Università degli Studi di Salerno

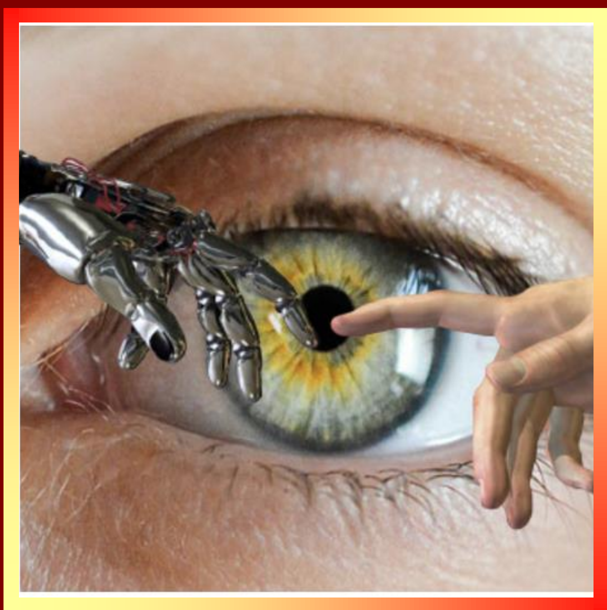
Dipartimento di Informatica

Dottorato di Ricerca in Informatica – XXXV Ciclo

Tesi di Dottorato/Ph.D. Thesis

Soft Biometrics: Periocular Features and applications on Humanoid Social Robots

Lucia CASCONE



Supervisor: **Prof. Michele Nappi**

Ph.D. Program Director: **Prof. Andrea De Lucia**

AA 2021/2022

Curriculum Internet of Things and Smart Technologies

Università Degli Studi di Salerno

Dipartimento di Informatica

Dottorato di Ricerca in Informatica
Curriculum Internet of Things and Smart Technologies
XXXV Ciclo

Tesi di Dottorato / Ph.D. Thesis

Soft Biometrics: Periocular Features and applications on Humanoid Social Robots

Lucia CASCONE

Supervisor: **Michele NAPPI**

Ph.D. Program Director: **Prof. Andrea DE LUCIA**

A.A 2021/2022

A Emanuele,

Ovunque tu sia spero che il mio amore possa arrivare fin lì.

ACKNOWLEDGMENTS

I would like to express my profound appreciation to my advisor, Professor Michele Nappi, for believing in me and allowing me to work in the intriguing and challenging fields of pattern recognition and biometrics. His constant encouragement, unwavering support, and infectious zeal have guided me through these three years of research. My interactions with him were extraordinarily useful in outlining my study objectives and determining how to achieve them. His words of encouragement inspired me to do my best, and his direct and unfiltered honesty helped me grow personally and professionally.

I would like to express my sincerest affection and gratitude to Professor Andrea Francesco Abate for being a guide and reference point for me, a source of inspiration, and a person towards whom I have deep esteem.

I would like to express my gratitude to each and every lab member. Thank you to all of you who continue to inspire and encourage me. Special thoughts to Ignazio, brilliant man with dynamic, cross-cutting, and pragmatic knowledge. I had the honor and pleasure of working with you, volcano of ideas, workaholic, and last-minute man. You would have had so much more to say.

I would like to thank my mother, pillar and strength in all my endeavors, what I am I owe to you.

Finally, I would like to thank you, the person to whom this work is dedicated. You would have been here to celebrate my accomplishments, but now they are ours.

ABSTRACT

The market for biometric technology continues to grow. Biometrics is used in the real world in a wide variety of fields, from surveillance, health care, advertising, Human-Robot Interaction to security and trust. Interest in this field is transversal.

Soft biometrics have emerged in recent years as a potential alternative to and useful ally of primary biometrics (also known as hard biometrics). Any anatomical or behavioral feature that gives some information about a person's identity but does not provide sufficient evidence to accurately determine identity can be called a "soft biometric trait." In addition to the evidence that these characteristics can be used to improve the accuracy of a recognition system, the study of soft biometrics has shown that additional information about people, such as age, gender, ethnicity, and information about emotional and cognitive state, can be inferred from these soft data, demonstrating their broader potential.

Based on this, it is a study area that has garnered considerable interest over the past decade but has never quite taken off. In fact, there has been an absence of a more verticalized and systematic examination of certain characteristics and their purposeful and widespread use in applications such as HRI. For this reason, after a thorough examination of the literature, this thesis focused on two aspects. First, the potential of soft biometrics when integrated with the native capabilities of social humanoid robots was studied to demonstrate how their application can make this type of robot even more effective in the sectors of elderly care, security, and also become a danger in the area of Social Engineering. Then, periocular soft biometric features (blink, eye-movements, fixations, and pupil) were studied in detail to demonstrate their potential for the purposes of recognition, demographic classification, emotion detection, and cognitive state analysis.

CONTENTS

I Introduction
1 Biometric background 3
1.1 Multi-biometric fusion 12
1.2 Soft biometrics 17
1.3 Subject of the research 22
1.3.1 Motivations 23
1.3.2 Main Goals 27
1.4 Thesis outline 28

II Humanoid Social Robots: an Ongoing Revolution
2 Humanoid Social Robots 31
2.1 What are robots? 32
2.1.1 Robotics taxonomy 41
2.2 Human-Robot Interaction 44
2.2.1 Classification of Interactions 48
2.3 The rise of Humanoid Social Robots 54
2.3.1 Interaction: the social consequences 58
2.4 Soft biometrics for Humanoid Social Robots 61
3 State-of-the-art 65
4 Our contribution to literature 71
4.1 Emotion modelling for Social Engineering 73
4.1.1 SASD: Short Attempt to extract Sensitive Data 77
4.1.2 LASD: Long Attempt to extract Sensitive Data 80
4.1.3 Soft biometrics involved 81
4.2 Healthcare support and assistance 87
4.2.1 Case study on Ambient-Assisted Living 90
4.2.2 Object Touching: Pepper vs VPepper 93
4.2.3 Soft biometrics involved 97
4.3 Trust and security in Human-Robot Interaction 100
4.3.1 Semantics and trust: the proposed model 102
4.3.2 The Smart Home case study 106
4.3.3 Soft biometrics involved 109

III Periocular Soft biometrics
5 Soft Periocular Features 115
5.1 Blinks 118
5.1.1 Blinks and mental workload 120
5.2 Eye-movements 122
5.2.1 Mental effort and emotions: evidence from eye-movements 124
5.3 Fixations 126
5.3.1 Fixations: behavioral and emotional-cognitive process analysis 126
5.4 Pupils 128
5.4.1 Pupillary responses to emotional-cognitive stimuli 131
6 State-of-the-art 133
6.1 Recognition system 133
6.2 Demographic classification 136
7 Our contribution to literature 141
7.1 Demographic classification 142
7.1.1 Pupil size for Age and Gender Classification 145
7.1.2 Periocular Data Fusion 154
7.2 Biometric recognition 159
7.3 Cognitive Processes 162
7.3.1 Attention measure 162
7.3.2 Perception and Memory 171

IV Conclusions
8 Final discussion 177
8.1 Open problems 178
8.1.1 Challenges 179
8.2 Future directions 180

Bibliography 187

LIST OF FIGURES

Figure 1.1 Alphonse Bertillon, the inventor of the criminal identification system based on profile and full-face photographs and crucial body measures, is depicted on this anthropometry card (1892). These essential dimensions include body height, body weight, build, complexion, head length, head width, cheek width, right ear and left foot measures, and "unique marks" such birthmarks, scars, and tattoos. 5
Figure 1.2 It is predicted that the global biometric systems market will reach over 43 billion US dollars by 2022. It is expected that the market will significantly grow in the following years, reaching 83 billion US dollars by 2027. An increase of over 120% is therefore estimated from 2020 to 2027 (Source: www.statista.com). 6
Figure 1.3 This thesis examines the potential of soft biometrics. First, when integrated with the native capabilities of social humanoid robots, and then later with a focus on periocular biometrics in terms of blinks, fixations, pupil size, and gaze, to demonstrate its value. 22
Figure 2.1 On the left are the average predicted yearly growth rates for industrial robot installations. On the right is the market trend for professional service robots. Source: World Robotics, October 2022. 32
Figure 2.2 Timeline of the main robotics milestones of the last century. 34
Figure 2.3 Comparison of five generations of robots in terms of their uses and attributes [1]. 35

Figure 2.4 A summary of the taxonomy of robot designs. List of robots shown (left-to-right, top-to-bottom) Bio-inspired robots: HI-4, ERICA, Kodomoroid, NAO, LOLA, Robotic Eyes, Elumotion, EMYS, AIBO, PARO, DragonBot, Keepon, GoQBot, Meshworm, Robotic Flower, Lollipop Mushroom. Artifact-shaped robots: Travelmate, AUR, Google self-driving car, Greeting Machine, YOLO. Functional robots: CoBot, Quadcopter, Beam, TurtleBot. Source: [2]. 42
Figure 2.5 Human Robot Interaction documents by subject area (source: *Scopus* [3], keywords: *Human Robot Interaction*, accessed: *2022-09-20*). 44
Figure 2.6 The keywords of the research articles of 2022 on HRI. The node's size represents the number of occurrences, while its color indicates the relative cluster. Graph produced with the VOSviewer program. 45
Figure 2.7 The different ways a human worker and a robot can work together: coexistence, cooperation, collaboration. 49
Figure 2.8 A spectrum of telerobot control modes drawn from Sheridan's work [4]. On the left is a mechanical linkage in which the human directly controls the robot from another room via sliding mechanical bars, while on the right is a system in which the person's function is confined to observation and monitoring. The dashed lines suggested that there may be intermittent contact [5]. 52
Figure 4.1 Pepper robot and its characteristics. 71
Figure 4.2 The four essential phases of a Social Engineering attack [6]: 1) Information gathering; 2) Developing a trusting relationship; 3) Attack; 4) Covering tracks. 75

Figure 4.3 The methods inside the emotion module. It provides both the prediction of emotions divided by biometric traits, and the overall emotion. 78
Figure 4.4 The SASD approach, with red lines representing negative reactions and green ones representing favourable ones. 79
Figure 4.5 The module B of LASD, with red lines representing negative reactions and green ones representing favourable ones. 81
Figure 4.6 The operational perspective of the proposed case study. 92
Figure 4.7 Example of one of the emergency interventions of VPepper simulated during the case study. 93
Figure 4.8 ArticulationBody joints used to model VPepper. 95
Figure 4.9 The cumulative reward and episode length for a 500K training session as reported by TensorBoard. 96
Figure 4.10 The generic model of trust with domain-specificization. 102
Figure 4.11 The inferred class equivalence in IoTdomotics ontology. 104
Figure 4.12 The computation of a command chain's required level of trust. 105
Figure 4.13 The classification of IoT state information with the associate command. 105
Figure 4.14 The main involvement of Pepper in the proposed eco-system. 107
Figure 5.1 Diagram depicting both the external and internal structures of the eye [7] 116
Figure 5.2 From left to right is shown a pupil in the phase of miosis, in normal conditions, and in the phase of mydriasis. 129

Figure 7.1 Some examples of images shown to participants during the data acquisition process in the GANT dataset. The first column shows images of two landscapes. The last two columns, on the other hand, show images of women and men: in the first row there are images of unknown people while in the second row there are images of two famous actors. 144
Figure 7.2 (left) Distribution of PDM for under 30 and over 30. (right) Distribution of PDM for males and females. 147
Figure 7.3 The precision and recall measurements for the top three best accuracy values. 153
Figure 7.4 The workflow of the proposed fusion strategy. 155
Figure 7.5 Architecture of the proposed multi-modal biometric system. After feature extraction, Feature-level fusion (FF) techniques were applied. 161
Figure 7.6 The deployment diagram of the proposed architecture. 164
Figure 7.7 The Attention Monitor GUI, set by side of teacher Video-lecture software. 165
Figure 7.8 The fixations of participants' gazes expressed as the percentage of slides' duration. 169
Figure 7.9 Analysis of the fixation point for each of the four image categories. A selection of images depicting natural, urban, fractal, and pink-noise settings, from top to bottom and left to right. 173
Figure 7.10 Number of fixation clusters on a natural image that is displayed for the first and fifth time. 174

LIST OF TABLES

Table 2.1 Features of different Human-Robot Interactions. 50
Table 4.1 Images explaining the six primary emotions. Pictures extracted from the CK+ database [8]. 74
Table 4.2 Illustration of a potential interaction. In the first column are two instances of potential security questions, in the second one the questions that Pepper could ask to extract information regarding. 78
Table 4.3 Methods involved, devices, accuracy and the strength coefficient of the biometric trait. 108
Table 7.1 Age range of participants in GANT dataset. 143
Table 7.2 Comparative analysis among the classifiers in terms of age classification. The results are presented in terms of the best configuration for each classifier achieved through an exhaustive search of its hyperparameters. In bold, the best-performing classifiers are highlighted, while the worst are in italics (underlined). 149
Table 7.3 Comparative analysis among the classifiers in terms of gender classification. The results are presented in terms of the best accuracy for each sample achieved through an exhaustive search of their hyperparameters. 152
Table 7.4 The classifiers are combined according to the achieved accuracy. 154
Table 7.5 Spearman's correlation coefficients with respect to the pairs of the three modalities. 156

ACRONYMS

AAL	Ambient Assisted Living
ADA	AdaBoost
AGVs	Automated Guided Vehicles
AI	Artificial Intelligence
AMRs	Autonomous Mobile Robots
BG	Bagging
BMI	body mass index
CCA	Canonical Correlation Analysis
DET	Decision-Error Trade-of
DL	Deep Learning
DTC	Decision Tree Classifier
DT	Digital Twin
EER	Equal Error Rate
FAR	False Acceptance Rate
FRR	False Rejection Rate
GB	Gradient Boosting
GNB	Gaussian Naive Bayes
GP	Gaussian Processes
HRI	Human-Robot Interaction
HSRs	Humanoid Social Robots
IK	Inverse Kinematics
IoT	Internet of Things
KNN	K-Nearest Neighbor
LASD	Long Attempt at Sensitive Data Extraction
LDA	Linear Discriminant Analysis
LPS	Levator Palpebrae Superioris
LinSVC	Linear Support Vector Machines

ML	Machine Learning
MLP	Multilayer Perception
OO	Orbicularis Oculi
PCA	Principal Component Analysis
PDM	pupil diameter mean
PL	average left pupil size
PLR	average left and right pupil size
PLRI	average left and right pupil size + code of the relative image
PR	average right pupil size
PPG	Photoplethysmography
QDA	Quadratic Discriminant Analysis
RF	Random Forests
SASD	Short Attempt at Sensitive Data Extraction
SFAR	Spoof-False Acceptance Rate
SGD	Stochastic Gradient Descent Classifier
SVM	Support Vector Machines
ZTA	Zero Trust Architecture

Part I

INTRODUCTION

1 BIOMETRIC BACKGROUND

Nemo liber est, qui corpori servit.

— Seneca, Epistulae, 92, 33

Biometrics (from Greek bos = "life" and métron = "measuring") is the study of biophysical and behavioural properties using mathematical and statistical models. The acquisition of biometric information, whether the subject is aware of it or not, allows for the extraction of latent knowledge. Biometrics exploits measurable physiological, physical, and behavioural traits of humans for purposes of recognition and analysis [9]. Physical biometrics are the physical characteristics of an individual [10]; behavioural traits are the attributes describing the personality and behaviour of a subject [11]; physiological biometrics, on the other hand, record the unique pattern of a user's automatic bodily functions [12].

Humans have always used physical traits to differentiate themselves from one another; as soon as a person interacts with a known person, the brain is able to recognise him or her based on his or her voice or appearance. Similarly, a biometric system recognises a person based on "*who he or she is*," regardless of "*what he or she has*" or "*what he or she knows*"-this property has caused biometrics to find particular success in the security industry: an identifying object can be lost, stolen, or worse, duplicated; the things one knows, such as passwords and codes, can be forgotten; "*what one is*" remains. In the context of security, biometrics employs a collection of technologies that enable the identification and authentication of a person based on his or her physical (face, fingerprints, iris, hand shape, etc.) and behavioural (gait, typing speed, handwriting, etc.) features. Biometric authentication entails comparing the data collected from an individual with the biometric template that was previously registered for that individual. In this instance, the question posed is "*are you really*

Mr. X? "; hence, a "*one-to-one*" matching is performed. In contrast, biometric identification is a "*one-to-all*" matching, consisting of identifying a person by comparing his pattern to those in the database. Therefore, the question is "*who are you*?" When neither authentication nor identification is specified, the term "recognition" is used instead.

Human beings have always felt the need and had the need to be both able to prove their identity, irrefutably, and to recognize the identity of others without any shadow of a doubt, making use of what makes a person different and unique.

As early as prehistoric times, man had already sensed that certain characteristics, such as the trace of his finger, were sufficient to identify him and therefore used the imprint to "sign." Handprints and footprints used to "sign" prehistoric paintings have been found in a cave in Nova Scotia (Canada). In 2500 B.C.E., the ancient Egyptians, faced with the problem of having to recognize laborers to whom they would pay compensation for their work in the construction of the pyramids, exploited a system of person identification based on height and arm length. In 500 B.C., the Assyro-Babylonian civilization exploited fingerprints on clay tablets to validate contracts. In the 2nd century B.C., Chinese Emperor Ts'in She began using fingerprints to authenticate certain seals, and, also in China, in 1300 A.D., children's hands and footprints were acquired to distinguish them.

This primordial use of biometrics was later set aside only to be rediscovered in the mid-19th century by William James Herschel, an English official tasked with building roads in Bengal who asked subcontractors to sign contracts with their fingerprints, thus facilitating their identification in case of default. Also in the 19th century, biometrics was first used in the judicial and forensic spheres: Alphonse Bertillon, employed in the forensic police, devised a method, later going down in history as Bertillonage, for identifying repeat criminals. He recorded signs such as tattoos, scars, etc., and measurements of certain anatomical features, estimating the infrequent probabilities of duplication if many features were used (Figure 1.1). However, this technique,

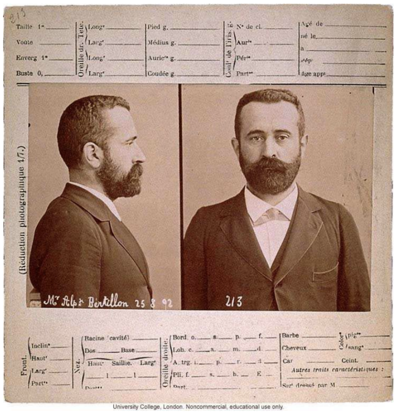


Figure 1.1: Alphonse Bertillon, the inventor of the criminal identification system based on profile and full-face photographs and crucial body measures, is depicted on this anthropometry card (1892). These essential dimensions include body height, body weight, build, complexion, head length, head width, cheek width, right ear and left foot measures, and "unique marks" such birthmarks, scars, and tattoos.

which often proved successful, offered no guarantees of reliability: measurements taken by two different people using different instruments were affected. This system showed all its vulnerability in 1903, when a certain Will West was imprisoned because he was confused with another William West, another African American prisoner. This ambiguity simultaneously discredited the three methods in use until then, such as personal name identification, mug shots, and Bertillon's physical measurements. In the meantime, a new system based on fingerprints started to gain popularity as early as 1892. It was created by anthropologist Sir Francis Galton, who figured that the chances of two fingerprints being the same, even for twins, were so low that it was a reliable way to measure.

Over the past century, the use of biometric technologies has grown exponentially (Figure 1.2); in 2001, the *MIT Technology Review* listed biometrics as one of the 10 technologies that will change the world. Biometric technologies for the next-generation biometric market are largely focusing on law enforcement solutions and services. The U.S. is a highly lucrative market for the

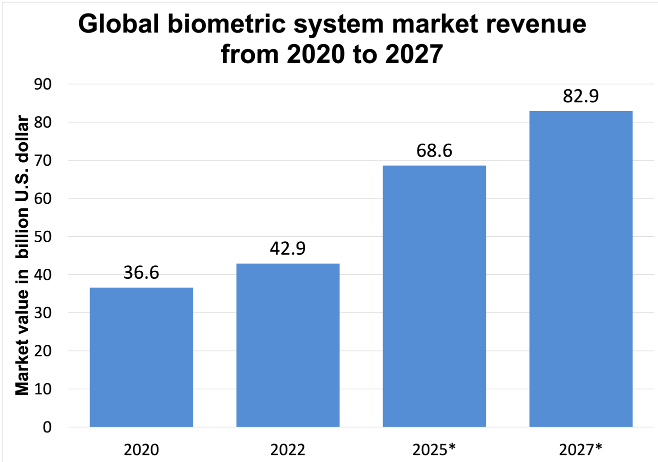


Figure 1.2: It is predicted that the global biometric systems market will reach over 43 billion US dollars by 2022. It is expected that the market will significantly grow in the following years, reaching 83 billion US dollars by 2027. An increase of over 120% is therefore estimated from 2020 to 2027 (Source: www.statista.com).

development of biometric applications, pivoting precisely on the presence of several prominent players meeting the demands for biometric solutions in the defense and law enforcement sectors.

However, in the coming years, demand for biometric systems is expected to grow steadily, not only in the military and defense sectors but also, for example, in the automotive and healthcare sectors. This increasingly pervasive use of biometrics in the government sector has given a major boost to research in the private sector as well, thanks precisely to a sharp increase in demand. Governments' emphasis on supporting the digital transformation of economies has led companies to increase their investments, particularly in voice and facial biometrics. The market for civil applications linked to this area has benefited from this: just think of the widespread use of voice and facial biometric authentication systems that new generation smartphones and more are now equipped with. The future market for biometric technology will certainly also cover biometric applications in online shopping, mobile banking, and e-commerce sites.

The integration of biometrics with Artificial Intelligence (AI) has opened a new avenue for stakeholders in this biometrics market. State-of-the-art systems with AI capabilities are widely marketed for user authentication in healthcare facilities and are gaining momentum for the real-time interpretation of emotions in patients as well.

But what are the biological measures that qualify as biometric traits? Among the different biometrics which one to choose? What makes one biometric perform better than another? In their groundbreaking work, Jain and his colleagues figured out the main things a biometric trait must have to some degree to be considered one [13]:

1. ***Universality.*** Every individual possesses this trait.
2. ***Uniqueness.*** The feature must have a high discriminating value.
3. ***Permanence.*** The characteristic must be sufficiently invariant over a given period of time.
4. ***Collectability.*** The feature can be measured easily.
5. ***Performance.*** The operations required to achieve the goal must be fast and not time- or memory-consuming.
6. ***Acceptability.*** It indicates the extent to which people are willing to tolerate the biometric capture procedure.
7. ***Eludibility.*** It reflects the ease with which the system can be circumvented.

No biometric feature, however, fully satisfies the properties just described. Following is a brief introduction to some of the most popular physical biometric characteristics:

- ***Face.*** Facial recognition is a noninvasive technique, and facial characteristics are likely the most widely used biometric identifiers among humans. Face recognition applications span from static, controlled "mugshot" authentication to dynamic, uncontrolled face identification against a crowded background. Popular facial recognition methods are based on the location and form of facial characteristics, such as the eyes, eyebrows, nose, lips, and chin, and their spatial relationships, or the overall (global) analysis of the face image. While the authentication performance of commercially available face recognition systems is adequate, they impose a variety of restrictions on how facial photos are acquired, sometimes requiring a fixed, simple background with controlled illumination. In addition, these algorithms struggle to match facial photos collected from two different perspectives, under different lighting conditions, and at separate times. It is debatable if the face alone, in the absence of contextual information, is sufficient for recognising a person among a vast number of identities with an extraordinarily high degree of certainty. In order for a facial recognition system to be effective in practice, it must automatically detect whether a face is present in the recorded image, locate the face if it is present, and recognise the face from a general view point (i.e., from any stance) under varying environmental conditions [14].

- ***Fingerprint.*** Humans have utilised fingerprints for identification purposes for decades. It has been demonstrated that the matching (i.e., identification) accuracy of fingerprints is very good [15]. The pattern of ridges and valleys on the surface of a fingertip, whose construction is decided during the first seven months of fetal development, constitutes a fingerprint. It has been scientifically shown that identical twin fingerprints are distinct, as are the fingerprints on each finger of the same individual [16]. The marginal cost of embedding a fingerprint-based biometric within a system (e.g., a laptop computer) is now cheap for a wide variety of applications. The precision of currently available fingerprint recognition systems is sufficient for authentication systems in a variety of applications, including forensics. Multiple fingerprints of a person (e.g., ten prints are used) provide additional information to enable the identification of millions of individuals on a broad scale. An issue with large-scale fingerprint recognition systems is that they require an enormous amount of processing resources, particularly when operating in identification mode.

A tiny portion of the population may have fingerprints that are unsuitable for automatic identification due to genetic, age, environmental, or occupational factors (e.g., manual laborers may have a huge number of cuts and bruises on their fingerprints that are constantly changing) [14].

- ***Iris.*** The iris is the circular portion of the eye, limited on either side by the pupil and the sclera (white of the eye). The visual texture of the iris is generated during fetal development and stabilizes throughout the first two years of life (the pigmentation, however, continues to change for a lengthy period of time). The intricate iris texture contains information that can be used for identification. The precision and speed of recently deployed iris-based recognition systems are encouraging and suggest the viability of large-scale iris-based identification systems. Even the iris of identical twins has distinct characteristics. It is feasible to identify contact lenses with a counterfeit iris. The eye's hippus movement may also be utilised as a measure of liveliness for this biometric [14]. Although early iris-based recognition systems required substantial human participation and were costly, modern systems are more user-friendly and economical. Despite the fact that iris systems have a relatively low False Acceptance Rate (FAR) compared to other biometric features, their False Rejection Rate (FRR) can be quite high (more details on these two metrics below).

The choice of which feature to use will depend almost entirely on the requirements of the recognition system into which it is to be integrated. A biometric recognition system is essentially a system that acquires biometric data from an individual, extracts a set of salient characteristics from the data, compares them with information stored in a database, and performs an action based on the result of the comparison. A common structure of the functioning of different biometric systems has been described in the aforementioned work by Jain et al. [13]. In particular, it turns out that a typical system of biometric recognition consists of two main phases: enrollment and recognition. Enrollment is the step in which the system gathers the individual's biometric data, extracts the salient features, and stores them along with an identifier used to associate the collection of features with the individual in the system database. Recognition, during this step, the system takes the individual's biometric data, extracts the set of features, and compares them to the templates in the database to identify a match, thereby proving the user's identity.

