlled study of phototherapy using blue, and blue-green light-emiting devices, and conventional halogen-quartz phototherapy. *Journal of perinatology*, 123-127.

Seshadri, D. R., Li, R. T., Voos, J. E., Rowbottom, J. R., Alfes, C. M., Zorman, C. A., & Drummond, C. K. (2019). Wearable sensors for monitoring the internal and external workload of the athlete. *npj Digital Medicine, 2*(1), 71. doi:10.1038/s41746-019-0149-2

Sharpe, M., & Wilks, D. (2002). Fatigue. *BMJ (Clinical research ed.), 325*(7362), 480-483. doi:10.1136/bmj.325.7362.480

Sheaffer, B. L., Golden, J. A., & Averett, P. (2009). Facial expression recognition deficits and faulty learning: Implications for theoretical models and clinical applications [Press release]

Shih, P., & Liu, C. (2005, 2005//). *Comparative Assessment of Content-Based Face Image Retrieval in Different Color Spaces.* Paper presented at the Audio- and Video-Based Biometric Person Authentication, Berlin, Heidelberg.

Shin, D. W., Yun, J. M., Shin, J. H., Kwon, H., Min, H. Y., Joh, H. K., . . . Cho, B. (2017). Enhancing physical activity and reducing obesity through smartcare and financial incentives: a pilot randomized trial. *Obesity, 25*(2), 302-310.

Silveira, P., van het Reve, E., Daniel, F., Casati, F., & De Bruin, E. D. (2013). Motivating and assisting physical exercise in independently living older adults: a pilot study. *International journal of medical informatics, 82*(5), 325-334.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Soar, J., Bargshady, G., Zhou, X., & Whittaker, F. (2018). *Deep Learning Model for Detection of Pain Intensity from Facial Expression.* Paper presented at the Smart Homes and Health Telematics, Designing a Better Future: Urban Assisted Living, Cham.

Spinsante, S., Ricciuti, M., & Scalise, L. (2018). *Contactless Measurement of Heart Rate for Exergames Applications.* Paper presented at the 13th IEEE International Symposium on Medical Measurements and Applications, MeMeA 2018.

Sun, G., Matsui, T., Watai, Y., Kim, S., Kirimoto, T., Suzuki, S., & Hakozaki, Y. (2018). Vital-SCOPE: Design and Evaluation of a Smart Vital Sign Monitor

for Simultaneous Measurement of Pulse Rate, Respiratory Rate, and Body Temperature for Patient Monitoring. *Journal of Sensors*, 1-7.

Sun, Y., Papin, C., Azorin-Peris, V., Kalawsky, R., Greenwald, S., & Hu, S. (2012). Use of ambient light in remote photoplethysmographic systems: Comparison between a high-performance camera and a low-cost webcam. *Journal of Biomedical Optics, 17*(3). doi:10.1117/1.JBO.17.3.037005

Sun, Y., Papin, C., Azorin-Peris, V., Kalawsky, R., Greenwald, S., & Hu, S. J. (2011). Comparison of scientific CMOS camera and webcam for monitoring cardiac pulse after exercise. *Applications of Digital Image Processing Xxxiv, 8135*. doi:Artn 813506

10.1117/12.893362

Tao, L., Burghardt, T., Mirmehdi, M., Damen, D., Cooper, A., Hannuna, S., . . . Craddock, I. (2016). *Calorie Counter: RGB-Depth Visual Estimation of Energy Expenditure at Home.* Paper presented at the Chen CS., Lu J., Ma KK. (eds) Computer Vision – ACCV 2016 Workshops. ACCV 2016. Lecture Notes in Computer Science.

Tao, L., Burghardt, T., Mirmehdi, M., Damen, D., Cooper, A., Hannuna, S., . . . Craddock, I. (2017). *SPHERE-Calorie*. Retrieved from: https://doi.org/10.5523/bris.1gt0wgkqgljn21jjgqoq8enprr

Thomson, E. A., Nuss, K., Comstock, A., Reinwald, S., Blake, S., Pimentel, R. E., . . . Li, K. (2019). Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. *Journal of Sports Sciences, 37*(12), 1411-1419. doi:10.1080/02640414.2018.1560644

Tran, Hieu. (2019). Survey of Machine Learning and Data Mining Techniques used in Multimedia System. 10.13140/RG.2.2.20395.49446/1.

Uchida, M. C., Carvalho, R., Tessutti, V. D., Bacurau, R. F. P., Coelho-Júnior, H. J., Capelo, L. P., . . . Marchetti, P. H. (2018). Identification of muscle fatigue by

tracking facial expressions. *PloS one, 13*(12), e0208834. doi:10.1371/journal.pone.0208834

Um, T. T., Babakeshizadeh, V., & Kulić, D. (2017). *Exercise motion classification from large-scale wearable sensor data using convolutional neural networks.* Paper presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), ancouver, BC, Canada.