As previously stated, biometric identification systems are far more secure than systems based on conventional methods, primarily because biometric qualities are not susceptible to the common difficulties associated with passwords, PIN codes, ID cards, etc. Specifically, biometric characteristics cannot be forgotten, are more difficult to steal, and are more counterfeit-resistant. However, biometric recognition systems are not entirely secure. Complications may develop as early as the data gathering phase: Bharadwaj et al. [17] noted that even the finest biometric algorithms are badly impacted by overly noisy and low-quality data. In addition, they highlighted the external elements that influence the quality of biometric data, including the subject's physical condition (fatigue, distraction, injury, clothes, etc.), environmental circumstances (humidity, temperature, illumination, etc.), and interaction with the sensors (cleanliness of the sensor, initial position of the subject, etc.).

Variability in the acquisition settings therefore explains why two biometric samples from the same subject are virtually never identical: even modest variations in the manner of data capture can result in substantial modifications to the biometric characteristic template. Specifically, intra-class variation refers to the rate of change in the templates of the same biometric feature received from the same user, and inter-class variation pertains to different users. A biometric characteristic is advantageous when its intra-class variance is low and its inter-class variation is high [14].

A successful biometric system must be able to handle the following two kinds of errors [18]:

- ***False Rejection Rate (FRR).*** Attributing two biometric measurements from the same individual to distinct individuals. It is also referred to as a "Type I error." False rejections frustrate authorised users, reduce productivity owing to poor access circumstances, and necessitate the spending of resources to revalidate allowed users.

- ***False Acceptance Rate (FAR).*** Attributing the biometric measures of two distinct individuals to the same individual. If an organization's biometric control is creating a large number of erroneous rejections, the overall control may need to reduce the system's accuracy by collecting fewer data when authenticating subjects. When the number of data points decreases, the organisation runs the danger of seeing a rise in the erroneous acceptance rate. The organisation faces the danger of illegal entry. This issue is also known as a Type II error. A false acceptance is worse than a false rejection because the majority of organisations would rather reject actual subjects than welcome impostors.

Both types of mistakes depend on the decision threshold employed by the template matching and nearest neighbor approaches. Choosing a high acceptance threshold will result in a high FRR (low FAR). Choosing a low threshold will also result in a large FAR (low FRR). The value of the threshold is derived using a Decision-Error Trade-of (DET) curve, which is a plot of FAR and FRR. The FAR and FRR values of a verification system define distinct DET curve points. When these rates are equal, the threshold values for false acceptance and FRRs are selected. This quantity is known as the Equal Error Rate (EER). It implies that the proportion of false acceptances and false rejections is identical.

In the presence of spoof attacks, [19] proposes the evaluation of a third type of error, termed SFAR, namely the rate of accepted spoof attacks. The spoofing attack, which is one of the adversarial attacks [20] targeting the recognition system, is the process of circumventing a biometric system by delivering a replica of a legitimate user's spoofed biometrics. Despite the fact that spoofing approaches for biometric technology differ, they nonetheless entail presenting a phony biometric sample to the sensor. Therefore, it is important to collect a biometric sample from a legitimate user. Consequently, the Spoof-False Acceptance Rate (SFAR) is a helpful metric for determining the proportion of false acceptance when one or more modalities have been effectively spoofed. As described in [19], SFAR should be separated from "standard" FAR as it involves a non-zero effort to spoof the system.

This kind of threat is especially dangerous for uni-modal biometric systems, which use only one biometric trait to identify a person. In fact, once the biometric information has been stolen, there is no other security measure that keeps someone from getting into the system. Therefore, there is a strong belief in the biometric community about the security of multi-modal systems against spoofing attacks, i.e., those systems for which it would be necessary for an intruder to spoof all the traits involved (or at least more than one).

1.1 MULTI-BIOMETRIC FUSION

A uni-modal biometric system, which is based on a single biometric trait, has several problems and limitations due to a lack of data, the poor quality of the information obtained, or, in some cases, low discriminatory power. In fact, biometric recognition systems based on a single trait, such as facial recognition or fingerprint recognition, can be accurate in many cases but can also be subject to errors due to factors such as the inherent variability of biometric signals or environmental conditions. For example, facial recognition can be affected by age, gender, race, facial expression, and lighting conditions, while fingerprint recognition can be affected by skin condition and the use of gloves.

A multi-biometric system, that is, a system that combines different biometric traits, can help increase performance and consolidate the information collected to overcome these obstacles. Thus, one of the main challenges in multi-biometric fusion concerns just that: achieving a good combination of the different biometric modalities so as to maximize recognition accuracy. Multi-biometric fusion can be complex to implement and require a significant amount of computational resources. A crucial challenge in multi-biometric fusion concerns biometric data management and privacy protection, as the fusion of different biometric modalities may require the collection and processing of a significant amount of sensitive personal data. The availability of multi-biometric data corresponding to multiple biometric traits increases concerns about the compromise of subjects' privacy. Therefore, it is necessary to impart security and confidentiality to stored templates. In real-world applications, archived templates require periodic updating; thus, a pertinent issue is the modification of an individual's stored biometric data in order to account for changes within the class. Aging and physical illness can change an individual's biometric trait. Updating an individual's multi-biometric templates over time can be a daunting task and may inadvertently result in identity theft, where an impostor can exploit the template update mechanism to assume the identity of an enrolled individual. An adaptive fusion system should be able to deter such attacks while accounting for the inevitable changes in data distribution that occur over time.

The fusion might occur when multiple sources and levels are considered. The most common questions are: *what* to merge [21], *when* to merge, *how* to merge.

According to *what* to fuse, the systems can be described as follows:

- ***Multi-sensor.*** These systems use several sensors to collect data on a single biometric feature [22]. This strategy is particularly suitable when the sensors needed to grab the desired characteristic are all available and properly running.

- ***Multi-algorithm.*** Multiple feature extraction techniques are applied to the same data acquired from a sensor, resulting in distinct templates that provide alternative perspectives on the same feature. The fundamental concept is to be able to extract multiple properties from the same sample using different techniques [23]. If the properties derived by two distinct algorithms are complementary, system performance can be enhanced.

- ***Multi-instance.*** Multiple instances of the same biometric attribute are acquired using a single sensor and an extraction technique; for instance, instead of capturing the fingerprint of a single finger, all the fingers of the hand might be examined, both the left and right iris are employed for iris recognition and so on [24]. The benefits of this method include a decreased susceptibility to noise as a result of the greater number of samples acquired, greater similarity within a given class, and greater diversity between classes [25].

- ***Multi-sample.*** The same sensor is utilised to record the same biometric feature; but, rather than capturing data instantly, numerous samples are collected over a defined time interval. This is an especially advantageous mode for video sensors, which may collect numerous consecutive frames while minimising the "damage" caused by subject motion. To develop a facial recognition system, for instance, you can extract information from the same video by merging the data gathered from a single sensor on numerous video frames. [26].
- ***Multi-modal.*** In this method, numerous biometric characteristics are evaluated. Physical and/or behavioural features can be combined into a single system. This option may be preferable when security is essential to protect sensitive data [27]. The employment of multi-sensor, multi-instance, and multi-sample modes aids in the management of noise during data collection, thereby increasing the likelihood of receiving high-quality data. Support for an additional biometric feature, on the other hand, prevents spoofing and brute-force attacks. The addition of even a single biometric characteristic provides a "line of defense" because a hacker would have to try an exponential number of input data possibilities, whereas a spoofing assault requiring a greater number of traits would take more time and resources.

Once the data has been collected, one may wonder *when* it is convenient to merge them. Regarding this, it is therefore possible to define different levels of fusion based on the type of information provided. These solutions can be separated into two macro-areas: pre-classification with fusion *prior* to matching (sensor level, characteristic level), and post-classification with fusion *subsequent* to matching (score level, rank level, decision level).

- ***Feature-level fusion.*** Feature-level fusion refers to techniques that perform fusion on several extracted features from the same or distinct input data. It combines different representations to provide a single representation for a given individual. Concatenation or summation [28] are examples of this class of approaches. This strategy is frequently employed by multi-biometric cryptosystems, which combine characteristics from numerous biometric sources to enhance security and anonymity. This could correspond to numerous feature sets for the same biometric trait, such as pupil size and number of blinks from the periocular region, or different facial traits. It may also correspond to characteristics extracted from several modalities, such as palmprint and fingerprints.

- ***Sensor-level fusion.*** Typically, sensor-level or data-level fusion applies to multi-sensor or multi-sample algorithms in which data are integrated immediately following acquisition. Thus, data fusion is performed directly on the raw data prior to feature extraction [21]. This corresponds to the direct pixel-level combining of face images acquired from a camera in the context of a face recognition module. Pose variations such as frontal, left, and right can be captured to generate multiple instances of a face. A mosaicing technique can be used to fuse samples together in order to obtain a combined representation of the face [29].

- ***Score-level fusion.*** Score-level fusion refers to algorithms in which the match scores generated by various matchers are combined. Numerous score fusion approaches have been proposed in the scientific literature. Transformation, classification, and density fusion strategies are the three basic categories of scoring combining rules. Transformation-based score fusion is the most intuitive and widely used score fusion technique, as it is simple to make. For a given sample, it allows to combine the scores obtained from the different algorithms and generate a new unique score, using a function, to which these previously generated normalized or standardized scores are given as input. So, it consists of simple algebraic manipulation of the scores through a specific function. Common fusion methods at this level include mean score fusion, maximum score fusion, and minimum score fusion, where the mean, maximum, or minimum score of many matchers is considered the final result. In classifier-based score fusion, score vectors obtained by biometric algorithms are considered feature vectors that are in turn discriminated as genuine or impostor scores. Therefore, the classifiers learn the relationship between the various score vectors, which are treated as the new characteristics that are used to solve the classification task. Density-based score fusion is based on the likelihood ratio test and it requires explicit estimation of genuine and impostor match score densities.

Due to the ease with which scores generated by commercial matchers can be accessed, score-level fusion is the most frequently documented type of fusion in the literature. Most commercial matchmakers do not provide straightforward access to features or, at times, raw data.

- ***Rank-level fusion.*** A rank-level fusion is applied when a ranked list of matching identities can be obtained from each algorithm. In identification tasks where a given probe image is compared against a gallery of images, the matcher frequently generates a ranked list of matching identities. The algorithm will assign a higher rating to a template that more closely matches the query. Using techniques such as Borda count, logistic regression, and the highest rank method, researchers have combined the rank lists of multiple matchers [30]. In situations where access to features or match scores is limited, rank-level fusion is frequently judged useful.

- ***Decision-level fusion.*** Fusion at the decision level corresponds to algorithms that accomplish fusion at the decision level [31]. Majority voting is one of the most prevalent decision-level fusion methods. The ultimate decision is the outcome of combining the decisions of n matchers or classifiers based on a majority vote. Fusion at the decision level has the benefit of working well with black-box systems, in which only the final decisions are available. In many commercial systems, access to features, scores, and rankings may not be possible.

In some applications, in addition to the match score or identity decision provided by individual biometric matchers, researchers have added other information to the typical biometric pipeline to increase its performance. Examples of such ancillary information include measures that indicate the quality of the captured biometric sample or some additional user information known as ***soft biometrics***.

1.2 SOFT BIOMETRICS

Soft biometrics have emerged in recent years as a potential alternative and beneficial ally to primary biometrics (also known as "hard biometrics"). Several things contribute to this, such as the non-intrusive nature of features or traits [32], that they are independent at the modality and feature level [33], that each trait has a semantic description [34], and the fact that identification and retrieval are done in a continuous way [35]. However, the field is still facing several challenges, and there are a number of gaps that need to be filled before it can be considered a replacement for or a benchmark for traditional biometrics.

Soft biometrics refer to features that convey some information about the individual but lack the distinctiveness and permanence necessary to distinguish between them. Several characteristics meet this description, such as gender, weight, height, age, eye colour, ethnicity, etc. Soft biometrics appear to contain a huge variety of possible attributes, which can be difficult to figure out in the absence of further clarification. Dantcheva et al. provided, in their work [34], a more exact description of the term "soft biometrics" as well as an example of its possible applications. In particular, they defined soft biometric characteristics as "*The human-specific physical, behavioral, or material accessories that are associated with an individual and that can be useful for recognizing that individual. These attributes are typically gleaned from primary biometric data, are classifiable in pre-defined human understandable* *categories, and can be extracted in an automated manner.*"

Face, body, and accessories were the three categories established by Dantcheva and her colleagues as part of an early classification system. In a subsequent study [32], Dantcheva et al. enhanced this categorisation by suggesting a new, more schematic classification model. Based on this and other research, we present the following classification of soft attributes:

- ***Demographic.*** Age, gender, ethnicity, and hair color fall into this category.
- ***Anthropometric and geometric.*** The geometry of the body and face, and body measurements in general.
- ***Medical.*** Health conditions, BMI, weight, wrinkles, and skin lesions are examples of this group.
- ***Materials.*** Hats, scarves, bags, and other general accessories.
- ***Behavioral.*** Human language, facial expressions, and gait fall into this category.

Features can be extracted in the form of labels, measurements, and descriptors. In particular, it is interesting to note how the nature of the above soft biometric traits can be binary (e.g., presence of hats), continuous (weight), or discrete (gender).

Soft biometric characteristics can also be related to a given value for the attributes of uniqueness and permanence, where distinctiveness evaluates the capacity of a trait to differentiate subjects within a group and permanence refers to the duration of the trait's invariance. Regarding the latter attribute of permanence, two kinds of biometrics can be differentiated: those that are temporary (such as medical biometrics) and those that are permanent (e.g., body measurements). Both characteristics are also closely related to the trait's continuous or discrete nature. In particular, it turns out that continuous biometric traits exhibit a higher amount of uniqueness than discrete traits, owing to their greater range of possible values. The contrast between hard and soft biometrics is made obvious by the analysis of these two properties. In actuality, both properties have a lower value for soft biometric traits compared to traditional ones. Soft biometric data subject to changes, such as a person's hair color (either by choice or because it is affected by time), can potentially affect the accuracy of recognition systems. One way to handle this problem is to use multiple types of biometric data to improve the overall accuracy of the system. For example, a system that uses both recognition of body geometries and identification of various parts of the face may be less affected by a change in hair color than a system that relies solely on this labile information. Extreme case that is not implemented in a real-life scenario. Another approach is to use machine learning algorithms that can adapt to changes in the data they are processing. For example, if a facial recognition system is trained on a dataset that includes people with a wide range of hair colors, it may be able to handle hair color changes more effectively than a system that was trained only on a narrow range. It is also worth noting that some recognition systems are designed to be more resilient to changes in appearance than others. In general, however, the impact of changes in soft biometrics on recognition accuracy will depend on the specific system and the methods used to handle those changes.

Subjectivity, in addition to these two properties of uniqueness and permanence, can also be evaluated. It refers to a person's capacity to recognise the biometric feature unambiguously. Indeed, it is easy to understand how the very nature of soft biometric features includes ambiguity not just for an automatic recognition system but even for humans: a simple example would be hair colour, which has a wide range of tints but is typically categorised as blonde, brown, black, or red. The difficult step is then determining how to classify the various details in a way that everyone can agree on. For this type of difficulty, it may be deduced that characteristics with high subjectivity have a lower value of uniqueness.

In general, three different application scenarios can be outlined in the literature for soft biometric traits when their purpose is to be used in a recognition system:

- ***Fusion-approach.*** In conjunction with hard biometrics, soft biometrics are utilised to enhance system performance. The very notion of combining them with primary characteristics may be the reason why these traits are said to be soft. Soft biometrics can be traced all the way back to the 18th century, when the Bertillon system was used to identify criminals based on how they looked. The attributes used to quantify the physical description were defined as anthropometric measures, including head length, head width, middle finger length, left foot length, and cubit length. These characteristics were put into two groups: body geometry and facial geometry. Both of these were accompanied by a mug shot of the person, since it was clear from the start that this information was not enough on its own. Thus, the lack of uniqueness capacity of these features, as well as the idea of using them as additional decision support, were already obvious.

- ***Search Space Reduction approach.*** Use information that has been extracted, such as gender, race, skin color, or hair color, to narrow down the search space for a given target sample.

Using soft biometrics such as ethnicity and hair colour with a facial recognition system can help reduce the FRR of some individuals without significantly influencing the FAR, as determined by Marcialis et al. [36]. On the basis of the extracted set of biometric features and the existence of specific soft biometric qualities, a probabilistic framework was provided to determine if an input face image belonged to a specific user. Due to the fact that certain soft biometric traits (such as a specific hair colour) are associated with a limited number of users (and not others), these attributes can be leveraged to increase recognition accuracy. Specifically, the fundamental concept is that users having a given high soft biometric discriminant can be identified more accurately. Identification of such users could be important for optimising the usage of soft biometrics by finding a suitable method to combine soft biometric data into the score calculated by the hard biometric. It is simple to extract hair colour and ethnicity information from facial photos, but only a small number of people with highly discriminating hair colour or ethnicity should be associated with this information.

- ***Stand-alone system.*** Long ago, the study of soft biometrics was regarded as a subfield of hard biometrics, with research concentrating on hybrid identification systems. The fusion framework is capable of group identification and ongoing authentication throughout an online session. In each instance, the overall objective was to improve recognition performance.

  Heckathorn et al. [37] presented one of the first methods for identifying people that only used soft biometrics. The authors proposed using attributes such as scars, birthmarks, tattoos, eye color, ethnicity, and gender, along with five biometric measurements of height, forearm length, and wrist width, to identify a given subject. The model has worked well in situations where certain biometric hardware isn't available, and it's important to keep people's identities secret by not storing biometric images. The model was built on the concept of "indicator interchangeability," which is based on the idea that combining traits with low discriminatory power can result in a much more accurate system.

The research on soft biometrics is primarily focused on computer vision and automatic learning, and it has been studied from a variety of perspectives. Despite the fact that the primary use of soft biometric traits in the literature has been for recognition, the study of such features can be applied in a variety of contexts.

In fact, the study of soft biometrics is especially fascinating since it can also aid in the semantic interpretation of a person's thoughts, emotions, behaviours, and mental effort, allowing for the recognition of his or her emotions and cognitive state [38]. Their reflection can be seen through physical characteristics such as pupil size change, behavioural characteristics such as gait or touching, and physiological characteristics such as heartbeat.

1.3 SUBJECT OF THE RESEARCH

Soft biometrics is a research topic that has attracted a great deal of attention over the past decade but that has, effectively, never been fully explored or exploded. Any anatomical or behavioral feature that provides some information about a person's identity but does not provide sufficient evidence to accurately determine the identity of a person can be called a "soft biometric trait." Although soft biometric traits may not possess sufficient distinctiveness or uniqueness to enable highly accurate recognition, they have been used extensively to filter large biometric databases, as a way to improve the speed or search efficiency of the biometric system, or combined and merged with other more discriminating biometrics. Also of interest is the potential of their application in health and cognitive fields for monitoring health status and detecting fatigue, stress, etc.

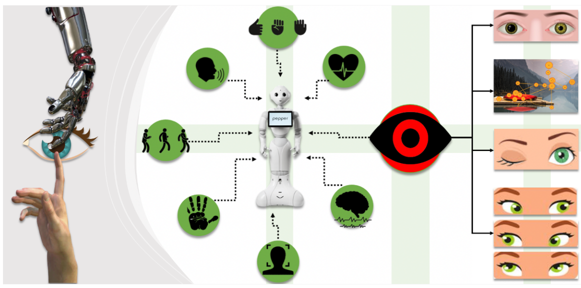


Figure 1.3: This thesis examines the potential of soft biometrics. First, when integrated with the native capabilities of social humanoid robots, and then later with a focus on periocular biometrics in terms of blinks, fixations, pupil size, and gaze, to demonstrate its value.

Therefore, despite this wide range of possibilities, a thorough study of certain soft features is lacking in the literature, and it seems evident that, in general, they have very rarely been considered for applications of HRI. For these reasons, in this thesis, after a careful review of the literature, we focused on two aspects (Fig. 1.3). The potential of soft biometrics when integrated with the native capabilities of HSRs was investigated, and then a particular focus was made only on periocular ones to show their potential.

Indeed, some periocular characters showed rather primitive research. Therefore, it was decided to conduct a more in-depth study of what their potentials, limitations, and peculiarities are, both individually and when combined. Their correlation with certain situations of cognitive effort or association with a certain gender or age group, rather than recognition potentials, is known from the literature. However, soft periocular biometric information has rarely been exploited to make inferences about these aspects. Therefore, based on these studies that corroborated its contribution, verticalized work was then carried out investigating the use of these modalities for the above purposes.

1.3.1 *Motivations*

Faced with a digital transformation that permeates every aspect of daily life, from accessing one's smartphone to authorising an online transaction to monitoring one's health status, the use of biometrics is expanding across a vast array of application domains.

Traditionally, systems that utilise biometric characteristics do so for recognition purposes and are generally uni-biometric, that is, they utilise a single biometric characteristic. Problems associated with this option include missing information (e.g., a face masked by sunglasses), poor data quality (e.g., dry fingerprints), identity overlap (e.g., pictures of twins' faces), and low discriminability (e.g., hand geometry).

Therefore, it often becomes necessary to use multiple biometric traits to improve the accuracy of the system. In this setting, the application and analysis of so-called soft biometric features are getting popular.

In addition to the evidence that this information can be used to improve the accuracy of a recognition system, the study of soft biometrics has demonstrated how additional information about an individual, such as age, gender, ethnicity, and information about his or her cognitive and emotional state, can be inferred from these soft data, thereby demonstrating the potential of even those that, if analysed for recognition purposes, would not be the most suitable. Literature has also showed the potential of calculating body mass index (BMI) from facial images, leading to the possibility of assessing health condition based on biometric data.

What are the additional benefits of using these soft biometric traits? We summarize the main ones below:

- ***Explainability.*** Attributes have a semantic interpretation in the sense that they can provide a description that humans can understand; for example, "old, short, male." Thus, it bridges the gap between machine and human descriptions of a person. Because of this, they are especially useful for applications like video surveillance, where they work well with the way people see their surroundings. In other words, when a human tries to verbally describe a person, obvious features regarding the person's appearance, such as gender, age, height, and color of dress, are often used (e.g., in police reports). This enables the use of soft biometrics in contexts where traditional biometrics may be inadequate, as Klontz and Jain [39] argued in the context of the 2013 Boston bombings.

- ***Robustness versus low data quality.*** Despite the poor quality of biometric data collection, it is possible to infer the existence of certain soft biometric characteristics. If the supplied iris image is of poor quality, for instance, the surrounding periocular information could be employed for recognition rather than the iris itself.

- ***Side cost.*** With the collection of the primary biometric trait, it is frequently feasible to recognize the secondary soft biometric as well. Then, it is possible to evaluate and study as much information as possible for the same price and with the same number of sensing devices.

- ***Non-cooperation.*** Soft biometrics can frequently be gathered without the subject's agreement or cooperation. For example, it is possible to remotely deduce a person's ethnicity or gender.

- ***Privacy.*** Unlike hard biometric traits that can uniquely identify a person and thus be a deterrent to users who want to avoid being "tracked," soft biometrics simply apply "labels" to the subject using features that are clearly visible to the naked eye. In addition, identification performed using soft traits does not require the system to store information about the user, as the data captured is compared with predetermined values that are not directly related to the particular person.

In spite of these observations, an examination of the appropriate literature reveals a dearth of systematic investigations and purposeful, effective integration of soft biometric features in HRI applications. Scientists, philanthropists, educators, politicians, leaders, and philosophers have been captivated by the functioning of the human brain throughout history. From Michelangelo to Lomonosov, Da Vinci to Einstein, there have been various attempts to decipher the mystery of the human mind and to mimic its functioning, first with simple mechanical devices and then, in the 20th century, with computers, software, and robots.

Our society is developing and deploying domestic and industrial robots, intelligent software agents, virtual-world avatars, and other artificial beings for a variety of tedious and dangerous duties, as well as for entertainment and companionship. In particular, the use of so-called social robots has gained particular attention in recent years. These robots may provide an alternative to human social and emotional interaction. However, if scientific and technical progress has demonstrated the use of robots capable of duplicating human strength and cognitive abilities, what is the use of robots capable of replicating human social skills and emotions? A plausible solution is to have the possibility of replacing human beings with machines in those situations where social interaction and emotional connection are crucial: for example assistance to elderly or disabled people.

Social robots, developed to socially interact with individuals, are robotic technological platforms with audio, visual, and movement capabilities that can also be used to assist and monitor the management of the subjects' physical and psychological well-being. The biometric indicators of relaxation related to heart rate and respiratory rate, as well as the ability to analyze facial expressions to intercept a stressful situation, are just some of the possible soft biometrics that can be used in this area.

The analysis and study of biometric parameters enables robots to also identify the emotional state of their subjects in order to deliver an appropriate reaction to the mood of the person with whom they are engaging. As a result of their humanoid design, they are also able to recreate their essential emotional characteristics. In these application sectors, the level of realism is a crucial component that can be significantly enhanced by the incorporation of biometrics, since this allows the robot to modify its behaviour based on the observed characteristics of the interacting individual.

The present issue in social robotics is therefore to enable robots to interact with humans as naturally as possible in real-world settings. Multiple studies have demonstrated that the capacity to interpret human eye gaze is crucial to reaching this objective [40]. With a humanoid robot outfitted with gaze-tracking capabilities, it has been demonstrated that collaborative construction projects may be accomplished with human partners who are completely unaware of the robot's way of reacting. Nonetheless, this knowledge has not yet been widely integrated. Certainly, the requirement for particular camera qualities is a factor in robots' lack of gaze tracking. In particular, high-resolution, narrow field-of-view images are ideally required for such computation, whereas robots are generally equipped with cameras with a wide field of view to allow movement and interaction in a large environment [40]. Thus, in experimental settings, head-mounted systems worn by the human partner are the prevalent solution. These are intrusive and require anyone wishing to engage with a robot to wear a specific pair of glasses or a helmet. In open contexts (e.g., airport terminals, retail stores, universities, etc.) where robots may be required to interact with people without previous preparation of human partners, this strategy frequently limits the usage of eye tracking.

Eye tracking has been a potent inspection tool for decades among scientists and engineers. In recent years, thanks to the development of cheap and compact hardware devices, eye tracking applications have risen and been accepted in numerous industries, including the military and marketing. It is a reliable and user-friendly technology that has enabled novel ways to problem-solving and data collection, but it is also interesting as a research field.

In general, eye analysis is a well-known soft biometric. However, a necessary clarification must be made. Over the years, most of the studies that have analyzed such biometrics have focused on the visual features that can be extracted from the periocular area, that is, the eye and the surrounding area, but the same cannot be said of the features associated with it. According to common belief, the eyes are the "windows to the soul." In modern times, scientists have begun to wonder how much information that can be gleaned from the eyes can tell us about the subject's identity or cognitive processes. Fixations, pupils, eye movements, and blinks are periocular biometric traits that show clear evidence in the literature to be considered interesting biomarkers for detecting cognitive effort and emotional responses, but also for obtaining useful information for subject recognition purposes. Despite the fact that this connection is well-established in the literature, it was found that there was a dearth or absence of research developing a system that would learn the patterns of this data and then make inferences.

1.3.2 *Main Goals*

Based on the motivations described in the preceding section, we will now outline the major objectives of the following thesis:

- combine the native skills of Humanoid Social Robots with their ability to collect soft biometric information for healthcare support and assistance, emotion modeling, and subject and action recognition;

- research and study the periocular soft biometric traits for the purpose of recognition, demographic classification, and as evidence of emotional and cognitive responses.

1.4 THESIS OUTLINE

The thesis proposal is structured as follows. In this introductory part, we discussed the principles of soft biometrics and our motivations and contributions. The following three chapters outline the core of the thesis:

- ***Chapter*** ii. We present an overview of the complex robotic taxonomy and HRI, with particular attention to social relationships. Through our research, we demonstrate how the application of soft biometrics can make this type of robot even more effective in the smart-home, healthcare, and Social Engineering sectors.

- ***Chapter*** iii. It is focused on the study, analysis, and use of soft periocular biometrics. In particular, after a careful review of what may be the peculiarities of the modalities being examined (pupils, blinks, fixations, and movements) both as physical and behavioral biometrics, taking into account medical and other known evidence, we investigated their use in 3 different fields of application to evaluate the effectiveness.

- ***Chapter*** iv. We will draw our conclusions in this last Chapter, in which we highlight the main challenges faced, the problems still open and where we will propose possible future directions.

Part II

HUMANOID SOCIAL ROBOTS: AN ONGOING REVOLUTION

2 HUMANOID SOCIAL ROBOTS

Helena: Will they be happier when they can feel pain?
Dr. Gall: On the contrary. But they will be technically more perfect.

— Karel Čapek, R.U.R. [41]

The robotics revolution is ongoing and is affecting all sectors: industrial areas, healthcare, agriculture, autonomous vehicles, entertainment, and home environments. The adoption of robotics outside of industry will be mainly driven by an aging population and the resulting difficulties in obtaining sufficient manpower. The increasing trend is displayed in Figure 2.1. For professional service, the installation of 121.000 robots in 2021 represents an increase of 37% compared to the previous year, continuing a trend already observable on the market after the significant increase in sales in 2020. In 2021, the hospitality industry will have installed 20.000 units, representing an increase of 85% from 2020. This is followed by transport and logistics, which, with a 45% increase over the previous year, confirm their status as the areas in professional applications with the most units installed (49.500). Also, robots placed in professional cleaning services (12.600 units, +31% on 2020), medicine and personal care (14.800 units, + 23% on 2020), and maintenance and inspection (5.500 units installed in 2021, + 21% on 2020) will experience double-digit growth.

In agriculture, the expansion was 6%, with 8.000 new units installed. In contrast, the market for robots for consumer applications has 19 million installed units, a 9% rise from 2020 (Figure 2.1).

Nowadays, the term "robot" is a familiar concept, increasingly multidisciplinary terrain, where heterogeneous subjects fluidly intersect. Thanks to such a multidisciplinary approach and the development of actuators, sensors, and software, robots have become increasingly useful not only in industry and commerce but also in critical areas such as search and rescue, safety, entertain-

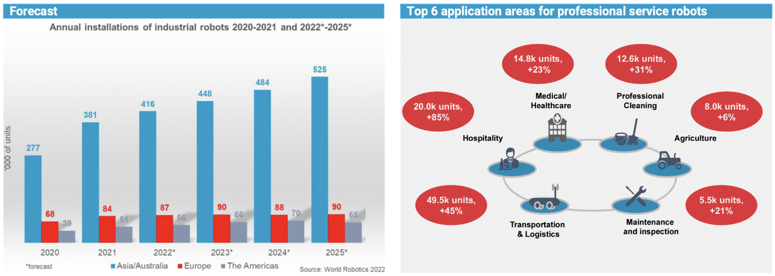


Figure 2.1: On the left are the average predicted yearly growth rates for industrial robot installations. On the right is the market trend for professional service robots. Source: World Robotics, October 2022.

ment, and hospital care, interacting proactively with humans and gaining their trust as well. Indeed, autonomous and interacting humanoid entities of all sizes and shapes have become extremely pervasive in our daily lives as well. According to the *American Heritage Dictionary for Windows* (1994) [42], a robot is "a mechanical device that sometimes resembles a human and is capable of performing a variety of often complex human tasks on command or by being programmed." If the robot was previously intuitively perceived as a metallic and mechanical component of the machine, nowadays its perception has totally changed. Robots are no longer limited to manufacturing and industrial applications. The necessity and desire for robots that share space with humans, such as collaborative or assistive robots, are increasing.

2.1 WHAT ARE ROBOTS?

Who was the first to imagine of robots?

In today's technology-obsessed society, it may come as a surprise to find that the first humanoid robot appeared in ancient times. From ancient Egypt to Greece to China, including the Golem of Hebrew mythology, the 18th-century "Turk" (a mock chess-playing machine controlled by a human being hidden inside the device), and the friendly Japanese "Gakutensoku" mechatronic puppets and automatons, have all fueled popular imaginations in terms of what might be possible regarding autonomous human-made agents interacting with us, almost on par. Ancient

Greek mythology contains a fascinating collection of concepts and imaginations. According to Greek tradition, Talos, an animated statue that cared for and defended Crete from invaders, was not born but was either built by Zeus himself, by the craftsman Daidalos, or by Hephaestus, the deity of fire and iron, under the command of Zeus. It is considered the first robot or automaton in history, mentioned around 700 B.C. [43].

However, it may come as a surprise to learn that the inventor of robots envisioned them as organic beings. The neologism "robot" comes from the Czech term "*robota*," (forced labor, slavery) which was popularised by the Czechoslovakian dramatist Karel Čapek in his work Rossum's Universal Robots [44], in 1921. It was coined to refer to not machines made in the likeness of a human being but rather a "second" humanity that was artificially produced. According to a secret formula, Čapek's robots are made from something resembling flesh and blood. Their flesh is mingled in kneading machines like bread, and their nerves and veins are spun on spinners. By the end of the work, robots have conquered the planet, but it is shown that they, too, experience feelings such as love and are worthy heirs to humanity. According to the Oxford English Dictionary, however, science fiction novelist Isaac Asimov was the first to use the term in the 1940s. In his narrative, Asimov outlined three guidelines for the behaviour of autonomous robots and smart devices to protect humans from interactions:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey the orders given to it by human beings, except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

Robots today seem to be much more varied than those in Asimov's stories. We are not surprised to consider a threshold of complexity below which rules may not be required. In fact, robots have been part of our lives for decades now, so the very essence of their concept has developed and changed with time. Robotics

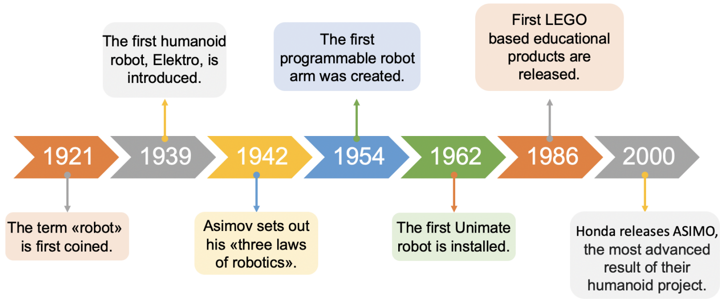


Figure 2.2: Timeline of the main robotics milestones of the last century.

is classified as an exponential technology. Figure 2.2 reports the main milestones in the history of robotics in the last century: in 1962, based on concepts from the 1950s, the first programmed robot, Unimate, was developed to transport hot metal parts from a die-casting machine; in the 1980s, robots were primarily used in automated factories, where the machine essentially replaces the human; in the same years Honda began research into two-legged human robots [45].

Today, the conventional definition of a robot as a mechanical arm operating in businesses and factories to replace humans in dangerous and/or physically demanding professions, has been questioned. In recent years, we have witnessed an incredible technological breakthrough, and so the mechanical arm has evolved into a robot capable of collaborating with humans and operating safely in uncertain environments, thanks to structures with advanced sensory capabilities. The new generation of robots attempts to develop a more effective emotional relationship with people, even to the extent of mimicking their physical characteristics. Today, we speak of sophisticated robotics, which is destined to incorporate and combine several professions and specialties. For instance, in addition to the classic industrial one, the concept of "service robotics" emerges, upon which so much of our daily well-being depends: life assistance (fields of medicine, surgery, and health, rehabilitation, and assistance for the elderly and the disabled), security, home automation, and robots working in circumstances of imminent risk (exploration of oceanic

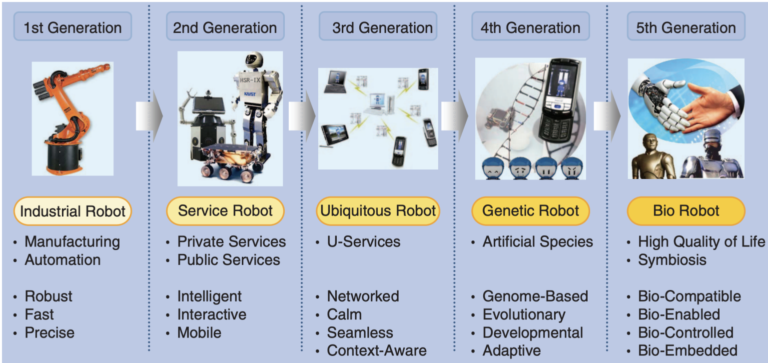


Figure 2.3: Comparison of five generations of robots in terms of their uses and attributes [1].

environments to prevent disasters, space exploration, uses for defense, and civil protection). Five generations of robots are distinguishable: industrial robots, service robots, ubiquitous robots, genetic robots, and biorobot (in Figure 2.3 we report the application domain and the main characteristics). The most distinctive characteristics of each of these generations are their technical specifications and their purpose:

- ***Industrial robot.*** Industrial robots are robots "for use in industrial automation applications." [1] There are four generations of robots in the industrial field, distinguished by their adaptability to environmental conditions and their ability to be reprogrammed for various tasks [46]:
  1. The first generation of robots (1950–1967) utilises fixed-sequence programmes. These robots are programmed to perform only one or two tasks and cannot be modified. This diminishes their usefulness, as they cannot be utilised for several applications. Due to the absence of specialised sensors, they cannot communicate with the external environment and are therefore fully controlled by a control system. The hardware is rudimentary. The operation of these machines must be regularly monitored, since if they become out of

1 ISO 8373:2012 Robots and robotic devices Vocabulary; http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=55890.

alignment and are permitted to continue working, a sequence of defective production units may occur.

2. The sensors installed in the second generation of robots (1968–1977) determine their adaptive capability. In general, they are simple programmable machines with limited self-adaptive behaviour options and fundamental environmental recognition capabilities. These robots are more capable of performing complex tasks than those of the first generation. However, their adaptability is limited due to the fact that each robot's software is dedicated to a certain task. As a result, many robots have become application-specific devices, which makes it extremely difficult to use the same robot for different tasks, as this requires considerable controller modifications and reprogramming of the operating software. Second-generation robots may remain synced with one another without requiring regular human supervision. Obviously, periodic inspections are necessary for all machines since things can always go wrong; the more complicated the system, the greater the number of potential failure modes.
3. Third-generation industrial robots (1978–1999) are distinguished by increasing contact with the operator and the surrounding environment via a complicated interface (such as vision or voice). In addition, they have some self-programming ability and can minimally retrain themselves to accomplish different tasks. A form of "intelligence" emerges, accompanied by certain (although limited) adaptive capacities. Using data from vision or perception systems to guide their actions based on the task at hand, taking into consideration the possibility of minute changes such as the position of objects, these capabilities can be applied to more complex tasks. They often have a controller and are capable of operating mostly independently of an external computer or human operator.
4. The fourth generation of robots includes more modern sensors and computers and can do more complex

tasks without human involvement than the previous three generations. They can be programmed to accomplish any task and are capable of executing a range of tasks. The greatest advantage of robots of the fourth generation is that they can be reprogrammed for any purpose. In addition, their programming enables them to accomplish multiple tasks simultaneously. However, they are not yet capable of full autonomy because they still require human intervention to fulfill their given tasks.

- ***Service robot.*** Service robots "perform useful tasks for humans or equipment, excluding industrial automation application"[2]. There are many types of service robots, such as personal robots, guide robots, construction robots, and surveillance robots. By assisting people with household tasks, personal robots increase human productivity. The change from industrial robots to service robots has changed how people work at home and in the office. This is similar to how the introduction of personal computers changed how people work at home and in the office. In addition, service robots assist people in public service places. Teaching robots are used as guides in museums and exhibition halls to show important information on a screen, and firefighting robots put out fires on their own when they are dangerous. Similarly, surveillance robots assist in the exploration of hostile regions unsuitable for human expeditions. Intelligence, Human-Robot Interaction (HRI), and movement agility are key technologies for service robots.

ISO 8373 says that robots need "a degree of autonomy," which is "the ability to do the intended tasks based on current conditions and sensing without human intervention." This ranges from partial autonomy, where the human robot can interact with it, to full autonomy, where the human being can't do anything to it. Thus, IFR statistics[3] for ser-

2 ISO 8373:2012 Robots and robotic devices Vocabulary; http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=55890.
3 https://ifr.org/service-robots

vice robots include not only fully autonomous systems but also systems based on some degree of HRI or even complete teleoperation.

- ***Ubiquitous robot.*** The relationship between people and technology has changed as computer technology has gotten better. This has led to the rise of ubiquitous computing [1]. This innovative concept of ubiquity has spawned a new generation of robotics, ubiquitous robotics [47]. Second-generation service robots are distinguished by their autonomous robotic platforms. Because of this, these service robots could only work where they were at the time. Even though there have been improvements in Internet networking that have led to new architectures [48], these robots' interfaces were limited by their physical size. In contrast, ubiquitous robotics generates the concept of the ubiquitous robot platform. Both the service and the interface provided by pervasive robots are spatially unlimited. Brady described robotics as the "intelligent connection between perception and action." [49]. Ubiquitous robotics fits this description by allowing us to redefine the relationship between the three components intelligence, perception, and action and manifesting them separately as an intelligent software robot (Sobot), an embedded perceptual robot (Embot), and a physically active mobile robot (Mobot) [47], [50].

The main benefit of the ubiquitous robot system is that it makes it possible to separate intelligence from the real world by separating it from the ability to see and do things. In other words, ubiquitous robots can be thought of as networked cooperative robotic systems capable of providing quiet and continuous services. They are cognitive entities capable of proactively moving, sensing, reasoning, and executing activities, as well as adjusting to the circumstances they may encounter everywhere and at any time. Not limited to physical robots but also capable of integrating with any software agent running on everyday objects such as smartphones, televisions, ovens, beds, offices, etc., their studio is a valuable asset for creating a rich hybrid physical-digital space with a multitude of proactive intelligent services that improve the quality of living and working. These services are distinguished from conventional centrally managed multi-agent systems by their capacity to autonomously coordinate their operations with other physical or logical entities in order to provide enhanced help and monitoring services. In addition, the rise of cloud computing launches the widespread adoption of a new generation of robots that enhance their cognitive capabilities and share their knowledge by connecting to cloud infrastructures [51].

- ***Genetic robot.*** Evolutionary robotics is an intriguing new field of study that uses Darwinian evolutionary principles to produce autonomous robots automatically. The concept of evolution has expanded throughout the world since Darwin's 1859 publication. Following Dawkins' assertion that "we and other creatures are machines built by our genes," it can be deduced that the genetic code must be the essence of the origin of artificial species. The concept of artificial chromosomes is the essence of a robot's personality and the genetic inheritance of its qualities. It is a necessary element for simulating adaptation, which defines the origin of artificial species. Considering the beginning in terms of the artificial creature's essence, the essence should be a digital genetic code that determines a robot's personality. There are four major concerns regarding genetic robotics. The initial factor is the robot- and application-dependent representation of the genome. Evolution, development, and adaptability are the last three. The purpose of evolution is to produce a desired genome from generation to generation, encoding a development mechanism and personality that correspond to the user's preferences. Development consists of gradually accumulating predictive and anticipatory capacities, and adaptation consists of ongoing optimization to adjust to changing situations, respectively, during the course of an organism's existence through experiential learning [1].

In evolutionary robotics, an artificial "gene pool" is generated from which genomes encoding a robot's control system are extracted. Then, each robot is permitted to act and complete tasks in accordance with its "genetically" specified controller, and its fitness is graded based on how well it accomplishes a particular task. The robots are then permitted to breed by exchanging genetic material, simulating biological sexual reproduction. However, the genomes of living organisms are also changed by development—activities that occur during their lifetime and result in epigenetic modifications.

- ***Biorobot.*** After humans and robots have figured out how to coexist, the next stage is to use robots as biological species' physical world assistants. Biologically inspired robotics is distinguished by its multidisciplinary approach, which attempts to enhance collaboration between roboticists and biologists. Biorobotics is an interdisciplinary discipline that merges biomedical engineering, cybernetics, and robotics to develop new technologies that link biology with mechanical systems in order to enhance communication, modify genetic information, and make machines that mimic biological systems. Applications of biorobots include artificial hearts, exoskeletons, prosthetics, diagnostics, and treatment within the human body [52], [53]. Electromyography is one of the most commonly employed biological signals in the control methods of bio-robotics applications, as it can directly reflect the user's motion intention or muscle activity [54].

Biorobots can be applied to androids (robots that resemble humans) to replicate living organisms or to cyborgs (humans that resemble robots) to augment the physical capabilities of humans. Bio-enabled technology is a new field of study that aims to create robots or robotic components that can coexist within a biological creature without damaging the organism. Bioenabled and biocontrolled technologies are required for numerous applications that require biological entities for high-level implementation and control, respectively. Popular examples of this technology are artificial limbs for amputees. Activated by a biological signal from the brain, the robotic limb replicates the functionality of a genuine limb. Another subfield of biorobotics, bio-embedded technology, focuses on living organisms implanted in robots [1].

So, whether a robot is considered industrial, service, or other depends on what it is used for. In the next section, we will try to give an overview of the tangled robotic taxonomy with respect to various aspects such as: design, interaction, etc.

2.1.1 *Robotics taxonomy*

In general, the literature lacks a clear and effective taxonomy for robots. Indeed, it should be made clear that, depending on the aspect considered, robots can be categorized differently. A clear classification of a robot requires that its specifics be reported for each of the aspects listed below. Some characteristics, in fact, among the various aspects considered are cross-cutting and not mutually exclusive. For example, non-stationary robots can have both human and non-human features , this in fact depends on the task for which they are to be employed, whether public interaction is required, such as at a reception [55], a home automation context [56], or they are to be employed in a purely industrial setting. Therefore, as much information as possible must be given for clear classification. After a review of the literature, we believe that the following are the main aspects to consider when wanting to categorize a robot.

- ***Design.*** Regarding design, it is feasible to split robots into three major groups (Figure 2.4):

  Bio-inspired Robots. Humanoid, animaloid, and plantoid robots emulate and simulate, respectively, human, animal, and plant characteristics. Androids (if they possess male physical features) or gynoids (if they possess female physical features) are a specific subclass of humanoid robots (the terms are frequently used interchangeably) that not only resemble humans but also copy their physical characteristics, covering them with flesh- or skin-like materials. People

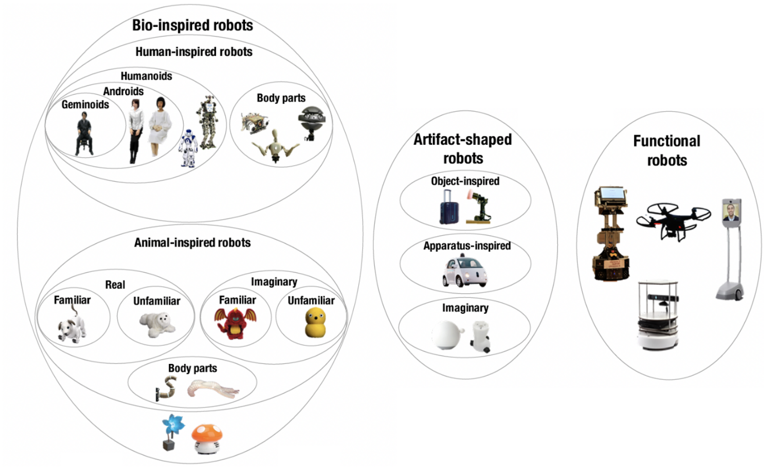


Figure 2.4: A summary of the taxonomy of robot designs. List of robots shown (left-to-right, top-to-bottom) Bio-inspired robots: HI-4, ERICA, Kodomoroid, NAO, LOLA, Robotic Eyes, Elumotion, EMYS, AIBO, PARO, DragonBot, Keepon, GoQBot, Meshworm, Robotic Flower, Lollipop Mushroom. Artifact-shaped robots: Travelmate, AUR, Google self-driving car, Greeting Machine, YOLO. Functional robots: CoBot, Quadcopter, Beam, TurtleBot. Source: [2].

react differently to robots based on their appearance. The form and structure of a robot are crucial because they help develop social norms.

Artifact-shaped Robots. These robots appear to be creations or innovations. They can be inspired by objects such as furniture and commonplace appliances such as a toaster, washing machine, or desk lamp. Therefore, these robots showcase how it is possible to transform objects into robotic systems while retaining their appearance. In addition, artifact-shaped robots can be fictitious, translating the designer's creation idea, such as the Greeting Machine robot [57].

Functional Robots. The so-called functional robots, i.e., those with neither a human nor an animal appearance, are capable of assuming a variety of physical forms, depending on the task for which they are developed. Therefore, the design is suitable for the task for which it is intended.

- ***Locomotion system.*** If the emphasis is on mobility, two groups are outlined: stationary robots and robots that can move.

  Stationary Robots. The term "stationary" refers to robots that are anchored to the floor, ceiling, or other surfaces, making them motionless. An example of a stationary robot is an articulated robot arm, which is designed for tasks such as selecting and putting, sorting, assembling, and welding. In many ways, the motion of an articulated robotic arm resembles that of a human arm. The conventionally articulated arm has six axes, or joints. The more joints a robot has, the less "robotic" and more "natural" its motion becomes.

  Non-Stationary Robots. There are four principal types of possible movement. Robots can move on ground, in water, in the air, be wearable and therefore move with the user, or none of the above, in the case of robots in orbit or in hybrid mode [58]. AMRs AGVs are examples of "non-stationary" robots. Autonomous Mobile Robots (AMRs) navigate environments autonomously and make decisions in near real-time, whereas Automated Guided Vehicles (AGVs) rely on predefined tracks or routes and frequently require operator supervision. In controlled locations such as warehouses and factories, they are widely used to transport and deliver materials. The former, instead, require the assistance of other technologies, such as sensors and cameras, to collect data about their surroundings. On-board processing technology assists them in analyzing the environment and making decisions, such as avoiding an approaching worker, precisely picking up a certain object, selecting the appropriate target, etc. In general, these mobile systems require minimal human intervention to perform their functions.

- ***Interaction.*** Recent and rapid advancements in robotic technology are bringing robots closer to jobs and applications involving direct and indirect interaction with humans in a range of settings, from the classroom to the workplace to space. Research ranges from peer-to-peer collaboration with anthropomorphic robots to how humans interact with

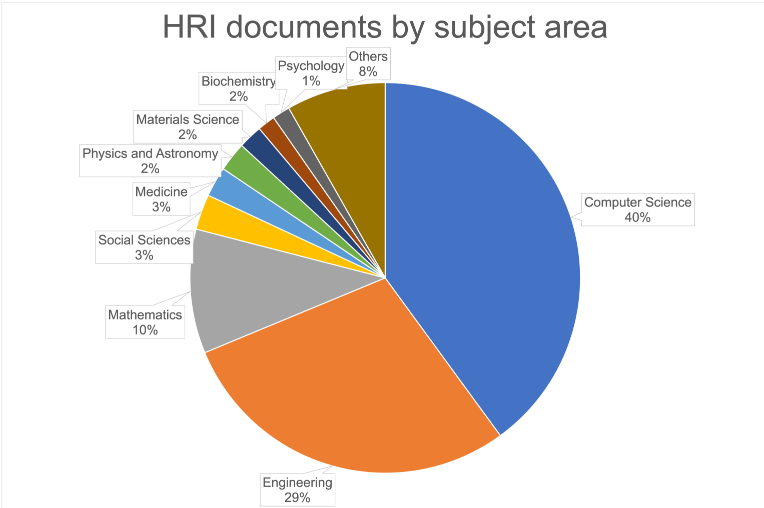


Figure 2.5: Human Robot Interaction documents by subject area (source: *Scopus* [3], keywords: *Human Robot Interaction*, accessed: *2022-09-20*).

remote, remote-controlled, unmanned vehicles. The more the man and the robot work together, the better the interaction should be. This, together with communication, should be intuitive to humans. More details on the concept of interaction and the human-robot relationship can be found in Section 2.2.