Viola, P., & Jones, M. (2001). *Rapid object detection using a boosted cascade of simple features.* Paper presented at the Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001.

Viola, P., & Jones, M. (2001). Robust Real-time Object Detection. *International Journal of Computer Vision*, 1-25.

Vojciechowski, A. S., Natal, J. Z., Gomes, A. R. S., Rodrigues, E. V., Villegas, I. L. P., & Korelo, R. I. G. (2017). Effects of exergame training on the health promotion of young adults. *Fisioterapia em Movimento, 30*, 59-67.

Wang, B., Tao, L., Burghardt, T., & Mirmehdi, M. (2018). *Calorific Expenditure Estimation Using Deep Convolutional Network Features.* Paper presented at the 2018 IEEE Winter Applications of Computer Vision Workshops (WACVW), Lake Tahoe, NV, USA.

Wang, S.-J., Yan, W.-J., Li, X., Zhao, G., & Fu, X. (2014). *Micro-expression Recognition Using Dynamic Textures on Tensor Independent Color Space.* Paper presented at the 22nd International Conference on Pattern Recognition, Stockholm, Sweden.

Wang, S.-J., Yan, W.-J., Li, X., Zhao, G., Zhou, C.-G., Fu, X., . . . Tao, J. (2015). Micro-Expression Recognition Using Color Spaces. *IEEE Transaction on Image Processing*, 6035-6049.

Wang, W., den Brinker, A. C., Stuijk, S., & de Haan, G. (2017). Robust heart rate from fitness videos. *Physiological Measurement, 38*(6), 1023-1044. doi:10.1088/1361-6579/aa6d02

Wiles, J. D. A., Simon R. Coleman, Damian A. Swaine, Ian L. . (2008). The relationships between exercise intensity, heart rate, and blood pressure during an incremental isometric exercise test. *Journal of Sports Sciences, 26*(2), 155-162. doi:10.1080/02640410701370655

Wu, B. F., Lin, C. H., Huang, P. W., Lin, T. M., & Chung, M. L. (2017). A Contactless Sport Training Monitor Based on Facial Expression and Remote-PPG. *2017 Ieee International Conference on Systems, Man, and Cybernetics (Smc)*, 846-851.

Wu, T., Fu, S., & Yang, G. (2012). *Survey of the Facial Expression Recognition Research.* Paper presented at the Advances in Brain Inspired Cognitive Systems, Berlin, Heidelberg.

Xie, K., Fu, C., Liang, H., Hong, H., & Zhu, X. (2019). *Non-contact Heart Rate Monitoring for Intensive Exercise Based on Singular Spectrum Analysis.* Paper presented at the 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR).

Xu, R., Zhang, C., He, F., Zhao, X., Qi, H., Zhou, P., . . . Ming, D. (2018). How Physical Activities Affect Mental Fatigue Based on EEG Energy, Connectivity, and Complexity. *Frontiers in neurology, 9*, 915-915. doi:10.3389/fneur.2018.00915

Yang, Z., Yang, X., Jin, J., & Wu, X. (2019). Motion-resistant heart rate measurement from face videos using patch-based fusion. *Signal, Image and Video Processing*, 1-8.

Yu, Y. P., Kwan, B. H., Lim, C. L., Wong, S. L., & Raveendran, P. (2013). Video-Based Heart Rate Measurement Using Short-time Fourier Transform. *2013*

International Symposium on Intelligent Signal Processing and Communications Systems (Ispacs), 704-707.

Zhang, Y., Xin, Y., Li, Q., Ma, J., Li, S., Lv, X., & Lv, W. (2017). Empirical study of seven data mining algorithms on different characteristics of datasets for biomedical classification applications. *Biomedical engineering online, 16*(1), 125-125. doi:10.1186/s12938-017-0416-x

Zhang, Z. (2014). *Heart rate monitoring from wrist-type photoplethysmographic (PPG) signals during intensive physical exercise.* Paper presented at the 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP).

Zhu, G., Jiang, B., Tong, L., Xie, Y., Zaharchuk, G., & Wintermark, M. (2019). Applications of Deep Learning to Neuro-Imaging Techniques. *Frontiers in Neurology, 10*(869). doi:10.3389/fneur.2019.00869

Zhu, Z.-A., Lu, Y.-C., You, C.-H., & Chiang, C.-K. (2019). Deep Learning for Sensor-Based Rehabilitation Exercise Recognition and Evaluation. *Sensors, 19*(4), 887.