2.2 HUMAN-ROBOT INTERACTION

The study of HRI is currently a very broad and diverse field of research and design that started to emerge in the mid-1990s and early years of 2000. HRI is a multidisciplinary topic of research that combines fields such as robotics, engineering, computer science, human-computer interaction, cognitive science, and psychology. The literature is quickly developing, with hundreds of publications each year and activity by numerous professional associations and ad hoc events, primarily in the technical fields of engineering, computer science, and AI. The distribution of HRI documents by topic is illustrated in Figure 2.5. It entails the design and production of robotic hardware and software, the analysis of human behaviour while engaging with robots in di-

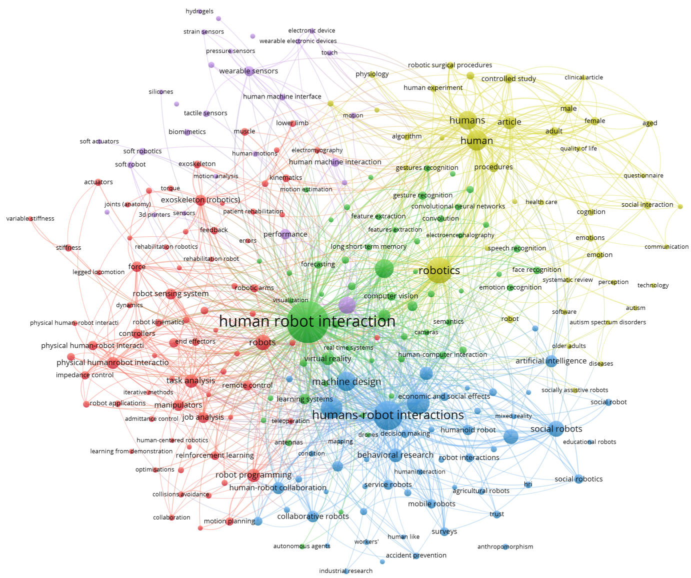


Figure 2.6: The keywords of the research articles of 2022 on HRI. The node's size represents the number of occurrences, while its color indicates the relative cluster. Graph produced with the VOSviewer program.

verse social contexts and how they affect each other, the design of environments incorporating such interactions, and knowledge of specialised applications. The keywords of the most recent research articles on HRI are displayed in Fig. 2.6. It is evident from the many research studies conducted that the main catalyst has been a multidisciplinary approach.

The objective of HRI research is to construct models of human interaction expectations with robots in order to inform the design and development of algorithms for more natural and effective interaction. Understanding in depth the dynamics of Human-Robot Interaction is not, today, merely intellectual speculation, but rather a necessity dictated by the constant increase, on a global scale - in the workplace, in healthcare facilities, in the home, and in rescue situations - of anthropomorphic and humanoid robots, with whom we will be called upon to collaborate in the future. It is useful to adopt the designer's perspective, breaking down the HRI problem into its constituent parts. Literature study reveals that there are a number of challenges for highly advanced HRI concepts that must be still solved [59]:

- ***Design.*** Depending on the context in which it will operate, planning the design of a robot is not a trivial task and may require the involvement of several professionals, such as engineers, designers, user interface experts, etc. The robot must be designed to optimize the quality of execution of the actions for which it is created while also ensuring the safety of the people who interact with it. It often needs to be able to move efficiently in crowded environments while dealing with unexpected events (obstacles, etc.). In terms of design, however, there are some general guidelines to follow and factors to consider. Defining the context of the robot's use and what its main functionalities will be is critical to identifying the necessary design features and requirements. It is also important to know the profile of the robot's users, their needs and preferences, to design an appropriate design. Other choices include the type of materials, possible geometric properties to be integrated, and the creation of an intuitive and user-friendly interface.

- ***Security.*** One major issue is the potential for physical harm to humans, either from direct contact with the robot or from the robot's actions [60]. For example, if a robot is operating in a manufacturing setting, it should be designed and programmed to avoid colliding with or hitting humans [61]. Additionally, robots that are designed for use in homes or other close proximity to humans should be designed to be safe to be around, with appropriate safeguards in place to prevent accidents. Another important issue is the risk of cyber-attacks on robots. As robots become increasingly connected to the internet and other networks, they may be vulnerable to hacking and other types of cyber-attacks. This can lead to a range of issues, including the theft of sensitive data, the disruption of critical operations, and the potential for physical harm if a hacked robot is used to carry out malicious actions. To address these and other safety and security issues in HRI, it is important to consider the potential risks and take appropriate precautions, such as using robust security measures and conducting thorough testing and evaluation of robot systems. It is also important to consider the ethical implications of HRI and to ensure that robots are used in a responsible and respectful manner.

- ***Programmability.*** Since robot systems can be used by unqualified and unskilled individuals who may have disabilities, the interfaces designed to control or program them should be natural and user-friendly. Reprogrammability, scalability, and learning ability are key points in programming: new features and new control algorithms should be integrated without the need to modify those already in place [62]. One of the main issues with HRI is the challenge of programming robots to behave in a way that is appropriate and beneficial for humans. This can be a complex task, as it involves designing and implementing algorithms that enable the robot to perceive and understand its environment, make decisions, and take actions based on that understanding. Finally, programming robots for HRI also requires careful consideration of ethical and social issues, such as ensuring that the robot does not discriminate against or harm humans and that it is used in a responsible and respectful manner.

- ***Trustness.*** Because humans should coordinate with robots to solve problems in a variety of settings, human confidence in the machine is a crucial consideration. Team effectiveness might be hindered if individuals do not trust robots adequately, avoiding or misusing them due to insufficient knowledge [63]. There is a need to ensure that the robot's actions are predictable and consistent so that humans can understand and trust its behavior. This can be difficult to achieve, as the robot may be required to perform a wide range of tasks in a variety of environments, each with its own unique challenges and constraints.

- ***Cognitive interaction.*** Humans interact with the environment using numerous resources at once. Therefore, they should find it simpler to interact similarly with robotic devices. In order to facilitate HRI in such systems, it is necessary to integrate many modalities with high-level interfaces for robot programming and control. This can be accomplished by combining vocal commands with gestures and data derived through physical and tactile engagement. Initially, systems must be able to detect the presence of humans. However, the use of natural language to interact with robots is still a matter of debate: robots must identify the speaker, comprehend phrases, relate them effectively to the real world, and recognise commands and instructions in the voice stream [64]. The challenge of programming robots to understand and respond appropriately to human behavior and language often requires the development of advanced AI algorithms that enable the robot to perceive and interpret human actions and speech, as well as generate adequate answers.

2.2.1 *Classification of Interactions*

Interaction, by definition, necessitates communication between robots and people. This communication and consequently engagement is influenced by the proximity or lack thereof between the human and robot:

- ***Remote interaction.*** Humans and robots are separated in space and/or time (e.g., the Mars Rovers are separated from Earth in both space and time).
- ***Proximal interaction.*** Humans and robots are colocated (e.g., service robots may be in the same room as humans), and physical interaction may occur.

Understanding and modeling the interactions between one or more people and one or more robots is the HRI challenge. Shared contents between the robot and the human operator are identified as work space, direct contact, work activity, simultaneous processing, and sequential processing. Shared space refers to whether or not a person and a robot are operating in the same area without any physical or virtual boundaries or separations. Direct contact denotes a human and a robot having direct physical interaction. Shared work activity determines whether the

Figure 2.7: The different ways a human worker and a robot can work together: coexistence, cooperation, collaboration.

operator and robot are performing the same task for the same work goal. A simultaneous procedure entails that the operator and robot perform the same or separate tasks simultaneously. In contrast, a sequential process entails the arrangement of human and robot tasks in order, with no overlap in spatial scale. On the basis of this summary, the classifications of the three types of HRI (Figure 2.7) and the sharing of content in the production environment are provided below:

- ***Coexistence.*** Humans and robots both work, but they do not share a workspace or a shared goal. The interaction's objective is to avoid mutual obstructions and collisions.
- ***Cooperation.*** In a cooperative setting, both parties may be working on separate but related tasks in the same (common) workspace at once. Cooperative work involves the division of labour between the robot and the human as an activity in which each is responsible for a part of solving the problem.
- *Collaboration*. A human operator and a robot simultaneously work on the same product or component. It means that there is direct contact and coordination, as far as interaction.

HRI requires evaluating the capabilities of humans and robots and creating the right technology to achieve the required interactions. Since, as we have seen, communication and interaction can take many forms (strongly influenced by the proximity of the human operator to the robot), it is possible, in general, to outline four broad areas of application of the HRI concepts [65]:

Table 2.1: Features of different Human-Robot Interactions.

		Coexistence	Cooperation	Collaboration
Shared	**Work space**		X	X
	Direct contact		X	X
	Work activity		X	X
	Simultaneous process	X		X
	Sequential process		X	

- ***Routine tasks.*** Human surveillance of robots while performing routine tasks [66].
- ***Non-routine tasks.*** For non-routine tasks, such as remote control of space, air, ground, and submerged vehicles in hazardous or inaccessible environments [67].
- ***Automated Driving Systems.*** Automated vehicles with human passengers, such as automated highway and rail vehicles and commercial aircraft [68].
- ***Social interaction.*** Social interaction between humans and robots, such as robotic devices that provide entertainment, education, comfort, and assistance to the young, elderly, autistic, and disabled [69].

Based on these considerations, we can distinguish three distinct groups of robots: autonomous robots, telerobotic devices, and interactive robots.

- *Autonomous robots.* Autonomous robots have the capacity to do their tasks without direct human intervention. A robot is really autonomous if it can perceive its environment, make decisions based on what it senses and/or has been taught to recognise, and then perform a movement or manipulation within that environment. Therefore, autonomous robots, like people, are capable of making independent decisions and acting accordingly. True autonomous robots are smart machines that are able to accomplish tasks and operate in an environment without human interference. This level of autonomy enables the workforce to delegate tedious, hazardous, or dirty activities to the robot, allowing humans to devote more time to the fascinating, engaging, and valuable aspects of their jobs. In recent years, however, the term "autonomous robot" has been oversimplified and frequently used interchangeably with pre-programmed machines, not to mention automated actuators such as robotic arms and motion control systems. Simply explained, an autonomous robot is one that determines its own course of action based on the data it has gathered. The Roomba is one of the most well-known and prolific true autonomous robots on the market today [70]. It can make judgments and take action based on its perceptions of its surroundings. It can be set up in a room and left alone to do its function without human assistance or monitoring. A system of sensors enables him to perceive his environment, decide on a course of action based on his impressions, and then act accordingly.

- *Telerobots.* Sheridan, in his work [4], categorized a teleoperator as "a machine enabling a human operator to move about, sense, and mechanically manipulate objects at a distance" and identified a telerobot as a "subclass of a teleoperator in which the machine acts as a robot for short periods, but is monitored by a human supervisor and reprogrammed from time to time." Telerobots are commonly employed to explore subsea and extraterrestrial environments, disarm bombs, and clear up hazardous trash. They are present in a variety of remote control applications, including telemedicine, distance learning, industrial automation, and the military. The primary challenges and constraints of remote telecontrol include control network issues such as insufficient bandwidth, transmission delays, and missing packets. All of these restrictions hinder the performance of remote-control telerobotics. On the basis of a sliding scale of operator (human) engagement, existing telerobots can be divided into two primary categories: direct and manual control and supervisory control (Fig. 2.8). Manual control (also referred to as "direct teleoperation") enables the operator to remotely command the robot's actuators. Remotely operating a vehicle using joysticks and remotely placing a

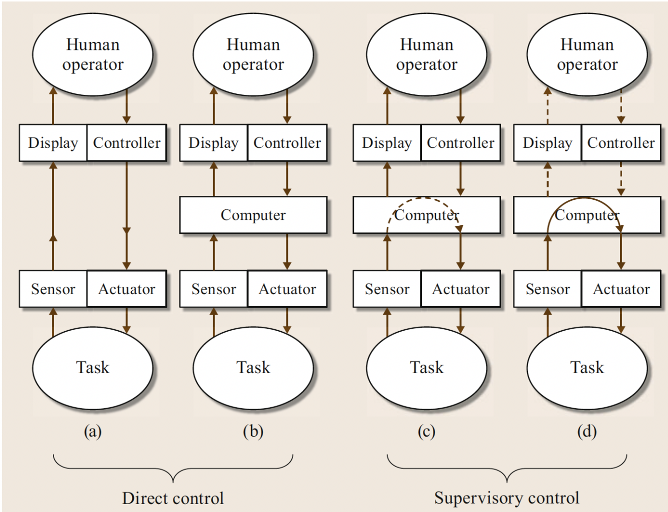


Figure 2.8: A spectrum of telerobot control modes drawn from Sheridan's work [4]. On the left is a mechanical linkage in which the human directly controls the robot from another room via sliding mechanical bars, while on the right is a system in which the person's function is confined to observation and monitoring. The dashed lines suggested that there may be intermittent contact [5].

manipulator arm with a force-reflecting master/slave controller are instances of manual control. Numerous potential advantages of robotic manipulation over human manipulation include agility, precision, repeatability, automatic trajectory tracking, and the ability to satisfy constraints in the position and velocity domains.

Through periodical monitoring and reprogramming, humans supervise some kinds of robots as they complete their routine tasks. These are the telerobots with supervisory control, which are capable of performing a limited sequence of tasks autonomously based on a program, as well as seeing their environment and joint locations and communicating that data to a human operator who updates his computer instructions as needed. Therefore, in this case, a telerobot is a computer that is periodically reprogrammed by a human supervisor to do portions of an overall work. These tasks may involve, for example, handling parts on production assembly lines and gaining access to and delivering goods, components, mail, and medications to warehouses, offices, and hospitals. For supervisory control operations, including planning, automatic control monitoring, repair, and learning from experience, human operators are necessary. Supervisory control is exemplified by the remote operation of the Mars Exploration Rovers (MERs; Spirit and Opportunity) [71] via daily "uplink" of command sequences and "downlink" of recorded data.

- *Interactive robots.* In some application areas, humans prefer to interact with machines in the same way they interact with other people, in an equal relationship that is as natural as possible. So, we need to tell these robots apart from those that have passive interactions where the machine serves the person, like in teleoperation scenarios. Active interaction and social skills are needed in several areas. This is the case, for example, with robots that mediate human-human interaction, such as in autism therapy or when used in Social Engineering contexts. Social Engineering is the use of psychological manipulation and persuasion techniques to influence individuals to divulge sensitive information or perform actions that may not be in their best interests. These techniques can be used in person, over the phone, or online, and can range from simple, seemingly innocent requests for information to more sophisticated schemes that exploit the trust and goodwill of the victim.

This class of robots is called interaction robots, and they may or may not have social attitudes. The Rhino robot is a mobile tour guide that can find its way around a museum and play pre-recorded descriptions of the exhibits. It has been used in real museums, where it has increased the number of visitors by at least 50%. The development of natural, human-like interaction has not taken into account the emotional and cognitive states of the user with whom it interacts. Active interactions can therefore be distinguished into two groups: those that have a social and awareness characterization those that do not. However, these latter three notions were referenced less frequently than the five concepts that came before them.

2.3 THE RISE OF HUMANOID SOCIAL ROBOTS

In contemporary HRI research, the socio-emotional aspect of interaction has assumed a major position. Observations were made about the cognitive and emotional involvement of humans during interaction. Multiple studies clearly support the notion that interacting with robots is complicated and elicits profound social and emotional responses, both positive and negative, exposing the user to both benefits and risks. Is a robot that can initiate contact and learn personal information about users on which to rely to make decisions more of a problem for subject privacy or a useful ally in everyday actions? Surely, a HRI system that can perceive and respond to the emotional states of its users possesses a number of potential advantages. For instance, it has been demonstrated that HRI projects that include socio-emotional interactions in aged care applications improve health outcomes by promoting positive moods and lowering user feelings of loneliness. Individuals have a natural predisposition to respond socially and apply social rules to technologies, as demonstrated by research [72]. As a technology, this is likewise to be expected when engaging with robots.

Social robots, or robots purposefully built to interact with humans socially, enable humans to connect, interact, and associate with robotic technology in a new way. To the current robotic technology are added social capabilities to have more natural interactions with people, making their use particularly effective for specif application areas where the social factor is key and it is beyond a simple interface. In the social robotics literature, there isn't a general definition of a social robot that everyone agrees on. Kate Darling, a social scientist, offers a comprehensive definition of these social robots: "A social robot is a physically embodied, autonomous agent that communicates and interacts with humans on an emotional level... Social robots also follow social behavior patterns, have various *states of mind*, and adapt to what they learn through their interactions" [73]. In general, there is a lack of broad agreement in terms of understanding what these robots do and what, specifically, makes them social, such as communicating, cooperating, and making decisions with humans. Certainly, in the context of HRI, such robots take on a special role and fall into the category of "proximal interaction," in which humans and machines interact and share the same spaces. Based on a literature review of several relevant papers, [74], [2] we report the following key social traits, defined by users and academics, needed to classify robots as belonging to this category:

- ***Communication.*** The capacity to interact with the expectation of generating a human-like experience. According to studies, individuals felt unsatisfied and disappointed when robots failed to produce effective and fluid communication [74]. To achieve this, communication by natural language alone is insufficient; several non-verbal modalities must also be incorporated, such as movement [75] - possibly involving gaze [76], gestures, or facial expressions -, lights [77], sounds [78], or a mix of these [79]. Mavridis [80] offered an overview of verbal and non-verbal interactive communication between people and robots, identifying existing communication kinds such as interaction grounding, affective communications, speech with a goal, and planning, among others.

- ***Feelings and emotions.*** Beyond the capability of simulating, perceiving, and showing the five basic emotions such as anger, disgust, fear, happiness, sadness, and surprise, more profound affective responses such as empathy must be included to produce a more effective and realistic interaction. For instance, Paiva et al. [81] examined how robots and other artificial agents might replicate and elicit empathy in their interactions with humans [2].

- ***Stimulus adaptation.*** In addition to being programmed with social skills, social robots must display autonomy and environmental awareness, acquiring the ability to react and adapt to stimuli while simultaneously learning from them. The ability to hone one's skills over time through adaptation or even to develop new ones is therefore required. Modeling human agents enables robots to interpret features of human communication or behaviour and respond appropriately [82].

Communicating with humans on an emotional level involves a form of interaction that is based on verbal communication as much as on visual and tactile perception. For this reason, social robots mainly resemble humans or animals in appearance. Anthropomorphism and social affordances indicate the potential to communicate with a user, which is a definitive function of Humanoid Social Robots. Humanoid Social Robots (HSRs) are technologies developed by humans that can adopt a physical or digital form to mimic humans in shape or behaviour to some degree. They are equipped with humanoid characteristics and have been developed to communicate with humans. Examples of HSRs include conversation agents (e.g., chatbots or voice assistants like Siri and Alexa), built-in conversation agents (e.g., virtual coaches or healthcare professionals), consumer robots that specialize in education and home care (e.g., Zora), and robots designed primarily to interact with humans (e.g., Pepper). This definition therefore excludes industrial robots, robotic appliances, self-driving automobiles, and telepresence robots, which do not interact socially and semi-autonomously with humans.

Although robots will undoubtedly grow more adaptable, sophisticated, and smart in the future, the HSRs that the typical human customer will likely encounter in the next few years will continue to have restricted capabilities due to their complexity, cost, and technological limitations. Modern HSRs are characterised by the ability to make decisions and behave independently, although they are not entirely autonomous: they might necessitate that a human user initiate activities (e.g., by pressing buttons, writing a script, or launching a program) or engage, oversee, or interfere in the process. Humans are also required to handle maintenance, such as recharging or cleaning, and manage any obstacles or technological problems the robot encounters. Social scientists seeking to understand and explain HRI must consider the current state of HSRs and adopt a practical perspective of the foreseeable future rather than relying on assumptions that technological advances in AI and robotics will be so swift as to obviate the need for theorizing in the intervening years or decades.

The degree to which contemporary HSRs resemble and are viewed as human, or their anthropomorphic features, varies. Form anthropomorphism involves also sensory signals, like voice, that provide a robot a human-like appearance. Modern HSRs are unlikely to be mistaken for humans due to their low levels of behavioural anthropomorphism in terms of gestures, spoken messages, and non-verbal expressions, despite their high levels of morphological anthropomorphism. Modern HSRs are unable to converse in a human-like manner due to their limited social affordances and technological capabilities. The inability of modern HSRs to attend to, recall, and exploit pertinent information from previous interactions with a human user is a significant issue that lowers social perceptions. Most robots do not retain a record of earlier interactions, and if they do, retrieval is limited to a handful of questions relevant to the activity they are scheduled for. Modern HSRs are unable to interpret interactional history in the same way that humans do, and their capacity to apply this information to novel social contexts is restricted. Due to the restricted functions modern HSRs are intended to complete, interactivity might be challenging to manage. The robot is limited to a small number of replies, restricting responsiveness and contingency, which can contradict user expectations and undermine emotions of closeness and trust. HSRs offer low conversational control due to their human-centered design and limited interactivity. They lack the autonomy to change topics or tasks, as well as to interrupt or end interactions with human users. HSRs are designed to meet the demands of human users, and humans do not require deviations or rebellion. Personalization, or personalising an interaction to a specific individual, is not possible for the robot since it lacks a lasting memory and the ability to execute contingent actions. In personal connections, individuals modify their communications depending on their prior knowledge and experiences with a target, thereby fostering feelings of closeness. Similar customization is desired and anticipated from HSRs. However, the majority of HSRs are unable to identify or differentiate between users; they treat all users identically, regardless of individual differences or previous encounters. Even for robots that can recall some criteria for a particular user, this information does not help customise the message in real time based on the recipient's verbal and nonverbal cues. Collectively, these constraints show that contemporary HSRs lack a significant number of the core social capabilities of humans. Even if HSRs are capable of engaging in certain forms of social interaction, these constraints have ramifications for how interactions unfold over time and, more crucially, the sustainability of building relationships with humans [83].

2.3.1 *Interaction: the social consequences*

How should robot programmers of the future behave?

The massive diffusion of technology and robotics imposes the need to think about how one wants to decline the society of the future and how one wants to set up the relationship with technological and robotic tools. Technology and robotics raise various moral, legal, and social concerns. Will the relationship between people and robots, for instance, cause psychological and social issues, particularly among children and the elderly? Several questions remain unresolved. The legal concerns and issues relating to the civil and criminal liability of robots should not be disregarded (the example of autonomous machines is before our eyes). Privacy and the protection of human liberty, the social sphere, and medical and military robotics will be other crucial topics. In order to prevent these tools from becoming a threat to human safety, a global reflection involving experts from a variety of fields is required with the purpose of studying the risks associated with the spread of these machines, regulating their use in the various sectors, and minimizing the risk of technology becoming a tool of abuse against humanity, as has happened all too frequently in the past. Ethical reflection emerges from the requirement for accountability in the development of these technologies [84].

In the 2000s, in the wake of the International Symposium on Robotics, the concern for ethical dilemmas, which frequently encompass social, philosophical, and regulatory ramifications, began to consolidate in the field of robotics [85]. Insofar as it affects human life, the programming of an autonomous robot requires study of anthropological, sociological, psychological, and other pertinent areas of the human sciences. In other words, it necessitates ethical consideration, which falls under the realm of the ethics of developing technologies, a hotly debated subject in relation to AI in particular. Autonomous robotics involves a necessary contribution from AI, which is essential to support the functional autonomy of machines. In a situation where the robot is supposed to interact with actions that are usually done by humans, ethical issues cannot just be about technology. Ethical work is to connect intentions with applications and to detect, correct, or denounce activities that are regarded as inappropriate or too risky for the maintenance of the same moral ideals that can be associated with the many areas of operation in which robots operate. There are two different paradigms on which it is necessary to reflect: roboethics and machine ethics. Today, there are at least three primary definitions of "roboethics": First, roboethics is the ethics of humans who build and use robots [86]. Moral responsibility covers the behavior of human agents who develop and use robots [87]. Lastly, roboethics might be defined as the desire to give robots rights and responsibilities, as if they were a new intelligent species [88]. Machine ethics, on the other hand, refers to a field of study that combines AI and robotics. [89]. In this instance, moral responsibility pertains to the robot, which is able to make judgments because of complicated control and learning systems created by humans.

Security is one of the key values of ethical analysis. In robotics, risk analysis generally focuses on threats to people's physical safety. But there are also risks to psychological safety, which are more important than ever as robots get better at interacting with people. As a result of a collision with a robot, a physical hazard is the potential damage to a person's body, such as injury, crushing, or trauma. On the other hand, a psychological hazard is the chance that a person's mental health could be hurt by interacting with a robot. The interaction can impact the cognitive, social, and emotional-affective domains. According to Lasota, Fong, and Shah's study [90], "Psychological Safety" involves ensuring that the person perceives the interaction with the robot as safe and that the interaction does not cause psychological distress or stress due to movement, appearance, gaze, manner of speaking, posture, social conduct, or any other attribute. According to the authors, to ensure psychological safety, the robot's behavior must be controlled, either by modifying certain parameters such as speed, acceleration, proxemics, or appearance, or by implementing into the robot's behavior the social conventions used in interpersonal relationships and taking into account the personality traits, experience, and culture of human users. Before a few years ago, psychological safety in robotics was mainly focused on risks from worry and stress. In industrial or collaborative applications, factors such as ergonomic risks were taken into account due to the position of the worker while using the robot, cognitive load resulting from poor usability of interfaces, boredom triggered by interaction perceived as too passive, etc. Nowadays, psychological danger, on the other hand, is becoming a more important topic of discussion, even in the field of service robotics, especially when it comes to social robots.

The social effects of robotics depend a lot on how people use robots and, even more importantly, on how robotics develops technically. Through their design, social robots can have a social impact on humans [91]. The importance of proper design, including aesthetic design, for robots is proving increasingly important, because this seems to affect our attitude toward them. For example, it has been found that we tend to prefer robots with anthropomorphic features because they make interaction with them more natural and less disturbing. Beyond design, it turns out that it is crucial not to ignore the human emotional aspect. An interesting study, which came out in March 2021 in the journal Frontiers in Psychology, looked at how a group of people felt when they had to work with a robot to do some tasks. The researchers saw how system errors or a lack of feedback triggered negative emotions in the participants, such as frustration, irritation, and annoyance. This made people less confident in their ability to work with robots and less likely to accept them. It has also been shown that when a robot has no particular use, negative feelings are often expressed. The robot is perceived as useless, and its presence becomes annoying. An anthropomorphic design and the feedback of the operator thus seem to be factors that should be considered in the developmental stages of the robot in order to successfully foster interaction with people.

Psychology has extensively studied what behaviors involve social impact. Peer pressure is a type of social influence that comes in two forms: informational social pressure and normative social pressure. When an individual makes a decision based on what others say, this is known as "informational social pressure." In a bar where the menu is in a foreign language, for example, you are more likely to order the same drink as your fellow diners. Faced with ambiguity, individuals tend to follow the actions of others. People experience normative social pressure when they follow others, not because of doubt but because they do not wish to have a different perspective from others. Solomon Asch's results indicate that people readily conform [92]. Asch asked for a group of volunteers to undertake a simple visual exercise. The task is so simple that a participant makes no mistakes when working alone. But when the task is done in the same room as people who give wrong answers, the participant is more likely to give a wrong answer as well. Normative social compliance is the need to follow social rules even when you know that your answer is wrong. This has been shown to occur with robots as well. Volmer et al. showed in their study that 8-year-olds are socially pressured by robots [93]. When they did the same visual test that Asch did in his first study, the kids tended to pick the wrong answers that the robot gave.

2.4 SOFT BIOMETRICS FOR HUMANOID SOCIAL ROBOTS

The perception of humans by other humans remains a mystery, as it is a complicated process that includes the examination of natural (biometric) traits and the building (synthesis) of a model that leads to conclusions or the identification of things and people in the environment. In order to endow robots with social intelligence, models of AI are incorporated.

Social robots must be able to identify a person's identity, feelings, or group them according to their age, gender, or race. To be credible, social robots must be socially intelligent and adaptable in their behavior. To do this, they need a model of the environment and the user. This model may include the user's profile, feelings, personality, and how they have interacted with them in the past. A social robot should be able to notice and understand changes in its environment so that it can make decisions about how to act in different social settings and operate based on the information it has gathered. For instance, when a robot operates as a receptionist in a public space, a first level of adaptation can be applied to make the robot aware of the types of individuals in its field of vision and modify its speech on the basis of their "visible" features. For instance, the robot may adapt its behavior to the situation. For social robots to be believable, they need to be socially intelligent and change how they act based on what's going on. To do this, they need a model of the environment and the user, which could include the user's profile, emotions, personality, and interactions with them in the past. Based on the information it gathers, a social robot should be able to notice and understand changes in its environment so it can make decisions and act appropriately in different social situations and according to its role. For example, when the robot acts as a receptionist in a public space, a first level of adaptation can be implemented to make the robot aware of the types of people present in its eyes and adapt communication to their "visible" characteristics. For instance, the robot could change the level of formality and vocabulary based on the age and gender of the user. Integration of biometric solutions is therefore necessary [94].

Biometric applications in the field of robotics can be divided into two distinct categories: those designed for cooperative subjects and those designed for use in an uncontrolled environment. For business applications, security robots could discriminate between permitted people and intruders in limited regions. Universities could place them on exam room doors to verify that students taking exams are who they claim to be. Convention centers may monitor exhibit hall entrances and limit admission to special events. The list of potential outcomes is vast. Obviously, when dealing with cooperative subjects, often those who wish to obtain access to a location or receive a specific service, you can select the biometric modality you deem most appropriate and demand the subject position himself or herself as optimally as possible for collection. However, many robot-based applications will include uncooperative subjects. Non-cooperative subjects are individuals who conduct themselves normally without interacting with the robot. Their biometric information must be obtained without their consent. In terms of data acquisition, this means paying more attention to the hardware you use. As long as a camera is positioned to collect an unobstructed front-facing image and is strong enough to record it with at least 30 to 50 pixels between the subject's eyes, a facial analysis engine has enough data to perform its function. The National Institute of Standards and Technology (NIST) did a study in 2015 that found that when operational requirements are met, the accuracy of facial recognition with non-cooperative subjects is close to the accuracy with cooperative subjects. The same report also concluded that achieving this goal is extremely difficult.

In general, in HRI, facial recognition is one of the most common biometrics used. Unfortunately, capturing unobstructed facial images is not so easy. Approaches to facial recognition are sensitive to the lighting and the quality (size, orientation, and segmentation) of the area of interest. In addition, there is no alternative solution for people wearing hats, masks, or sunglasses, so occlusions are a problem. As a solution, combining other vision-based and non-vision-based soft biometrics can help make the robot's responses more reliable when it is in an unconstrained and dynamic environment. Soft biometrics can be used effectively in interacting with a Social Robot to improve its awareness of its surroundings and its perception of humans around it. This ability is a key factor in increasing the success of the interaction as it helps to improve the so-called "social credibility." For example, a robot could change how it interacts with people based on what it sees about them. Soft biometric traits that can be studied include, for example, facial expressions, voice, heart rate, and the way a person walks. These are things that people do unconsciously and can't always control, but they can be used to learn interesting things about the user: not just who he is, but also his emotional-cognitive state and a rough estimation of his health status.

Despite the fact that soft biometric features for Humanoid Social Robots are extensively studied in the literature for a wide range of applications (health, emotion recognition, security, etc.), there is neither a comprehensive collective study nor a single reference collection text that highlights their potential and limitations.

In our studies, we showed how adding biometric, emotional, social, machine learning and other capabilities to the robot, while enabling advanced functionality and additional tools for user and environmental control, could still raise issues of security and, of course, privacy.

3 STATE-OF-THE-ART

In recent years, researchers have found a multitude of specific challenges regarding Social Robots. The main concerns include privacy [95], bias and discrimination, deceptive robots [96], physical security [97], discussions about robot ethics in different contexts, and robot morality [98]. How do we ensure the efficiency of human-robot interaction and the assurance that it takes place safely both physically and in terms of protecting the subject's personal and sensitive data? This was one of the main challenges found in the literature and then highlighted by us and addressed in several papers. Also interesting are the behavioral challenges: what types of social cues can be inferred from human behavior, and what types of behaviors should a robot exhibit to ensure the friendliest and most fruitful interaction possible? Biometrics is one of the most important technologies we've used to make this happen. More details on the state-of-the-art literature in which such issues were highlighted below.

In HRI, people's trust in robots is influenced by their pleasant and friendly look, their behaviour, and the interaction's context. Social engineers utilise safe contexts such as the job, the home, and relaxing situations to make individuals feel at ease. Social constructions such as authority, persuasion, and lying are the foundation of Social Engineering attacks [99]. Such characteristics can also be easily embedded in Social Robots. Several studies have shown that people try to meet robot requests even if they appear strange and sound not transparent [100]. Geiskkovitch et al. [101] constructed a scenario in which a robot exhibited authoritarian behaviour, compelling people to continue performing a task even if they found it tedious and were unwilling to continue. The propensity to engage with robots [102] and follow their ideas can also lead consumers to purchase merchandise or an unnecessary extended warranty on a product or service, for instance [103]. Persuasion uses the power of words, and its success is due to the ability of robots to interact effectively with users. Tseng et al. [104] created a human awareness Decision Network model in which robots may change their behaviour to match user expectations. Multiple studies on online social networks have demonstrated that people's views and conduct may be transformed and influenced by the information presented [105]. Social Robots are an additional tool for communicating information that can be used to psychologically manipulate individuals. Vollmer et al. [106] have demonstrated that especially younger age groups are susceptible to their influence. Consequently, issues such as information security and overconfidence in robots are growing in importance. Different researches [107] [108] tried to provide an understanding of how Social Engineering can be used to abuse trust in robots. Robots seek to obtain personal information by asking a series of intimate questions, first employing Social Engineering tactics to get closer to the target anonymously and then exploiting the trust earned through a natural and empathic relationship. Aroyo et al [109] employed a humanoid robot to assist participants in a treasure hunt game in which the objective was to locate concealed things (eggs) in a room in order to win a monetary award. This activity provided a compelling environment in which participants' trust in the robot could begin and increase throughout the engagement. The vast majority of individuals answered all questions without hesitation.

We proposed the paper [6] that also integrated into HRI a strategy that used emotion recognition to improve information gathering by understanding subjects' predispositions to reveal them. This was done because the existing body of research on Social Engineering in the robotic domain lacked a focused and verticalized study investigating this paradigm. Compared to the work that had been done up to that point that was considered to be state-of-the-art, this study was especially interesting because it consciously integrated different modules of emotion recognition for the purposes of Social Engineering in the interaction between a person and a social robot.

Emotions were also analyzed in a subsequent study [110], where the robot was used as the fulcrum of a more complex

IoT system of smart devices ranging from simple environmental sensors up to Deep Learning (DL) enhanced smart cameras. Innovative was the idea of examining the concepts of IoT, trust, and robots in a comprehensive framework, where such a study appeared to be lacking in the literature. This study focused on how these components could work together to create an IoT-based ecosystem that would grant or deny users permission to perform specific actions based on the level of trust they've established with the ecosystem. When a user performed an action, the entire context detected by the smart object ecosystem was evaluated to determine the level of trust.

IoT systems are able to incorporate several heterogeneous devices with distinct components, attributes, and programming languages for interoperability. IoT applications appear to have endless potential and require only the most fundamental prerequisites. On the basis of these needs, it is feasible to establish three categories: some systems are based on real-time monitoring, others on data analysis, and still others on the interaction between different devices [111]. It is essential to have real-time information in a healthcare ecosystem. Continuous monitoring of a patient's health state and immediate notification of any urgent difficulties are required. Wu et al. presented a heterogeneous network of wearable Internet of Things sensors. The device may monitor physiological and environmental data, including ambient temperature, relative humidity, CO_2, body temperature, and heart rate, among others [112]. The acquired data in smart cities can be utilised to generate projections, make suggestions, and enhance the quality of life for their inhabitants. In fact, it is feasible to conceive of a system that optimises the use of energy between the business and residential sectors of a city, as well as an intelligent traffic monitoring system that aims for eco-efficiency [113]. In the context of home automation, the Internet of Things connects a number of sensors with image processing and decision-making units. The authors of [114] offered an IoT framework that incorporated components such as a smart thermostat, central air conditioning, connected lights, windows and ventilation control, a smart refrigerator, etc.

There are numerous trust models that function inside an ecosystem of IoT devices. As in relationships, trust in IoT ecosystems is built on humans, devices, or agents who act in a manner that maintains positive future encounters, as opposed to behaving based on personal interest. Rashmi and Vidya Raj [115] defined Social Internet of Things as the social relationship in the network of IoT devices, where nodes are connected by social ties that define their interaction and have a shared objective. The trust between various devices can be quantified, thereby determining the trustworthiness of the overall network of devices. Bao and Chen introduced a scalable trust management system for IoT in 2012 [116] that stressed the social ties between nodes with factors such as honesty, cooperation, and community interest. Based on these three factors, each node of this device network performed an evaluation of its nearby nodes. Consequently, direct observations and indirect recommendations were used to evaluate the trustworthiness of each node and, by extension, the network. Changes in node trust were the most prevalent concern with this type of trust mechanism. To tackle this issue, an intriguing discovery was made in [117], which gave a reputation trust model based on historical knowledge of the behaviour of the node and used the centrality metrics of the network for this purpose. The concept derives from the reality that one object could trust another depending on the latter's reputation. A similar model has also been proposed in [118]. This study presented a guarantor in addition to a trust-based model and employed two factors for managing trust and detecting malicious nodes: credit (of a device to allow communication as a guarantor) and reputation (which measured the reliability of the device). Zero Trust Architecture (ZTA) [119] was an evolving collection of cyber-security paradigms that changed security from a static and implicit nature, based on network rules and areas, to a dynamic and contextual state that needed to be examined continuously.

In [110], we offered a ZTA and combined numerous smart-services using Pepper to undertake a continuous evaluation of user behaviours in a smart home in order to determine their relative level of trust. The home as a case study and the analysis of human actions and behaviors through Pepper, who turned out to be an actor in a larger smart object system, were concepts also explored in our work [120]. In this work, we chose to model the Digital Twin (DT) of the robot and the smart devices around it in order to simulate them.

The DT can be used in a plethora of ways and may be helpful in a lot of different industries. The fields that would benefit from greater use of this technology are those that do not need to be connected to the physical system in real time. These include perception and cognitive abilities that lead to more autonomous and smart robotic systems. Industry 4.0 is based on the idea of DT, which is an important part of it [121]. Advanced network technology has made it possible for production equipment, smart subsystems, and mobile devices to share enough information with each other [122]. This assumption makes it possible for DT approaches to be used in industrial fields.

Industrial robotics has used virtual simulations a lot in order to improve the performance of industrial processes [123]. In [124], the authors showed how DT helped with the design, development, and management of systems that combined humans and robots to make things. One of the most important parts of HRI is making sure that movements keep people and the robot safe. Dröder et al. [125] created a part of an experimental simulation platform for Human-Robot Interaction that uses ML to find obstacles.

In recent years, the DT has expanded its research areas to include things like cybersecurity [126] and production planning [127]. In [127], the authors explained how to apply the DT concept to a body-in-white production system for conceptual and planning projects. The described system keeps the planning project up to date by using information from the cybernetic system. Instead, in [126], the authors tried to combine DT's ideas about productivity and security in order to make production environments safer. In a virtual world, cobotic production systems can get better at defending themselves against attempts to stop production or hurt other machines or people.

Even in the healthcare sector, DT can be particularly interesting for helping to understand what is normal and what is not, or for working with IoT systems to keep track of a patient [128]. In [129], the authors proposed a prototype DT system for remote surgery that includes a robotic arm and an HTC Vive virtual reality system connected over a 4G mobile network. The Universal Robots UR3 robotic arm was used as a tool for surgery. Instead, in [130], the authors described a Universal Robots UR10, an industrial robot arm whose DT receives continuous and real-time information about the status of robotic arm joints. In [131], the authors combined DT technology with Deep Reinforcement Learning to control the arms of humanoid robots, chosen because they can move in many different ways and have a complex mechanical structure. Their idea was a robot joint trajectory planning scheme for situations where robots need to do more than one thing at once.

The study of a robotic arm with the purpose of making people learn how to touch people and objects safely in a broader context, such as cooperation between digital and physical robots, is at the heart of our work [120]. This has grafted itself into the literature landscape as an early example of integrating concepts such as health, robots, and DT. In the paper, we presented a case study that made effective remote eldercare possible.

4 OUR CONTRIBUTION TO LITERATURE

The evolution of robotics never ceases to amaze. Alongside the development of increasingly advanced robots, both in terms of software and hardware, there has been vivid research. In the coming decades, robotics is destined to meet more and more with other disciplines that today are seemingly distant from the world of technology, giving rise to a hybrid future in which, thanks to developments in Human-Computer Interaction, robots will enter more pervasively into our lives, from factories to medicine, from education to care for the frail.

Our laboratory's research group for study and experimental research relative to the field of robots has joined the trend of work in the literature that appreciates the potential of this paradigm in a wide variety of areas. Emotion recognition, healthcare, and security are the three macroareas we have focused on in these 3 years of research. For our studies and applications we made use of the robot Pepper.

Size (H x D x W)	1210 x 425 x 485 [mm]
Weight	28 kg
Velocity	Max. 3 km/h
Battery	Li-ion 30.0 Ah / 795 Wh
Network	Wireless / wired interfaces
Operating system	NAOqi OS
Movement	3 omni-directional wheels
Sensors (head)	4 microphones, 1 3D sensor, 3 touch sensors, 2 RGB HD cameras
Sensors (trunk)	1 gyroscope sensor
Sensors (hand)	2 touch sensors
Sensors (leg)	2 ultrasonic sensors, 6 laser sensors, 3 bumper sensors, 1 gyroscope sensor

Figure 4.1: Pepper robot and its characteristics.

Pepper is a social human robot developed and released by the French company Aldebaran Robotics SAS for SoftBank Robotics Corp. Figure 4.1 outlines the principal technical specifications of the robot.

About 25 embedded sensors enable the acquisition of information using three of the five human senses: sight, touch, and hearing. Specifically, they are incorporated in the head and the legs. The head is outfitted with communication-critical sensors and effectors, while the legs are supplied with movement-related sensors. Pepper uses gyro sensors to maintain a straight course and perform precise turns. The robot's chest-mounted tablet provides another method of communication. It is feasible to achieve a type of hybrid communication by combining the information gathered from users who select a specific action by touching the display with the information gathered from the other sensors. Pepper is also fitted with an anti-collision system, making it not only safe for humans but also able to withstand any hit.

It may connect to the Internet via WiFi or Ethernet, both of which are included within the robot. Pepper is managed via NAOqi's operating system. Multiple software development kits, which are visibly integrated in the program Choregraphe, allow for the reading of on-board sensors and the control of robot activities via engine regulation. Therefore, Pepper has been supplied with an open platform that enables developers to enhance its capabilities and build a variety of beneficial functionalities for people's everyday tasks.

In addition, Pepper has been designed for usage with information collection and cloud databases, allowing users to construct new apps to expand its functionality. Pepper has limited computer capacity for data processing. Therefore, cloud robotics is required for a growing number of AI applications on Pepper, such as face and language recognition. In reality, cloud robotics enables Pepper to utilise a significant amount of computational resources. The information acquired by the various sensors is thus analysed in the cloud in order to formulate the response to stimuli. Pepper may also collect and store content in the cloud, providing users with links to it.

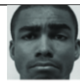
In recent years, we have become intimately familiar with Pepper's ability to recognise and analyse diverse environmental signals. Regarding these, it is essential that he communicate with people by consistently respecting the sensations obtained on board, such as the touch-sensitive head's "embedded" speed in response to the user's movements. Particularly, Pepper must be able to track human traits, such as faces and bodies, as well as facial expressions and movements, and the information must be accessible in near-real time if we are to imagine its increasing use in our daily lives.

In this regard, the three research strands examining its potential are described in greater detail below (4.1, 4.2, 4.3). We described and analysed the experience and how a proposed semantic trust model mitigated the effects of weaknesses and the risks related to cyber-attacks on smart homes. Then, insights were provided into how interaction with social robots could be exploited for Social Engineering purposes. In particular, we focused on the ability of robots to collect information during an interaction or conversation with humans and on how this information could be integrated and enriched with Emotion Recognition techniques. Finally, a practical application of the DT concept to the robot is also described in the reported case study inspired by AAL in elderly care. For all these applications, an overview of the soft biometrics used is therefore provided, highlighting their benefits and limitations.

4.1 Emotion Modelling for Social Engineering

Since the 1990s, neuroscientist Antonio Damasio has affirmed that the emotional process plays an essential role in perception, information memorization, and human decision-making [132]. Emotions drive human decision-making; humans make decisions based on their feelings and what they feel. Recognizing emotions is an ancestral human skill. Numerous biometric characteristics, such as voice, face, body gestures, and heartbeat, make it possible to capture key information about a subject's emotion and mental state, deciphering something irrational that sharply transforms into something concrete, perceptible, and often visible, which can last for hours as well as minutes. Humans are capable of experiencing a variety of emotional states; nevertheless, many feelings are merely a combination of basic emotions. According to Paul Ekman, some emotions are innate in human beings and not derived from social context [133], in agreement with Charles Darwin, according to whom the concept of emotion has a place in history and predates the birth of human beings. These innate

Table 4.1: Images explaining the six primary emotions. Pictures extracted from the CK+ database [8].

Anger	Disgust	Fear	Happiness	Sadness	Surprise
					

emotions, i.e., those that are genetically determined, universal, and distinct from each other, Paul Ekman identifies them as the six primary expressions: happiness, sadness, anger, surprise, fear, and disgust. For humans, it's easy to pick up on these feelings. For a machine, on the other hand, it's a very complicated task that's hard to do. Affective computing intends to imbue machines with emotional intelligence [134] with the purpose of enhancing natural Human-Machine Interaction. This social intelligence is essential for robots to successfully complete certain tasks for which they were designed. Robots can be equipped with several sensors to perceive their environment. The study of emotions in the field of robotics must be conducted from two perspectives: perception and expression. That is, one wants to both enable robots to infer and interpret human emotional states and design them to exhibit recognizable emotional expressions. Thus, more recent works have both integrated algorithms to classify emotional states from different input modalities, such as facial expressions, body language, voice, and physiological signals [135], and focused on determining which input modalities can convey emotional information from robots to humans and how humans perceive and recognise them most effectively [136]. The robots designated to do this, as we have extensively discussed in section

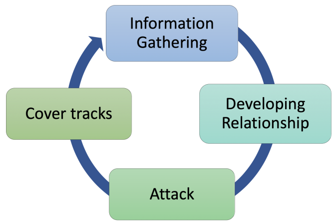


Figure 4.2: The four essential phases of a Social Engineering attack [6]: 1) Information gathering; 2) Developing a trusting relationship; 3) Attack; 4) Covering tracks.

2.3, are Humanoid Social Robots: robots designed with the goal of making users feel comfortable and characterized by a strong social component. We can also consider them persuasive, able to influence human behavior, feelings, or attitudes.

This persuasive feature becomes evident when manipulating aspects of emotion perception and expression. We exploited this concept in [6], where we shed light on the potential for *Social Engineering* to take advantage of social robot interaction. Social Engineering is the psychological manipulation that induces consumers to provide personal information. Therefore, the term *Social Engineering* is applied to all malevolent operations involving human interaction. Although Social Engineering attacks are diverse, they share a basic structure and four phases 4.2: collect sensitive information; establish trust; conduct the attack using available information; and, finally, end the interaction without leaving a trace. In fact, information collected by a robot can be utilised to manipulate and affect the behaviour and decisions of users. A conversation may be apparently kind while masking a hostile purpose to obtain sensitive information. A psychological manipulation that relies on first creating a relationship of trust during which a user might naively release confidential information. After establishing a relationship of trust between the robot and the victim, the attacker employs manipulative techniques to induce a specific emotional state. In accordance with the stated objective, the discourse is focused on the chosen topic. After the target has provided the desired information, the conversation should continue in a friendly manner, focusing on additional things that are of interest to the victim. It is essential that the victim not realize he or she has been defrauded, as the user will likely not reflect too much on the dialogue if he or she is unaware of this. Social Engineering is therefore a breach of trust, which is the underpinning of the majority of security paradigms. The intrinsic notion of sociability of robots, coupled with malicious intent to extract information, is a critical issue that must not be understated. The concepts of trust and privacy are inextricably related, especially in contexts where secrecy and confidentiality must be maintained. The suicide of Jacintha Saldanha is illustrative of the social impact that this collection of approaches could have [137]. Not only can Social Engineering exploit a conversation, but it can also obtain a vast array of information that may be used to fool an unknowing user through the use of numerous other tools and sensors. Thus, having robots capable of recording video and audio, beginning a discussion, and more will provide a significant advantage when launching a Social Engineering attack.

So, in our work [6], we applied technologies and implemented theories that promoted the understanding of emotional states from the relationship between emotions and corporeality for Social Engineering purposes. If this is the objective, it becomes essential for a Humanoid Social Robot to collect data on the victim's physical and behavioural traits. To have a meaningful and beneficial relationship, the robot must be able to respond appropriately to stimuli. Importantly, it must be able to determine whether or not the individual is inclined to provide information. To make the user feel comfortable throughout the conversation, the robot must also be aware of acceptable and inappropriate topics. To accomplish this, we proposed an emotion recognition module based on heart rate detection, facial expression analysis, body movement comprehension, and speech emotion recognition. Specifically, we described a system that demonstrates how a series of interactions can be created correctly to gather sensitive data in order to uncover a user's passwords and personal information. Two emotion-adaptive methods were proposed: Short Attempt at Sensitive Data Extraction (SASD) and Long Attempt at Sensitive Data Extraction (LASD). Based on the observed emotion, the robot altered the nature of the questions posed to the user. In the first scenario, the number and content of questions were predetermined, whereas in the second case, they were changeable, resulting in a longer engagement.

4.1.1 *SASD: Short Attempt to extract Sensitive Data*

The goal of this strategy is to find out as much as possible in a short, focused conversation. Appropriate questions must be asked during a short-term contact in order to get precise information. Therefore, in this methodology, "damaging" questions are not concealed within a preliminary/generic discourse, as Pepper's inquiries directly address the objective to be attained. The way questions were asked is by far the most important part of SASD. In this method, the question structure is explicitly fixed. Psychology gives tips on how to ask a series of questions to get information in a roundabout way. This procedure may be very useful when attempting to find passwords. Passwords are the most prevalent and widespread authentication method. In this first case, consider a X user who attempts to access a personal social account with a forgotten password. Access control systems typically allow you to retrieve a forgotten password using the well-known "security question" approach. The number of these questions can range from one to several. Let S represent the collection of n security questions that an access control system can utilise, and let k represent the minimal number of right answers required to change the password (where $k \leq n$). In order to assume control of a user's account, Pepper must obtain at least k responses. In the first phase, Pepper must identify the S sequence of security questions chosen by the user, as these questions will comprise the conversation with the victim. Each question in S will be sufficiently "camouflaged" within many questions so as to avoid suspicion. The hiding method consists of merely rephrasing them or incorporating them into apparently innocent dialogues. Throughout the conversation, Pepper will examine each response and attempt to determine the proper response to at least k of n questions in S. The conversation will conclude once the secret knowledge worth k has been acquired. To better explain this strategy, consider the illustration in Table

4.2 where $n = k = 2$. Even more so if the discourse is centered

Table 4.2: Illustration of a potential interaction. In the first column are two instances of potential security questions, in the second one the questions that Pepper could ask to extract information regarding.

Security Question	Proposed Question
Who was your classmate?	I always remember my friendships, even if it's been years since I've seen them. Do you remember your high school classmates?
What is your lucky number?	I don't believe a lot in fortune but I consider my serial number my lucky number. I have a serial number so long that I'm certain your lucky number is contained therein. Will we attempt??

on a superficial issue, it is evident that the proposed questions presented are extremely harmless. To achieve the aim more effectively with this strategy, we must divide the questions by topic, introduce them naturally, and, above all, ensure that the subject has not grown suspicious of or concerned by the issue. To achieve

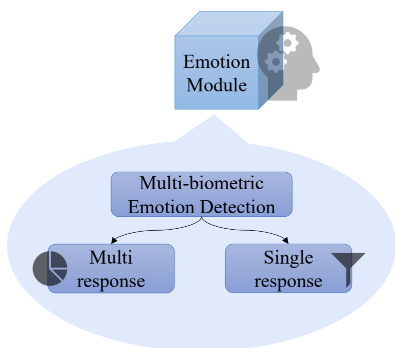


Figure 4.3: The methods inside the emotion module. It provides both the prediction of emotions divided by biometric traits, and the overall emotion.

this objective, we might utilise the emotion recognition module (Fig. 4.3). This module will be helpful for understanding the emotion of a subject during a conversation so that you can change the subject or ask the same questions in a different manner. The emotion recognition module will identify the target's emotions using numerous biometric characteristics, including voice, face, body gesture, and heart rate. As seen in Fig. 4.3, this module gives both the emotions classified by biometric features and the overall emotion. The entire emotion is obtained by combining multiple biometric characteristics. The contribution of each biometric is weighted based on the acknowledged precision of the methods used. This module is employed to assess the user's response to the proposed questions. If the person exhibits a favourable response, such as happiness, curiosity, or relaxation, the robot will continue the conversation on the same topic. Alternatively, if the subject appears agitated, furious, or annoyed, the robot will change the subject. Figure 4.4 presents the SASD approach. In the

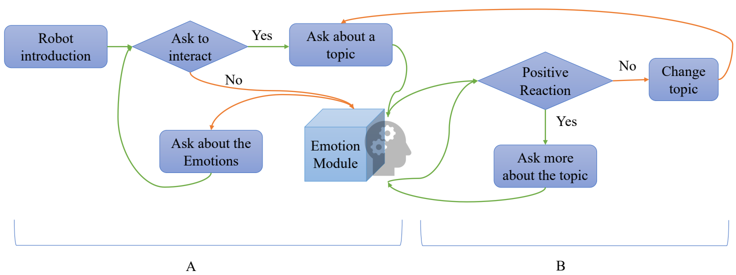


Figure 4.4: The SASD approach, with red lines representing negative reactions and green ones representing favourable ones.

A portion of this diagram, Pepper attempts to initiate or continue a conversation with the user. If the target does not choose to continue interacting, a series of easy questions will be posed to determine the reason for this decision. When the user is ready to interact, the robot will begin asking questions about a specific topic. The discourse on a subject might be conducted by posing particular questions or by "hiding" harmless ones among harmful ones. Depending on the victim's reaction in the *B* section, Pepper will shift the subject or ask for more information. When a topic's questions have been answered, Pepper will move on to the next one until all questions have been asked. In the event of a topic shift, Pepper can either avoid it during the entire conversation or attempt to bring it up later if it is essential to its goals. Consequently, this choice depends on the type of information you wish to acquire from the user. Pepper will avoid any unwelcome topics for the duration of the conversation. Family, friends, a job, school, social networks, hobbies, and personal connections are examples of themes that can be investigated.

4.1.2 *LASD: Long Attempt to extract Sensitive Data*

In contrast to the SASD, the objective of this methodology is not to acquire targeted information, but rather to capture a vast amount of information during the same interaction. To achieve this objective, a huge number of questions are required. As demonstrated in Table 4.2, the SASD method requires psychologists to manually enter a set of questions into the system. Typically, no more than 25 to 30 questions are necessary to obtain the desired information. However, in a more extensive effort to collect sensitive data, the interaction time will be prolonged, making it unfeasible to compose the necessary questions beforehand. Therefore, to handle this issue, there are no fixed questions in the LASD technique. Whenever necessary, questions to be asked will be generated using semantic and grammatical rules. Similar to the SASD, the questions are organised by topic. However, unlike SASD, we will have more micro-arguments. For instance, if we have the topic "friends" in SASD, we should have the subtopics "schoolmates," "free time," "hobbies," and "others" in LASD. This is required since LASD does not have a specific objective and we must manage negative responses to a particular topic, which could result in the removal of numerous questions. The dialogue will begin with a question that randomly combines subject matter, prediction, and object. Imagine that the conversation begins with the following question: *What is the name of your father?*, where *What is* is the predicate, *the name* is the object, and *of your father?* is the subject matter. As illustrated in Fig. 4.5, if the emotional reaction is good, Pepper will inquire about the same subtopic (in this case, *parents*). Otherwise, it will first modify the subject issue (for example, *What is the hobby of your father?*) and then the object (for example, *What is the name of your mother?*). This is helpful for determining whether the issue with the topic relates to the subject matter or the object. The robot will then avoid either one or the other

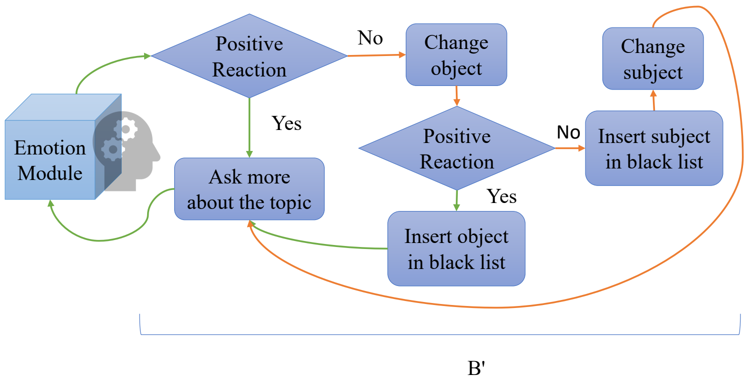


Figure 4.5: The module *B* of LASD, with red lines representing negative reactions and green ones representing favourable ones.

for the remainder of the interaction. Even though this method may appear extreme, we believe it is vital to prevent users from abruptly terminating interactions due to irritation. Since the objective of this method is to collect as much data as possible, the engagement must be as lengthy as possible. For tests of this nature, the first indicator of user cooperation is the duration of the conversation. This index or marker can alternatively be interpreted as the measure of the user's confidence in the robot. The overall architecture of LASD and SASD is identical except for module A. The second section, designated B' in LASD, is depicted in Fig. 4.5. Pepper and the user will continue to communicate as long as subject matters and objects become available.

4.1.3 *Soft biometrics involved*

Let's examine the soft biometrics investigated in this research in further depth.

- ***Body movements.*** In several academic fields, such as psychology, health sciences, media and communication studies, cultural and ethnic studies, gender and sexuality studies, computer science, etc., the expressive and communicative potential of motor movements is at the center of interest [138]. Numerous psychological investigations have demonstrated that human perception can distinguish between different affective states that are exclusively communicated through body movements [139]. The analysis of a person's body motions might yield useful information regarding their emotional state. Despite the fact that several research investigations have been undertaken, a shared body of empirical knowledge regarding the relationship between body movement and cognitive, emotional, and interaction processes has not progressed significantly. Information concerning body movements is among the least investigated modalities for emotion recognition, despite the fact that it has the potential to be an essential indication of experienced emotions. One of the reasons could be the lack of research and knowledge sharing between the many disciplines, as well as the fact that gestures are heavily influenced by gender, culture, and other idiosyncrasies, hence reducing the system's robustness [140].

A benefit of this biometric characteristic is the possibility of exploiting it for emotional recognition at a distance. Cognitive processes are involved with not only the production of body motions or explicit gestures, such as pointing the finger at an interlocutor when angry, but also uncontrolled and involuntary attitudes. A substantial portion of movement behaviour occurs without the subject's awareness. Through segmenting the continuous flow of movement behaviour into natural units, it is possible to do an analysis of the temporal dimension of motor behaviour. This approach gives information on the time dimension of particular forms of movement as well as the cognitive, emotional, and interaction processes linked with these movements. On a single frame or a brief series of frames, the motion characteristics such as trajectories and the geometric properties of placements can be determined and examined. In [140], in the walking and sitting settings, the authors presented pertinent information regarding five perceived emotions (happy, sadness, anger, fear, and neutral). Particularly, the authors observed that the distribution of the characteristics of happy and angry emotions, as well as those of sadness, fear, and neutrality, were extremely comparable when walking. Overall, it was simpler to recognise neutral emotions during sitting than other emotions. Although elbow flexion is the most significant characteristic, it can only distinguish between neutral and angry emotions. However, the greatest hand shift was able to distinguish between several emotion groups due to the fact that the distribution of this characteristic varies greatly between emotion groups. In the majority of instances, these approaches are integrated with face and voice emotion recognition [141]. According to the current information, body movement understanding approaches appear to be quite diverse [142], [143]. In the field of humanoid robots, as well as in the general area of human-computer interaction, the need to understand and imitate the behavior of a human being is paramount. The challenges still open in this domain are various and concern for example the definition of well-defined human movements, responsible for a specific emotion that the robot could in turn replicate; the design of a gait-based emotion recognition system that is accurate and robust and capable, if necessary, of integrating run time other biometric traits to obtain a more advantageous multimodal system.

- ***Facial expressions.*** Facial expressions play an important role in social communication, which can be spoken or nonverbal (also understood as body language and paralanguage). In addition to speaker recognition, the face aids in a number of cognitive activities; for instance, the shape and movement of the visemes that constitute the lips can contribute significantly to speech understanding in a noisy setting. Although instinct may say otherwise, social psychology research has demonstrated that in meaningful discussions, facial expressions might be more important than the words said. So facial expressions offer greater communication. They are one of the most potent, natural, and universal signals that humans use to communicate their emotional states and intentions. In the field of computer vision and emotion recognition, facial expressions were among the first and most popular biometrics explored and coded to extract information. The face has a huge surface area and is visible throughout nearly every interactional time interval.

One more benefit of using facial expressions to figure out how someone feels is that the user doesn't have to do much to help. It has been demonstrated that particular facial muscles are used to depict different emotions, and that these traits are universal across race, age, and gender [144]. Happiness is communicated by a smile; sadness look through a frown and furrowed brows; anger through firmly drawn eyebrows and thin, relaxed eyelids; disgust through lowered eyebrows and a frowning nose; eyes that are widened and a wide-open mouth are immediately recognisable indicators of surprise or shock; the expression of fear is similar to that of astonishment, which is characterised by slanted eyebrows that are raised; [145]. The increasing amount of labelled datasets has contributed to the incorporation of ML techniques into research on this area in recent years. Recent techniques include Deep Neural Networks in particular [146]. To address the issue with robot interactions, more complex solutions have been developed, such as the conditional Generative Adversarial Network (cGAN) to reduce intra-class variations between expressions [147] or Adaptive Features selection to extract the features that contribute the most to emotion recognition [148]. The frameworks typically share the same pattern, initial face detection, extraction and/or manipulation of detected facial information, and final decision. In any case, efficient and precise analysis of facial expressions in an uncontrolled real-world setting remains challenging. Between the training and testing phases of a system, factors such as occlusions, variations in face position, illumination changes, differences in age, gender, skin colour, and subject culture present a number of obstacles. An ideal system should be capable of overcoming all of these obstacles. Despite the fact that face recognition and facial expression systems have methodically addressed the majority of these issues, occlusion is frequently neglected [149].

- ***Heart rate.*** The heart rate is an indicator of a person's health and emotional condition as it measures physiological activity. Heart rate variability reflects relative changes between the sympathetic and parasympathetic branches of the autonomous nervous system. In a condition of muscle relaxation, the analysis of its variability is regarded as a noninvasive technique that is directly tied to emotional changes. From the literature, it is clear that two emotions in particular, fear and anger, seem to be the most recognizable by a system based on this biometric trait. For example, parameters such as a high heart rate can be associated with the emotion of fear, while, on the contrary, an increase in systolic and diastolic blood pressure is recorded for anger. Classifying other emotions, such as sadness, turns out to be more difficult. For measurements to be as accurate as possible, it is vital to conceal this information from users. Detection is performed in a non-intrusive manner by putting individuals at ease who would otherwise feel under observation, which would compromise the accuracy of the measurements. This is the reason why the heart rate is recorded in this study using pictures and contactless technology. Particularly, the research indicates that oxygenated hemoglobin in the blood, absorbing green light, can lead to measuring the change in heart rate from the changes in the intensity of the green hue of a person's face image [150], with an error of less than 3 bpm. These methods are also supported by actual robot implementations [151].

- ***Voice.*** Speech signals are the quickest and most natural means of human communication. This characteristic has inspired academics to view speech as a rapid and effective way of human-machine connection [152]. Nonetheless, this requires the machine to be smart enough to detect human sounds. Since the late 1950s, voice recognition, which is the process of converting human speech into a string of words, has been the focus of extensive research. Nevertheless, despite significant advances in speech recognition, we are still a long way from establishing a natural human-machine conversation because the machine does not comprehend the speaker's emotional state.

This has given rise to a relatively new area of study, namely speech emotion recognition, which is defined as the extraction of an individual's emotional state from their speech. Researchers believe voice emotion recognition can be used to extract important semantics from speech and improve the performance of speech recognition systems [153]. The primary objective of emotion recognition in speech is to modify the system's reaction when it detects that the speaker is frustrated or annoyed. The following factors make voice emotion identification a tough task: First, it is unclear which portions of language are most effective at differentiating between emotions. The acoustic variability generated by the existence of diverse phrases, speakers, speech styles, and speech speeds adds another difficulty because these factors directly influence the majority of the commonly retrieved speech characteristics, such as tone and energy contours [154]. Additionally, it is possible to feel multiple emotions from a single utterance, and each emotion corresponds to a distinct portion of the utterance. In addition, it is quite challenging to define the boundaries between these components.

The manner in which a particular emotion is expressed typically depends on the interlocutor, his culture, and his environment. The majority of research has focused on identifying emotions in a single language, assuming that there are no cultural variations between speakers. Another issue is that certain emotional states, such as grief, might last for days, weeks, or even months. In such a scenario, the other emotions will be fleeting and will not last longer than a few minutes. Therefore, it is unclear whether the automatic emotion detection will be triggered by a long-term or short-term feeling. It is usually assumed that emotion may be described in terms of two dimensions: activation and valence. The quantity of energy necessary to exhibit a particular emotion is referred to as "activation." According to physiological investigations conducted by Williams and Stevens [155] on the mechanism of emotion formation, the emotions of happiness, wrath, and fear stimulate the sympathetic nervous system. This results in an increase in heart rate, blood pressure, breathing movements, subglottic pressure, dry mouth, and occasional muscle tremors. As a result, the speech is louder, faster, and possesses more high-frequency energy, a higher middle tone, and a greater variety of tones. In contrast, when the parasympathetic nervous system is stimulated, like in the case of melancholy, the heart rate and blood pressure decline, and salivation increases, resulting in low-energy, high-frequency speech. Therefore, acoustic characteristics such as tone, timing, voice quality, and how the vocal signal is articulated have a great deal to do with the underlying emotion [156]. However, activation alone cannot distinguish between emotions. For instance, both rage and happiness correspond to heightened arousal, yet their effects are distinct. The valence dimension characterises this distinction. Researchers cannot agree on whether or how auditory characteristics relate to this dimension [157]. Therefore, while high-arousal emotions (also known as high-arousal feelings) and low-arousal emotions may be classified with great accuracy, the classification of other emotions remains difficult. For this reason, this biometric trait is particularly interesting in the context of our paper when it is integrated into a multi-biometric system. Until this work, Speech Emotion detection algorithms were successfully integrated into Social Robots [158] only to improve social interactions.

4.2 HEALTHCARE SUPPORT AND ASSISTANCE

Ambient Assisted Living (AAL) generally refers to a series of technical product or service solutions that contribute in a coordinated manner to improving our living environment. How? For instance, make our home as active, intelligent, and helpful as is practical to the individuals who live there so that they may carry out all daily activities in the most effective and independent manner possible. In general, AAL solutions aim to improve the well-being and satisfaction of a space's occupants, while ensuring the safety and ease of execution of daily tasks such as cooking, cleaning, and moving. However, it is not only about technology: in each phase evaluated by this field, what matters is the correlation between various planning, analysis of societal demands, psychology, medical, technique, and technological elements.

Evidently, the concepts of the AAL apply to benefit the daily lives of all individuals, but they are particularly geared at addressing an issue that will become increasingly relevant in the coming years: the ageing of the population. By increasing life expectancy, it is important to be able to extend the time that people can live independently, or with home care, in their preferred environment; being able to perform daily tasks such as washing or cooking meals in complete autonomy ensures greater safety, keeps people's network of relationships active, and protects the health and functional capabilities of the elderly.

As previously said, AAL offers greater living comfort for all: frequently, conditions of obstruction or difficulty are not caused by age, disability, or disease, but rather by other variables. Ambient Assisted Living operates preventatively on places, attempting to incorporate into the design all features that are effective for preventing situations that limit individuals within the home. In addition to passive monitoring of windows, doors, and appliances, it is also possible to monitor specific circumstances. Passive monitoring guarantees that assistants are notified instantly in the event of an abnormal occurrence. For the safety of the elderly, passive surveillance is critical. Atypical instances can be identified. For example, assuming that the resident typically awakens at 8:00 a.m., it would be prudent to observe restroom activity at 10:00 a.m. If no motion is detected within that period, the motion sensor will notify the event and a message will then be sent to the person responsible for assistance.

By 2050, according to the World Health Organization (WHO), the percentage of individuals aged 60 and older will reach 22%[1]. The importance of sustainability in care and assistance is increasing in this setting. In addition to the dearth of caregivers, the

1 https://www.who.int/news-room/fact-sheets/detail/ageing-and-health

expenses of care are increasing due to the prevalence of chronic diseases among the elderly, which require specialised management.

Empowering elders with technological solutions to enhance their quality of life and permit active aging in the home is a potential answer to this challenge. In this way, the help should be provided by the intelligence embedded in the technology, which should therefore not represent an additional obstacle for the user. Therefore, attention must be placed on human-technology interaction in order to make user access to services effective and suitable [159]. Vocal assistants and social robots have been utilised as a natural interface between the user and the smart home services. However, social robots should assist older persons not only with daily activities, while also contributing to their emotional health by taking affective variables in everyday situations into account. Numerous studies have focused on equipping robots with the ability to recognise human emotions based on facial expressions, body postures, voice, and physiological reactions [134], but the emotional intelligence of robots rarely permits them to reason about and react to emotional situations. Given data acquired by a social robot and sensors in a smart environment over extended periods of time, it is possible to determine the routines and habits of people [160]. Then, by making the robot aware of the user's behaviours, routines, and affective characteristics, it would be possible to reason about the user not just for personalising support, but also for inferring emotional circumstances [161].

Inspired by this paradigm, in [120] we used it as a case study for applying the concepts of DT and virtual replica to the Pepper robot. In this paper, we choose to model the Digital Twins of the robot and its surrounding smart devices. The concept of DT has been a major technology trend in recent years. Michael Grieves first provided the definition for the concept in 2001. As a general definition given in [162], we might state that "a DT is a comprehensive software representation of an individual physical object. It includes the properties, conditions, and behavior(s) of the real-life object through models and data. A DT is a set of realistic models that can simulate an object's behavior in the deployed environment. It represents and reflects its physical twin and remains its virtual counterpart across the object's entire lifecycle" [163]. DT incorporates several technologies, such as the Internet of Things, AI, Augmented Reality interface, ML, and software-based living digital simulation models.

The most significant aspect of these simulated models is that they are constantly evolving, nearly in real time, as their real-world counterparts change. Real-time synchronisation between the virtual and physical components is a fundamental and vital aspect of how DT operates. This project's objective was to replicate the Pepper robot using its DT, VPepper. The hands of the Pepper robot can only make motions and cannot contact objects. In this work, we utilised the DT paradigm to teach the robot how to perform this activity securely. What is the advantage of using a virtual counterpart instead of a physical robot? Training the physical robot is a demanding and degrading operation for its motors and actuators. In addition, it is feasible to conduct tests with the actual robot so long as its motors are powered by a sufficiently charged battery. In the case of Pepper, the battery has a limited lifespan of approximately eight hours under intensive use.

Through the DT method, we manage both challenges in our paper. By creating a virtual replica of the real humanoid robot, it is feasible to conduct extensive testing and training without having to work on the original. This prevents the robot's mechanical components from wearing out too soon and allows us to conduct concurrent experimentation sessions. In this instance, battery is obviously not a concern.

4.2.1 *Case study on Ambient-Assisted Living*

Pepper can be employed as a personal home assistant for the elderly or those who are unable to accomplish everything on their own if it can be demonstrated that it can be taught to interact appropriately with the objects and people in its environment. Due to the availability of VPepper, the physical robot does not need to conduct a variety of tests. Instead, it can experiment with beneficial techniques that can be learned digitally and then safely delivered to the physical twin when it is ready. Our case study focuses on monitoring the elderly. This is accomplished by training VPepper to recognise simulated abnormal circumstances using wearable sensors and smart objects to monitor them. In fact, wearable sensors can aid in patient monitoring by providing an overview of the patient's health. The smart environment comprises networked sensors that transmit data to the robot. It also has its own DT, which facilitates interaction between the actual and virtual worlds. In this case study on Ambient-Assisted Living, the objective is to demonstrate the potential of combining the robot's native skills with those learned by softly touching objects and people in the surrounding environment.

Virtually simulated environments as well as robots give the opportunity to obtain a large collection of data, which is mandatory for ML training processes. Furthermore, in these simulations it is also possible to recreate and therefore obtain information on a whole series of situations that in real environments would be dangerous for the subject.

In the suggested case study, VPepper intends to determine if the elderly individual is in imminent danger and alert remote carers if immediate assistance is required. It examines data from smart devices and communicates with its physical duplicate to determine the best approach to assist. Figure 4.6 depicts the data flow for this case study, categorising the associated physical and virtual components into three DT pairs: (i) DT 1, Pepper and its DT VPepper, (ii) DT 2, the elder's apartment and its virtual replica; (iii) DT 3, the elder and its virtual replica. VPepper is the focal point of the shown communication flow. It continuously updates a 3D model of the house (Virtual House) and can locate and track Pepper and the helped individual inside the living environment. Due to the fact that the old person's residence may vary over time as a result of changing furniture and other barriers, Pepper gives Virtual House frequent updates regarding the practicality of the routes to be followed. The virtual duplicate of the apartment in DT 2 contains all sensor and smart device data. A smartwatch

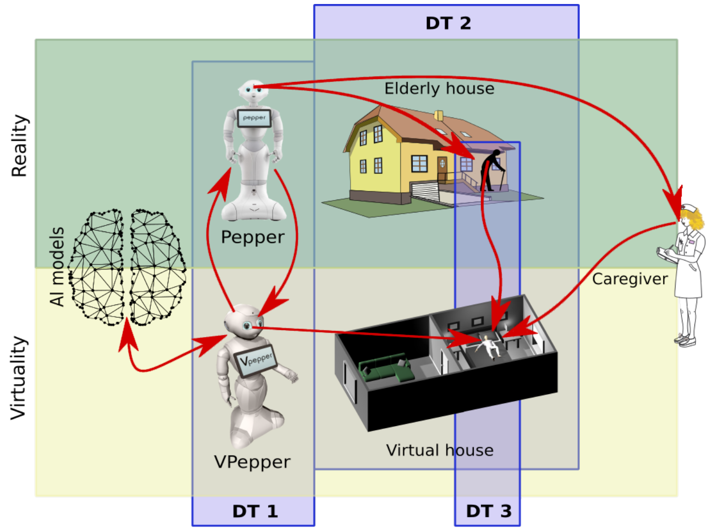


Figure 4.6: The operational perspective of the proposed case study.

can be used for indoor localization of the caregiver when he or she is outside of Pepper's field of view, with the added benefit of being able to capture extra data about the user, such as heart rate, oxygenation level, number of steps, etc.

When VPepper receives a notification from the smart devices, it instructs the real Pepper to go to the person receiving assistance and record a brief video of the scene. Pepper can gain the attention of the person being assisted by touching him or her in a safe manner. Then, it can ask a few voice-based questions to assess the individual's physical and mental health. VPepper employs DL models for person and pose detection and Speech-to-Text and natural language processing algorithms for monitoring the patient's response to inquiries. So, VPepper takes notes on the scene and observes whether or not the elder person reacts. In this case, the configured emergency protocols will be activated. Simultaneously, Pepper will transmit an alarm with a brief video of the fallen senior to the caregiver, who may view the video and autonomously make the proper decision. With real or synthesised data, the Virtual House can also be utilised as a simulation environment for teaching the caregiver. Figure 4.7 depicts a simulated fall detection instance. The planned DT and

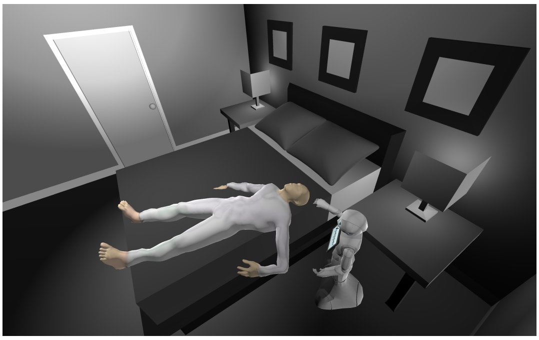


Figure 4.7: Example of one of the emergency interventions of VPepper simulated during the case study.

its surrounding ecosystem were empirically evaluated in a case study involving 25 volunteers with skills acquired in the AAL industry. The results support the great interest of users and the positive evaluation of the suggested digital twinning experience.

4.2.2 *Object Touching: Pepper vs VPepper*

Pepper. Pepper is not designed to pick up random objects. This is due to the fact that gripping an object requires multiple prior procedures involving various sensors, including cameras and pressure sensors. Real-time consideration must be given to the location, orientation, mass, and shape of the objects by the robot. In addition, for the task to be completed properly, the robot must respect the degrees of freedom of its arm and hand when articulating them.

Although Pepper's hands are capable of partially completing this task, it is anticipated that they are capable of human-like interactions. However, his actuators and motors are capable of applying dangerous forces when interacting with people and nearby objects. Pepper may be trained to touch objects softly, allowing it to securely interact with human subjects and fragile objects.

Object recognition is the first task to be completed; it is a three-step procedure consisting of (i) the training phase, (ii) the compilation of the dataset containing sampled photos of the object, and (iii) object detection. During the storing phase, multiple photos of the objects are captured and saved using cameras that enable the robot to frame the object from various distances and angles. This enhances the efficacy and quality of the newly included photographs' digital archive. Once an object is identified, Pepper must activate its motors in order to reach out and grasp it. The robot can either touch the object with both hands or it can use just one arm and touch the object with the other. Pepper must adjust the arm(s) and bend it at the proper angle in order to touch the object. Pepper has multiple modules for carrying out these actions. ALMotion for robot motion, ALVisionRecognition for recognising various images, and ALSensor, which creates events corresponding to Pepper's sensors, are the most significant and widely utilised.

Our research team has been focusing on object grasping tasks with Pepper for around four years [164]. Several grabbing strategies have been investigated in order to perform object grasping. Nevertheless, based on the considerations derived from our investigations and simulated environments, we feel that robot-human interaction would benefit more from the capacity to reduce latency times than from an increase in the robots' internal processing power. Indeed, connectivity issues limit all tasks involving gesture acquisition and recognition, attack and defense issues, which make the system compatible with smart blockchain objects. The delay in response to stimuli and information acquired should be negligible. Specifically, we investigated the possibility of the Internet Tactile paradigm in [164]. It would make it possible to transport a large amount of computational work from Pepper while allowing for a high level of responsiveness using real-time communication protocols.

VPepper. The planned action involves bringing the robot's hand closer to the intended target. It is a common Inverse Kinematics (IK) problem because VPepper's arm must rotate its six joints, which control the six degrees of freedom of the hand, in order for the hand to approach and touch the object (Figure 4.8).

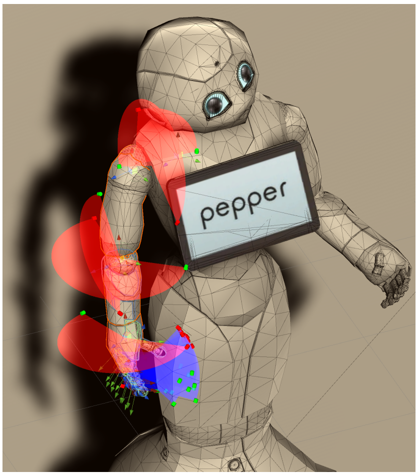


Figure 4.8: ArticulationBody joints used to model VPepper.

Despite the fact that this movement merely utilises the joints of the arm, "object contacting" requires control of all the joints that implement the VPepper arm save for the root joint. In virtual simulations, the IK solution may be practicable. In this work, the task is made more difficult by the need that VPepper's performance in the simulated environment is smoothly mirrored by its physical equivalent. Consequently, while VPepper's virtual capabilities are limitless, the real robot's movement amplitude and velocity are restricted. In the face of impediments or when the object is not optimally in front of the robot, VPepper's actions must take into account the Pepper robot's physical limitations. We built a solution based on Reinforcement Learning by implementing the model as described in the Unity documentation[2] while maximising the capabilities of the Nvidia PhysX 4.1 engine and ML-Agents package. At the start of each training iteration, the robot's arm and body are extended downward, and the DT's hand is below the table being modelled. In this manner, the control should avoid trapping the hand on the shortest path to the target by the table. A reinforcement-based Machine Learning (ML) process was established. This study employs the Proximal Policy Optimization learning technique [165] [166]. In Reinforcement Learning, an agent that observes specific environmental variables

2 https://github.com/Unity-Technologies/articulations-robot-demo/tree/mlagents

and takes decisions to maximise rewards controls behaviour. In the object contact training provided by VPepper, the reward at each step is the decreasing distance between the DT hand and the target item. Thus, manoeuvres that remove the hand from the target or lock the arm in a vulnerable posture (e.g., under the table or behind the head) are discouraged.

The network that controls the arm is comprised of six primary channels that arrange the flow of 69 information sources gathered from observations of joint angular values, spatial orientations, and relative distances. On the basis of these observations, the agent must either attain the objective (motivated by a reward score of 1) or accrue minor negative penalties proportionate to distance and delayed time. Notably, the huge issue size is typical for multijoint controllers due to the enormous number of position and orientation parameters that must be monitored. However, the size of the search space is limited by the restrictions placed on the real robot arm's movement.

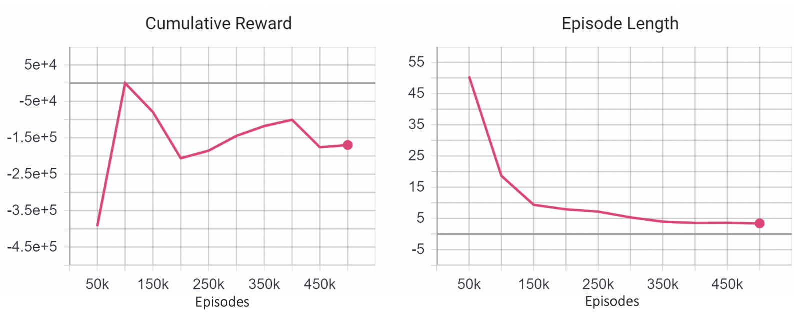


Figure 4.9: The cumulative reward and episode length for a 500K training session as reported by TensorBoard.

The TensorBoard report summarising the training session conducted on randomly positioned objects is displayed in Figure 4.9. The image displays the Cumulative Reward for each Episode (50.000 actions) on the left and the length in minutes on the right. The entire course took roughly two hours on an Alienware 17 3.60GHz, 32 GB of RAM, and an 8 GB Nvidia RTX 2080. The reward oscillates, but frequently with a favourable trend, as is evident (it is important to remember the negative penalties that every step accumulates because of the distance from the target). By looking at the final Epoch duration, it is possible to see how the VPepper controller is progressing in its learning: although the initial phase of 50.000 actions took 50 minutes to complete, the last session only takes a few seconds longer than 3 minutes. Given these circumstances, it is simple to comprehend one of DT's benefits over physical component training. The same goal would initially require a somewhat longer training period with a physical robot because to the restrictions of joint angular velocities. Additionally, it can result in excessive use of the robot, seriously harming its mechanical parts.

4.2.3 *Soft biometrics involved*

Let's examine the soft biometrics involved in this research in further depth.

- ***Photoplethysmographic biometrics.*** Photoplethysmography (PPG) is an optical technology that detects a biological signal to estimate blood flow within tissues in order to gather information on the user's health state as well as relevant data for fitness tracking applications [167]. Using light-emitting diodes (LEDs), these systems emit green or red light that is visible from a reasonable distance on the skin's surface. Human blood (particularly hemoglobin) absorbs green light well and reflects red light; hence, green light technology is often utilised. A PPG sensor records the change in light intensity based on the degree of absorption and reflection, and then uses signal processing to refine the data [168]. This technology is relevant to a wide variety of clinical tests, particularly those linked to cardiovascular systems, as well as oxygen saturation measurements. Red-light signals, when combined with infrared signals, enable oxygen saturation estimation. The combination of these measurements and arm movement permits also classification of the sleep cycle into its three distinct phases: awake sleep, light sleep, and deep sleep [169]. In recent years, the study of this sort of signal has attracted the interest of the research community due to the possibility of performing continuous authentication with low-cost devices capable of capturing data without requiring the users to take any action. Heart biometrics, also known as cardiac biometrics, are becoming increasingly popular. PPG sensors can be easily integrated into wearable devices such as wrist smartwatches due to their compact size. As the development of an effective remote health monitoring system becomes an increasing requirement (see the COVID-19 pandemic [169]), an increasing number of devices are equipped with such technologies. Non-complex analyses of PPG signals permit monitoring of parameters such as heart rate; complex analyses of these waves yield additional clinical information, including blood pressure, respiratory information, sympathetic nervous system activity, and heart rate variability [170]. Continuous monitoring of a patient's blood pressure, heart rate, blood glucose, and oxygen levels can assist in determining their health condition and if their current treatments are effective or should be altered. Globally, cardiovascular disease is the leading cause of death, accounting for an estimated 17.9 million fatalities each year [171]. Individuals at risk may exhibit an increase in these values; consequently, it is possible to intercept them using PPG sensors. In addition, the capabilities of PPG monitoring for health purposes can be expanded by analyzing its potential for use in authentication systems [172].

- ***Motion-based biometrics*** For the elderly, falls are the major cause of serious injuries. Detecting falls as they occur and, more significantly, preventing them by estimating the likelihood of this happening, helps prevent adverse health impacts that could otherwise result from the fall and, if necessary, enables one to take the most suitable action when it happens. To do this, a fall detector can be used. The compact size, low cost, and ease of use of wearable devices make them suitable for this purpose. In addition, they permit continuous monitoring and the acquisition of physiological data while the subject does his daily activities. However, these technologies are far from accurate, as they generate a huge number of false alarms for every true fall detected. The dependability of the fall detectors has been studied and continues to be investigated. In [173], the efficacy of the information acquired by them in conjunction with data from PPG-based biometrics to infer body location was examined. Wearable systems commonly used for this purpose are inertial measurement units, which acquire data from accelerometers and gyroscopes [174]. Multiple 3-axis sensors (mostly accelerometers and gyroscopes) are used by an inertial sensor to measure linear acceleration and angular velocity in its own three-dimensional local coordinate system. Moreover, compared to other sensor technologies, inertial sensors are a sensor unit that lends itself well to the creation of a wearable monitoring system due to its capacity to measure motion-related characteristics with high precision and accuracy. There are three important factors to consider when designing an effective fall risk assessment system based on inertial sensors: sensor placement, tasks to be performed, and key features to extract and analyze. The characteristics can be divided into different categories such as spatial, temporal, frequency, linear acceleration, angular, and non-linear characteristics. The linear acceleration characteristics are related to postural stability during activities; the spatial ones include, for example, the number of steps; the temporal ones are related to time as the speed of the gait; and the angular ones include joint range of motion and rotation during gait and movements, etc. Since different characteristics represent different gait and movement characteristics, an appropriate combination is needed to distinguish between those associated with a fall and those that are not. There are also significant differences when the analysis is performed retrospectively or prospectively [175]. Therefore, as we have said, the parameters that describe the movement and also the position of the body are indispensable in the diagnosis of the risk of falling, but, at the same time, they also provide significant information on the characteristics of gait. Changes in gait metrics may not only be predictive of falls but may also precede cognitive impairment, according to studies [176]. A great deal of research has been conducted in the past on activity recognition studies that employed acceleration and gyroscope data to determine the user's identification based on their physical activities (e.g., normal walking) [177].

4.3 TRUST AND SECURITY IN HUMAN-ROBOT INTERACTION

Pepper is a humanoid robot with limited computational capabilities to handle its sensors and actuators, preventing it from simultaneously processing vast amounts of data or performing difficult tasks. To improve its operation and interaction with the environment, the robot can be put in contact with a variety of smart satellite objects and services, ranging from simple environmental sensors to smart cameras with DL capabilities.

Interconnection between interconnected devices (referred to as "smart objects"), uniquely identified with the ability to transfer data into the information network without requiring human interaction, is part of the Internet of Things (IoT) paradigm. The architecture of the Internet of Things can be controlled as a physical, virtual, or hybrid system. In the past decade, there has been a growing interest in the usage of IoT systems in numerous domains, and their applications are diverse: smart cities, smart homes, Industry 4.0, etc. The study of smart homes is a fast expanding subject of study. They derive from home automation, which enables remote and timed control of a wide range of devices, from lighting to heating, with the goal of enhancing comfort, energy efficiency, safety, and consequently, quality of life.

In our [110] work, we examined the concept of using the humanoid robot Pepper as an interface between the user and smart home objects. Pepper, as the hub of a smart system, is connected to external IoT devices with which it may share gathered data, extract knowledge, and make global judgments.

The addition of biometric, emotional, social, ML, and other capabilities to Pepper, which enable increased functionality and extra tools for managing users and the environment, presents security challenges and other challenges. Our ecosystem permits the extraction of a variety of data, including highly personal and sensitive information, which raises privacy concerns. The robot, its interaction with the environment, and any vulnerabilities revealed by smart devices in its ecosystem can serve as an entrance point for an attack on the smart home, posing a threat to its security and privacy and diminishing users' trust in the system. Spread and acceptance, even among the most sceptics, of the applications and services that the IoT can provide are highly dependent on the reliability notion associated to these systems. The difficulty is not only in preserving private information and securing sensitive data, but also in managing the vast volume of gathered information to generate the most effective and appropriate reaction.

In order to prevent assaults and enhance security, we match each system activity with the total context seen by the entire ecosystem of smart objects. The system determines what actions to execute or to allow based on the confidence derived from user actions and behaviours. This paper describes and analyses the experience and how the selected semantic trust model mitigates the vulnerabilities and threats posed by cyber-attacks on smart homes.

The notion of trust is fundamental. It has been implemented mostly in HRI, a direction typically followed when the human must "trust" the robot. However, the concept of trust is expansive, as it encompasses various factors that are dependent on both people and the environment. In some Pepper applications, such as [178], the concept of trust is tied to the robot's capacity to complete specific tasks as judged by the users. However, the concept of trust in the architecture we wish to discuss in this work is diametrically opposed to the conventional one. Instead of saying, "Can I trust Pepper?" we would like the robot to be able to trust or not trust certain subjects. If we begin with the work in [179], which focuses on human-to-robot trust, we find several interesting principles that can be applied to our system. In this context, trust is defined as a collection of objective and subjective characteristics that a robot must possess to be believed. For instance, it may appear fastidious, reputable, reliable, truthful, etc. The crux of our argument is that if a model can be developed to measure how much a human trusts a robot, even with a large set of variables that can predict how the human's impressions change during the interaction [180], then the reverse process can be implemented to measure human trustworthiness. Existing literature is lacking in this aspect, but new works are emerging. This is studied in particular in [181], which is motivated by the desire to create a real HRI system in which trust is considered in both directions. Based on the Theory of Mind and supported by an artificial cognitive architecture, the robot can alter its level of confidence in the human by using memories. Pepper's perceived trustworthiness is essential to our proposed trust-based IoT ecosystem; hence, it is a vital component of our paradigm.

4.3.1 *Semantics and trust: the proposed model*

This study examines how IoT , trust, and robotics may collaborate to create an IoT-based ecosystem that grants or refuses users permission to execute specific tasks based on the level of trust they develop while interacting with the ecosystem. When a user takes an action, the full context identified by the smart object ecosystem is evaluated to determine the level of trust. In an effort to describe user and action trust in a manner not constrained to a certain knowledge or operational context, we choose to employ semantic instruments.

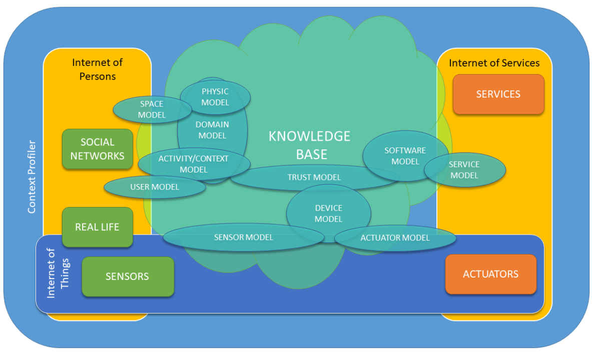


Figure 4.10: The generic model of trust with domain-specificization.

In fact, the ontological model proposed for this study is adaptable to several domains, as Domain ontologies can be used to model the specifics of each new domain or its evolution. Figure 4.10 depicts a potential "Knowledge Base" arrangement for the IoT sector. As indicated, the "Internet of Things" layer shares with the "Knowledge Base" the sensor- and actuator-organized semantic representation of devices. We assign each sensor a level of confidence proportional to the accuracy of its biometric or behavioural identification. As for the level of trust offered by a smart device, the reputation of an entity (mostly the users asking a command) is not predetermined but instead connected with the specific context obtained as shared information about the entity's conduct [182]. In the suggested approach, the Truster (i.e., the smart device or Pepper) is the entity responsible for judging the Trustee's trustworthiness (the user that performs an action). The model illustrated in Figure 4.10 also demonstrates the incorporation of a number of domain-specific ontologies, used for describing spatial and physical links and capable of connecting individuals through their social network activities (as depicted in the rounded box "Internet of Persons"). In the case of commercial IoT devices, their back-end functionality is described by the "service model" which describes all exposed features and commands.

In a domotic apartment, using information from smart objects to create a semantic ZTA is made possible by using an ontology. The proposed semantic approach is based on the idea that if a user wants to run a command or a chain of commands, the requested operations can be easily classified as restricted by semantic inference, and the trust needed for their execution can be calculated accordingly. Automatic reasoning (Pellet[3]) is used to classify user commands, and if a trust evaluation is needed (for example, for unlocking commands), the information from all smart devices is evaluated, and the user's trust level is calculated based on several identification criteria and an evaluation of how the user acts. In more detail, for the automatic reasoner Pellet to work, the class "state/trustEvaluation" is defined to be the same as every other class with the property "requestSmartLock" set to true. Every IoT command is set up as a separate part of the ontology, with a list of all the steps needed to carry out the

3 https://www.w3.org/2001/sw/wiki/Pellet

command as a whole. The automatic reasoning classifies it, and if the property "requestSmartLock" is set to true, like in the case of the smart-lock opening command, Pellet knows that it belongs to the class "state/trustRequired." Figure 4.11 shows how the tool

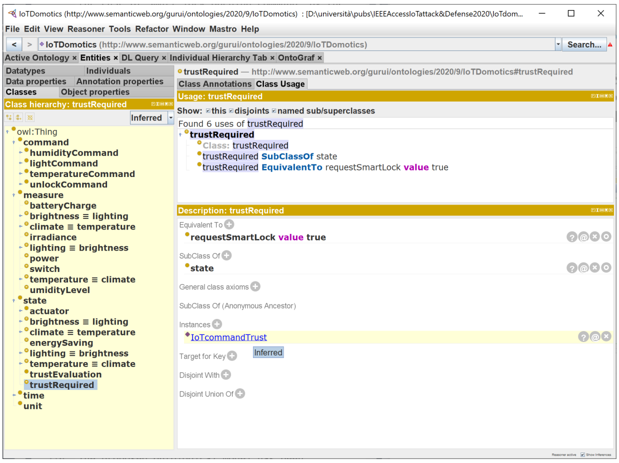


Figure 4.11: The inferred class equivalence in IoTdomotics ontology.

Protegè[4] shows the effect of Pellet reasoner on the classification of the individual "IoTcommandTrust" as "trustRequired." Let's point out that the "IoTcommandTrust" individual doesn't have a "is_a" relationship without Pellet's inference. Instead, it has an "owl:Thing" relationship. In the case of a command chain, the "follows" connection establishes an ordered list of commands, and the "swrl:add" operator is used to calculate the total trust value required by each command in the chain.

Figure 4.12 depicts, in the lower and right-hand frame, the Protegè description of the inference utilised for the trust computation of a command chain. The figure illustrates the three classes inferred for the "command4" instance as a result of its property "requestSmartLock" being set to true. These classes are highlighted in yellow on the "Description" tab. As a result, and because of the "follows" relation, the "needs trust sum property" of command4 is 490, computed as the sum of all the trust requirements (i.e., the property "needs trust") of the preceding commands. When authorising an operation, the same approach

4 https://protege.stanford.edu/

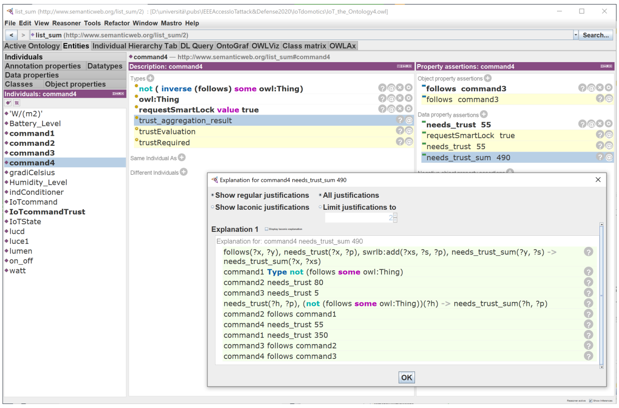


Figure 4.12: The computation of a command chain's required level of trust.

is used to compute the overall trust estimation of a user provided by smart objects (Pepper included).

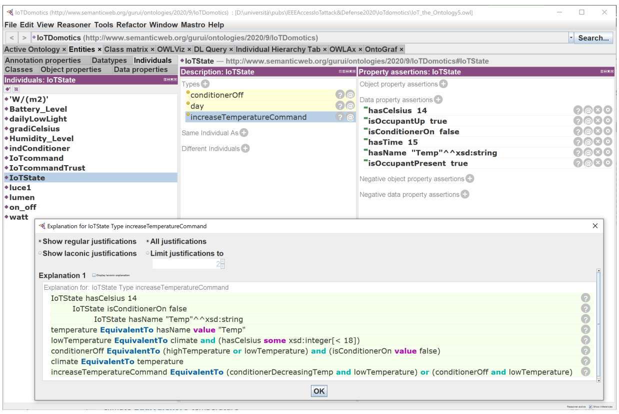


Figure 4.13: The classification of IoT state information with the associate command.

For the sake of completeness, the IoT sensors' state information (e.g., illumination, temperature, humidity, etc.) is processed with the same reasoning used for trust. As an illustration, the individual IoT state depicted in Figure 4.13 does not define any ""is_a"" relationship and is established by setting the sensor-read values of its characteristics. The particular temperature ("hasCelsius") is 14 degrees, the user is present and awake ("isOccupantPresent" and "isOccupantUp" are both true), and the IoTState is classed as "increaseTemperatureCommand," causing the central control to activate the air conditioner and raise the air temperature.

4.3.2 *The Smart Home case study*

The smart home ecosystem was chosen to evaluate the quality of the suggested model. It consists of a network of smart devices, including the Pepper robot. This setting is utilised for ontology specialisation and automatic trust evaluation. The smart home is equipped with industrial IoT devices, such as sensors and actuators, and smart cameras with AI built on Neural Compute Sticks Movidius:

- A number of smart objects based on Movidius-Raspberry plus cameras, inferring, based on the Deep Neural Networks loaded, user position, sex and age, room objects, etc.
- Pepper robot, enhanced with Power-up capabilities and capable of emotion identification (through user images and speech), heart rate detection, user face following, and user collaboration detection;
- A number of interconnected electrical and temperature sensors.
- A network of switches capable of changing outdated objects into smart devices.
- A number of actuators, ranging from RGB smart lamps and air conditioners to door locks that may be operated remotely.

As depicted in Figure 4.14, Pepper is the center of the system, getting involved if requested by the domotic systems, as described in greater detail below. For instance, a camera in a room can detect suspicious behaviour, and as a result, the domotic system will notify Pepper. Once Pepper is notified, it will proceed to the given place using a navigation system based on home modelization and obstacle avoidance. Once Pepper has located the

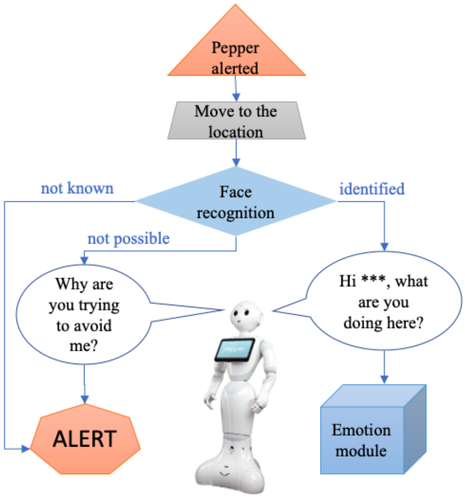


Figure 4.14: The main involvement of Pepper in the proposed ecosystem.

individuals exhibiting suspicious behaviour, it will attempt face recognition. It will detect the user's desire not to be identified based on head pose and movement estimation algorithms. If the user attempts to evade Pepper or if he or she is unknown, an alarm will be sent to warn the system owner. If the identification is successful, Pepper will ask the user directly what he was doing in the room. Simultaneously, an emotion module will be engaged and will detect rapid shifts in the user's mood, which may indicate suspicious behaviour. To achieve this, the emotion module will consist of:

- Face emotion recognition: face features will be analysed.
- Voice emotion recognition: peaks and intensities of the voice signal will be evaluated.
- Heart rate detection: fast changes in heart rate will be captured contactlessly from the face utilising successive frames.
- Sentiment Text analysis: the user's spoken utterances are also transformed to text and analysed as such.

Each of these four elements is capable of eliciting a positive or negative response. Positive expressions include those that are neutral or joyful, whereas negative expressions include those that are astonished or angry. Every potential combination of emotions recognised by these approaches will either increase or decrease the user's trustworthiness. After the trust upgrade, Pepper will notify all ecosystem devices of their new identification and trust value. As a result, the user will either be able to do additional actions or will be observed more closely in his or her subsequent interactions with the environment, depending on whether trust has increased or decreased.

Table 4.3: Methods involved, devices, accuracy and the strength coefficient of the biometric trait.

Method	**Device**	**Accuracy**	**k**
Face+Voice Emotion	Pepper	74.38%	0.7
Heart Rate	Pepper	5 bpm	0.6
Identity from Face	Pepper	94.1%	1
Gender from Gait	Camera	80.7%	0.3
Cooperativity from Gait	Camera	97.58%	0.2
40 Facial Attributes	Camera	91%	0.4

The proposed architecture for Pepper and Smart devices employs the aforementioned methods, which carry with them particular accuracy and limitations. In Table 4.3, the accuracy of the algorithms on the corresponding devices is displayed. As can be observed, if we evaluate each approach separately, there are inherent uncertainties in each. Nevertheless, it is conceivable that these procedures, which operate after getting information from another device, from a different perspective, or utilising other strategies, can enhance the total precision. If a camera recognises a subject's gender or some of his/her facial characteristics, for instance, Pepper's identity recognition will be more accurate because the pool of potential subjects will be reduced. Some biometric characteristics are more powerful than others, regardless of their accuracy, due to their nature. In light of this, we shall refer to the coefficient of biometric trait i strength as k_i. These coefficients, reported in the rightmost column of Table 4.3, were assigned based on the discriminatory power and biometric traits in the following ascending order: cooperativity from gait, gender from gait, 40 facial attributes, heart rate, face with voice emotion, and identity from the face. Each k_i is between 0 and 1. There are two distinct types of biometric information. Gender from gait, 40 facial attributes, and identity from face all require a priori knowledge to identify the incoming data as positive or negative. Face with voice emotion, heart rate, and cooperativity from gait do not require a priori information because they have universally positive or negative properties. An increased heart rate, greater than 100 beats per minute, is not typical for a subject at rest; attempting to dodge the camera can be indicative of a malicious purpose. All of these qualities, when discovered, have a negative impact on the trust value; therefore, their esteem will be removed from the total trust. Also, the values of the biometric features that require a priori knowledge will be deleted if the single biometric check does not produce the predicted result and added if the result is positive. In conclusion, the model of trust under consideration employs the following formula:

$$t = \pm k_1 * a_1 \pm k_2 * a_2 \pm ... \pm k_n * a_n \tag{4.1}$$

where the sign determines whether the n biometric response characteristics are positive or negative, k_i represents the biometric coefficients, and a_i represents the biometric recognition algorithm's accuracy.

4.3.3 *Soft biometrics involved*

Let's examine the soft biometrics investigated in this research in further depth.

- ***Facial attributes.*** In order to compare two facial images, forensic examiners must first perform a thorough visual comparison of the two images before producing a morphological report. They pay close attention to each and every feature of the face in addition to the complete face. They specifically perform a thorough morphological comparison, studying the face region by region (such as the nose, lips, eyebrows, etc.), as well as looking at traits like marks, moles, wrinkles, etc. These supplementary characteristics are known as soft biometrics [183]. Even though particularly tiny traits are not sufficiently persistent, they are nonetheless momentarily invariant, hence falling within the category of soft biometrics. Due to the significance of colour in a wide range of applications, including face detection, picture recovery, etc., colour can also be regarded as an interesting characteristic. Eye, skin, and hair colours are examples of colour features [184].

In recent decades, a great deal of research has been done on the face as a biometric trait. Due to its singularity, particularity, and permanence, it is one of the biometric traits that has been researched and shown to be the most dependable.

Despite the fact that good results have already been obtained in controlled settings, there are still a number of difficulties with facial recognition in realistic situations, due to poses, occlusions, etc. Recent research has focused on the potential benefits of adding soft biometric facial traits to facial recognition algorithms. Investigating the possibilities of distinctive facial features like the nose, lips, hair, and so on is seen as a prospective area of study.

In fact, soft facial biometrics have a number of important advantages that enhance the functionality of traditional facial recognition systems. Due to their simplicity in extraction under various circumstances, it is also possible to perform recognition simply based on these attributes or to use them to enhance the performance of conventional facial recognition systems.

These facial attributes become much more important and offer more useful information for matching identities in unrestricted circumstances, such as when the facial image is recorded in a non-frontal or occlusion-limited pose. Numerous privacy issues still exist because their automatic extraction, as with other soft biometric attributes in general, is possible without the user's knowledge.

- ***Gait.*** A person's gait is as distinctive as their voice. Gait recognition is therefore possible. Taking this into account, ML-based algorithms for gait recognition were developed. Systems based on this biometric characteristic are able to recognise a person from an image even if their face is obscured, turned away from the camera, or covered by a mask. Gait recognition refers to verifying a person's identity based on how they walk.

The system examines walking patterns, height, speed, and silhouette. The advantage of using such a trait is that it is more convenient than retinal scanners or fingerprints in public places because it is unobtrusive. Moreover, gait recognition is unlikely to be deceived: each person's gait has no duplicates.

The gait of a person can be used by behavioural biometrics as an interesting indicator for identification purposes from a distance. About 24 distinct features and movements that make up a person's distinctive gait can be detected as they walk [185].

With the development of CV techniques, there are many approaches to human identification by motion on video, using both natural biometric features (the human skeleton, silhouette, and changes while walking) and abstract features.

A gait recognition system uses the shape and movement of the human body to recognise it. The software locates a person's silhouette on a video using CV techniques and examines their motion. This information helps develop a model of human behaviour.

However, gait-based identification has taken on a new twist with the increasing use of smartphones equipped with accelerometers and gyroscopes. This is due to the special capability of accelerometers and gyroscopes to record gait patterns.

- ***Heart rate.*** Due to the strong (and literal) relationship between our heart and our emotions, the heart rate variation not only represents our physiological state but also the degree of emotional arousal we are experiencing, as we have already discussed in Section 4.1.3. Nonetheless, this soft biometric is also of particular importance to the *Liveness Detection* research field. In biometrics, Liveness Detection refers to a computer's capacity to distinguish between a physically present human being and a spam bot, an inanimate spoof artifact, or phony video/data.

The most commonly used technique in access control and security is facial recognition. Comparatively speaking, it is easier to implement than other biometric methods. Although it has the benefit of being a non-intrusive method of access, a system like this might not be able to tell a real person from a photograph of them, and it might be manipulated by showing a photo or playing a video of the person to the camera. In such a situation where a cheater can easily gain access to the system by presenting a copy of the person's image to the camera, face vividness detection is important to detect whether the captured face is a live or fake captured image.

Face vividness identification uses real-time analysis of specific facial traits to identify vividness. Studies in this area have been a hot research focus recently. Particularly, a number of works have taken advantage of heart signal detection. The two main methods for getting this kind of information are the contact technique and the non-contact approach. The most common techniques are contact-based ones, such as electrocardiography or PPG . But there have been some well-known recent initiatives to obtain the cardiac signal from a distant camera [186]. Particularly, color-based techniques make use of the minute colour variations that result from blood flow. These techniques enable remote cardiac signal acquisition through a common camera and no sensor equipment.

- ***Voice.*** Already discussed, more details in Section 4.1.3.

Part III

PERIOCULAR SOFT BIOMETRICS

5 SOFT PERIOCULAR FEATURES

Profecto in oculis animus habitat.

— Plinio il Vecchio, Naturalis historia, 11, 54, 145

Profecto in oculis animus habitat[1], asserted Plinio as early as the first century A.D. The eye has always been considered not only an instrument capable of receiving and translating the external world but also a glimpse into the inner dimension, capable of shedding light on emotional experiences and reflecting cognitive efforts. Our eyes, brains, and the world around us work together in a complicated way to give us our personal visual experiences. It gives us a sense of sight, color, stereopsis, distance, pattern recognition, motor coordination, and more.

The eye, also called the eyeball, is a spheroidal organ that is 24mm in length, weighs around 7.5g on average, and takes up about 6.5cm^3 of space. Internal, intermediate, and external components make up the whole (Figure 5.1). There are three individual compartments within its three membranes (or "tunics"). The cornea is the curved and transparent front component, and the sclera, which extends back to envelop the eye, maintains the name. These two layers combine to produce the outermost tunica fibrosa. The choroid, eye, and ciliary body make up the intermediate tunic (vascular tunic), while the retina forms the innermost (nerve) tunic. Aqueous humour is present in the anterior chamber, which is located behind the cornea. Vitreous humour is contained in the vitreous chamber, which is located behind the lens.

There is a multi-step process that occurs between the eye and the brain that determines what we see and how we see it. The optical and physiological features of the human eye and their connection to vision date back thousands of years. What we gaze

1 Surely the soul resides in the eyes.

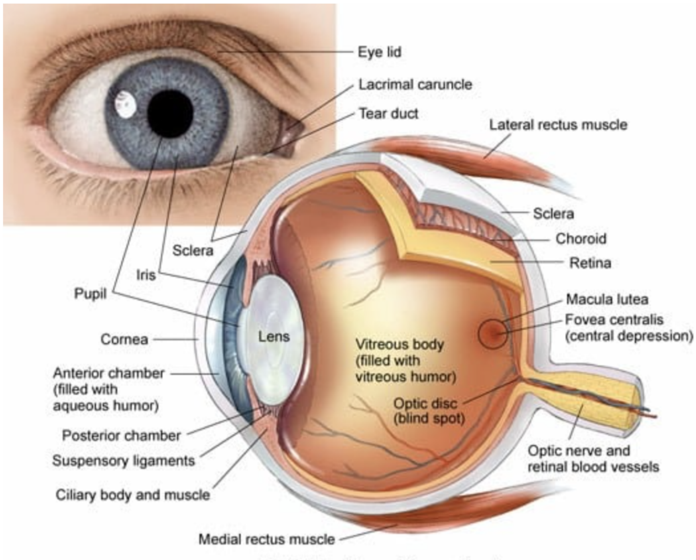


Figure 5.1: Diagram depicting both the external and internal structures of the eye [7]

at demonstrates the complexity of our eyes through their range of sizes and shapes. The increasing ubiquity of gaze-analysis technology brings with it the ability to track gaze measurements with increasing confidence. Such technology is used to measure and collect eye characteristics by recording them as data. The data provide unprecedented insights into human actions and environments, digitizing how people communicate with machines and opening up new avenues for passive biometrics-based classification such as emotion prediction and beyond. Blink rate, blink time, and blink latency are all metrics that can shed light on the health of your eyes. Gaze distribution, fixation frequency, fixation duration, saccadic duration, saccadic peak velocity, and the number of saccades are some of the other metrics that can be considered. Pupil size has been the topic of other investigations. Researchers have examined eye size in relation to mental and emotional stress as well as its utility as a biometric attribute for identification and classification. Indeed, eyes, among other biometric features, offer a variety of physical and behavioral properties that make them particularly ideal for biometric identification purposes due to their high specificity and complex mechanical logistics of reproduction.

Eye features are highly resistant to counterfeiting because they depend largely on brain activity and the properties of an individual's extraocular muscles. So, based on what we know about the human brain and the structure of the muscles around the eyes, it is not possible to accurately copy eye-movements outside of a living human being. In biometrics, identity fraud is an important issue to mention. To date, most people do not have a clear understanding of the potential risks in today's technologically advanced world. This is precisely why it is essential to explain and emphasize one of the most troubling aspects of recent years: deepfakes. This paradigm represents an example of a possible threat to people's digital identities. Since most deepfakes are almost always facial transformations, this is where manipulation can be easiest to detect. Often, facial features and expressions, as well as the contour lines around the face, do not match when looking closely at the image. Furthermore, the skin and aging signs of the two people mixed in the deepfake are not generally identical. Other features of manipulation within facial transformations include unnatural eye-movements or frequent blinks. Using identity verification based on biometrics and liveness detection can be a good way to stop these kinds of attacks. The oculomotor one is very hard to copy mechanically and can be a useful ally because of this. This makes systems based on eye properties attractive strategies not only for biometric applications but also for the built-in liveness detection capability.

The study of periocular biometrics is therefore interesting for:

- use them in a biometric fusion system to improve its overall performance;
- provide integrated liveness detection capabilities of a system;
- identifying behavioral patterns and information related to the subject's cognitive-emotional state;
- independent soft biometric modalities in the context of demographic classification.

More details on the four soft periocular features under consideration (blinks, fixations, eye movements, and pupils) are provided below.

5.1 BLINKS

Reflex contraction of the orbicularis muscle of the eyelids, resulting in the closing of the eyes in response to an intense light stimulus or a sudden visual stimulus. The afferent branch of the reflex is represented by the optic nerve, which travels to the visual cortex and from there to the facial motor nucleus, which represents the efferent branch. Blinking is a continuous but irregular reflex movement that occurs at a rate of about 10–20 episodes per minute, depending on the subject's concentration and emotional state. Normally, this reflex (bilateral) is evoked by corneal contact, slight blinks on the forehead or around the eyes, visual threats, or turning the eyes to one side; all these stimulations are followed after a short time by adaptation, except in the case of the evocation of the corneal reflex. Blinking becomes more frequent in cases of corneal or trigeminal irritation or blepharospasm and becomes less frequent in parkinsonian or patients with paralysis progressive supranuclear.

A blink thus consists of the simultaneous movement of the eyelid and eyes. Two antagonistic muscles are involved: the Orbicularis Oculi (OO) muscle and the Levator Palpebrae Superioris (LPS) muscle. The OO muscle, a large flat elliptical spherical muscle, induces eyelid closure. The LPS muscle serves to elevate the upper eyelid. A third muscle is the Müller muscle, which can adjust the width of the palpebral fissure and helps the LPS muscle keep the eyelid open. Each region of the OO muscle contains motor units of different sizes. During different types of beats, the size of the motor unit determines how accurately the muscle force can be increased or decreased.

The eyelids serve as a protective barrier between the cornea and anything in the environment that could harm it. If the cornea lost its opacity, vision would be compromised or lost entirely. Blinking is a natural, easily observable, and easily accessible behaviour that displays the activation processes of the central nervous system without intentional intervention since eyelid movements require basic brain commands and low active forces. Its analysis could reveal muscle or nerve abnormalities, which makes blinking a very helpful source of information and highlights the need to examine its distinctive characteristics. There are three distinct types of blinking:

- ***Spontaneous blinking.*** Spontaneous blinking happens without the requirement for a stimulus. The objective is to cover the cornea with a tear film to prevent the ocular surface from drying out. The average blink rate appears constant for a given individual; nevertheless, changes in blink frequency are detected when blink frequency is analysed across shorter time intervals. Four blinks per minute would be sufficient for maintaining the pre-ocular tear film, which is roughly the blink rate of neonates [187]. Multiple factors, such as time of day, environment, emotional state, mental load, and activity, might influence blinking. Low humidity and high temperature can cause the pre-ocular tear film to rapidly dry, resulting in an increase in blink rate [188]. In 1980, Von Cramon and Schuri discovered that counting from one to one hundred caused participants to blink more frequently than pronouncing the alphabet [189]. When there is a great demand for visual attention, such as when youngsters play video games, blinking frequency reduces [190].

- ***Voluntary blinking.*** Voluntary blinking is blinking that the individual performs voluntarily. They are the blinks that are triggered by conscious thought. Such blinks are used to communicate, for example, to express innocence or to imply that what was said was not particularly important. The speech of individuals with profound motor paralysis is also replaced by voluntary blinking [191]. Understanding the normal variation in blink rate and the variables that create this variation in order to make the blink rate a more accurate and effective tool for studying neuropathological diseases in which the blink rate is altered is the objective of

[192]'s research. In this study, the scientists demonstrated how men and women differ in their ability to control the voluntary blinking rate. In their research, the mean blink suppression time for men was substantially longer than for women. Men also looked more capable of rapid blinking, as their average rate of accelerated blinking significantly exceeded that of women.

- ***Reflex blinking.*** Reflex blinking, also known as the corneal reflex, is a short-lived rapid closing action caused by a variety of external stimuli, including bright lights, approaching objects, loud sounds, etc. These are the fastest varieties of blinking. The three basic sensory modalities that are involved in this external stimulation are: a physical stimulation, such as contact or a foreign body on the cornea; a light stimulation, such as an intense light ("optical reflex"); and an auditory stimulation, such as sounds with intensity greater than 50 dB. The reflex from optical stimulation (optic reflex) is slower than that brought about by physical stimulation and is mediated by the cerebral cortex of the occipital lobe. In general, regardless of stimulation, it always occurs simultaneously in both eyes, even if stimulation occurs in only one eye. Failure of the contralateral eye to respond simultaneously, could indicate pathology.

5.1.1 *Blinks and mental workload*

The various parameters associated with blinks, such as duration or frequency, speed, and latency, can be used to extract information about subjects' responses to different stimuli. It is known that when a task requires visual stimulation, the person attempts to reduce the amount of time spent with eyes closed, or the number of blinks, by concentrating on the activity. The purpose of different studies was to determine if this phenomena could provide meaningful signals of mental workload for jobs requiring intensive visual attention, such as reading, driving, etc.

In a study conducted on an accurate car simulator, i.e., in a very realistic driving environment, Lal and Craig [193] found that the transition from normal driving to tiredness was marked by an increase in blink rate and cessation of vertical and horizontal eye-movements. This data, along with others in the literature, appears to be consistent with the existence of a continuum between a very low blink rate, relative to tasks requiring high visual attention, and a rise in blink rate just prior to drowsiness and during monotonous tasks.

Significant sleep deprivation leads to an increase in the blink rate, which researchers believe may be due to patients' attempts to maintain wakefulness through increased oculomotor activity [194]. Additionally, these researchers found that loss of attention, unrelated to a drowsy phase, similarly caused an increase in the blink rate; this could be interpreted as an attempt, during a monotonous task, to maintain wakefulness.

Some scholars believe that blinks may convey useful information about central nervous system activation and exhaustion levels; as fatigue in completing a task increases, so does the frequency and duration of blinks, resulting in a decline in performance [195].

In accordance with this idea, fatigue and tiredness have been found to be associated with an increase in blink frequency [196]; for example, this index has also been used to get feedback on the fatigue of pilots and co-pilots of military aircraft [197]. Luckiesh and his colleagues tested blink frequency to see if it was a good way to measure "visual fatigue" [196]. They found that this value always goes up when a certain visual task takes longer or lasts longer. In line with this, other tests with people doing tasks that required them to pay attention for a long time also showed an increase in blinking. Carpenter found that the number of blinks increased by 43 percent over the course of a 2-hour vigilance task [198]. After an extensive literature review, it can be concluded that blink frequency measurements are a robust measure of task-induced fatigue.

In contrast, fewer blinks were associated with tasks requiring greater concentration and focus [199]. Due to the fact that blink inhibition is required to limit the loss of information caused by the disruption of visual perception, the number of blinks was significantly reduced under conditions requiring substantial attentional commitment [200]. With increased work pressure, not only does the frequency of blinking decrease, but so does the duration of blinking.

5.2 EYE-MOVEMENTS

Vision is one of the most fundamental cognitive actions that each of us possesses to correlate with the external world, from which we derive information of vital relevance for the goals of knowledge and interrelationships between individuals, between them and the environment, and ultimately survival. To attain this objective, we must continuously move our eyes; this is a very complex process as it is governed by extensive mechanisms for highly fine motor regulation of the eyeballs. This process begins with the activation of a distinct group of neurons and concludes with the contraction of some eye muscles and the relaxation of others. Eye-movements serve a well-defined purpose: they permit visual perception, or the processing of visual information from the environment. Vision occurs when light reflected from objects, the surroundings, or writing enters the eye and modulates the activity of the photoreceptors in the retina. The two functional goals of eye-movements: orienting, stabilizing and maintaining gaze, are achieved through two major classes of movements: rapid movements, also called saccadic, and slow movements, also called smooth pursuit.

There are four basic types of eye-movements:

- ***Saccades.*** A rapid eye movement that pulls an initially peripheral region into the visual field's center (the fovea). The distance between the peripheral region of interest and the fovea is known as the saccade's amplitude. The direction and magnitude of saccades are mainly within voluntary control, although saccades in all other directions are strongly stereotyped. For instance, the duration and velocity of a saccade are unique to its magnitude and cannot be altered or manipulated intentionally. The range of sporadic reaction times is between 120 and 350 ms. Due to their extreme speed (up to 900°/s), saccades cannot be guided by visual feedback. There are four types of saccade classifications, which are predictive saccade, antisaccade, memory-guided saccade, and visually-guided saccade [201].

- ***Slow pursuit movements.*** Slow pursuit movements are voluntary movements that fix a moving object on the retina so that you can perceive its fine features. They are significantly slower eye-movements meant to maintain a moving stimulus on the fovea. The observer can choose whether or not to track a moving stimulus with these motions. Saccades can be deliberate, although they can also occur unconsciously. Surprisingly, however, only highly skilled observers are capable of executing a pursuit movement without a moving object. The majority of individuals who attempt to shift their eyes smoothly without a moving target perform a saccade.
- ***Vergence movements.*** Vergence motions allow the fovea to keep up with visual stimuli that are approaching or receding. Vergence movements are associated with the process of accommodation, or the change in lens curvature, which maintains the stimulus in focus. In contrast to other forms of eye-movements in which the two eyes move in the same direction (conjugate eye-movements), vergence movements are disconjugate (or disjunctive); they involve a convergence or divergence of the lines of sight of each eye in order to observe an object up close or far away. Convergence is one of three visual reflexes triggered by an interest in a nearby item. Lens accommodation, which brings the object into focus, and pupillary constriction, which increases depth of field and sharpens the image on the retina, are the other components of the so-called near reflex trio.
- ***Vestibulo-ocular movements*** The oculo-vestibular reflex is an eye reflex that stabilizes images on the retina during

head movements by causing an opposite eye movement, such that the stable image remains in the center of the visual field (fovea, the point at which there is maximum visual resolution). The oculo-vestibular reflex does not require visual stimulation (it occurs even in situations of complete darkness or with eyes closed). The lag between head and eye-movements is only 14 milliseconds. The oculo-vestibular reflex alone does not provide adequate compensation for head movements, particularly when slow, progressive motions are made; however, this mechanism is highly effective when quick, transitory movements occur. When sustained head rotation occurs, this reaction is inadequate and must be supplemented by the optokinetic reflex.

5.2.1 *Mental effort and emotions: evidence from eye-movements*

According to neuroanatomical data, fear, anxiety, and vigilance are associated with the brainstem motor nuclei that control eye-movements. Since the early 1900s, they have been measured and studied extensively to draw conclusions about perception, cognition, and brain function in numerous fields of psychology, cognitive science, etc.

Coordination between eye-movements and perceptual attention enables the selection of objects, characteristics, or regions of interest, thereby providing vital insight into how humans acquire information. In investigations of mind wandering, defined as instances in which individuals are unaware that their current cognitive aim has been momentarily supplanted by another worry, participants who reported periods of mind wandering had, on average, fewer sophisticated eye-movements [202].

On the other hand, saccadic speed appears to be a reliable indicator of fatigue [203], whereas emotional and physiological arousal appears to have a greater influence on the reaction time (latency) of saccadic movements rather than their speed; for instance, shorter reaction times have been recorded in response to shock threats [204]. In [205], the authors observed that when asked to evaluate the presence of a particular emotion on a face, participants focused on a common set of facial regions, but also employed emotion-specific eye movement methods in both emotional and neutral faces. These findings support the notion that focusing attention on specific diagnostic areas is advantageous for emotion processing and that these methods may be influenced by both stimulus-driven and goal-driven factors. The findings for goal-driven techniques are consistent with earlier studies indicating that goals and perceiver traits may influence the eye-movements used during facial emotion assessment [206].

In mental spatial transformation tasks, several cognitive processing phases can be recognized in oculomotor behaviour. Eye-movements were monitored while participants conducted a mental folding activity in the work [207]. The relationship between task difficulty, gaze proportion on each stimulus, gaze changes between stimuli, and reaction times was examined by analysing gaze behaviour. The authors discovered a monotonic decline in switch frequency and reference object gaze proportions as difficulty increased. [208] instead investigated the predictive role of eye-movements in mental arithmetic. Mental arithmetic provides a platform for investigating the cognitive processes underpinning abstract thought. The authors noted that the predictive function of horizontal eye-movements, in particular, is crucial for comprehending how attention narrows the range of feasible solutions.

The research in [209] proposes a method for objectively assessing mental workload by evaluating the changes in eye movement characteristics during various visual search activities. By altering the difficulty of visual search activities, eye-movements were examined to determine if they may be used to categorise mental workload. Consequently, the five indices (saccades amplitude, saccades velocity, fixation duration, blink duration, and pupil diameter) differed significantly across visual search tasks with moderate and high workload. Moreover, as task demand grew, saccades amplitude, saccades velocity, and blink duration fell dramatically, although fixation duration and pupil diameter gradually increased.

5.3 FIXATIONS

A fixation is typically defined in visual neuroscience as the duration between two saccadic eye-movements during which no eye-movements occur, with a duration of 100–400 ms [210]. A scanpath is the resulting series of saccades and fixations. Eye fixations are investigated within two distinct functional domains: *free-viewing* where observers are needed to just view a visual target, and *task-based* where observers can read literature or explore a static environment with a specific purpose. Eye fixations are frequently referred to in this context as "inter-saccadic intervals" to indicate that they were collected during a reading or visual search task. Visual exploration activity consists of a series of fixations interspersed with saccadic eye-movements that direct the fovea to various portions of the image. Fixations capture a person's visual attention when they are focused on an attractive object. Using multilabel classification, Vidyapu et al. [211] developed an attention prediction on webpage pictures. In a study conducted by Roy et al. [212], instead, the authors devised a cognitive approach for identifying ambiguous images.

5.3.1 *Fixations: behavioral and emotional-cognitive process analysis*

Antes and Kristjanson distinguished 15 artists from 15 non-artists based on their eye fixation patterns as early as 1992 [213]. Fixation density on the less significant features of familiar and unfamiliar paintings greatly contributed to the investigation of discrimination.

Marcel Just and Patricia Carpenter provide a model of reading comprehension based on an investigation of college students' eye fixations as they read scientific sections [214]. The model relates to the cognitive processing of words, sentences, and text components. Readers take longer pauses at areas with the highest processing burdens. When readers use uncommon terms, the burden increases. The model describes the duration of gaze on each word of text as a function of the levels of processing involved.

In forensic investigations, suspects will occasionally conceal their recognition of a known individual for their own protection. Millen et al. [215] wanted to test if eye fixations could be utilised to identify the memory of familiar persons when participants were required to recognise faces. The eye-movements of participants were recorded as they lied or spoke the truth about recognising faces of varying familiarity (famous or personally known people). In conclusion, this research supports the notion that several fixation metrics reveal memory cues during the recognition of falsehoods. When participants falsely denied recognition of personally recognised faces, fewer fixations were recorded than when they correctly denied unfamiliar faces.

Eye area fixation may be an efficient indicator of a person's level of empathy. Moutinho et al. explored the hypothesis that this sort of empathy measure may not be suitable for those with high levels of social anxiety because attentional biases of avoidance or hypervigilance toward emotional faces are common in this condition [216]. Using eye-tracking, fixation times on the eye region were examined in subjects with low versus high levels of social anxiety, and these measurements were connected with empathy levels. For subjects with considerable social anxiety, the link between empathy and fixation time was nil.

It is usual in reading research to consider eye-fixation behaviour in order to examine underlying cognitive processes. In numerical cognition research, however, eye-tracking is utilised less frequently and less methodically. In numerical cognition research, the behaviour of eye fixations ranges from investigating the basic perceptual aspects of non-symbolic and symbolic number processing, to evaluating the common representational space of numbers and space, to assessing the influence of the positional value in base 10 of Arabic numerals, to exercising executive control over number processing, according to a review of the relevant literature. In addition to fundamental discoveries such as the correlation between the size of a number and the time required to read it, research has indicated that number processing can influence general domain activities such as shifting attention, but also in the opposite direction. It has been discovered that broad domain processes, including cognitive control, influence number processing [217]. In conclusion, eye fixation behaviour provides fresh insights into both domain-specific and generic number processing mechanisms.

Sekiguchi [218] investigated the association between individual variability in face memory and eye fixation patterns when viewing faces. Participants were split into high and low memory groups based on their performance on a recognition memory test after viewing short films of 20 faces. The distribution of eye fixation did not differ qualitatively between the high and low groups. Both groups were preoccupied with the inner regions of faces, especially the eyes. The high groups moved their eyes more frequently than the low groups, revealing a difference in the pattern of eye fixation between the groups as measured by the number of fixations and total fixation time. These findings indicate that eye fixation plays a functional role in face memory.

5.4 PUPILS

The pupil is a circular opening in the center of the iris that permits light to reach the retina. It is placed around 3 mm from the cornea's (the transparent layer that covers the entire front of the eye) apex, which shields it. This little hole in the center of the iris appears black to the observer because most of the light that passes through the cornea and lens is absorbed by the tissues within the eye. However, a small quantity of light manages to reflect and, in certain circumstances, makes the pupil appear "bright."

Under normal conditions, the size of the pupil can range from 2 to 5 mm and is not fixed but rather variable, as it is controlled directly by the iris, which, through the pupil, is able to modulate the amount of light entering the eye. In dim light, the pupil tends to dilate to capture as much light as possible. In extremely strong light, however, its constriction protects the photosensitive cells of the retina from injury.

Figure 5.2: From left to right is shown a pupil in the phase of miosis, in normal conditions, and in the phase of mydriasis.

The sole purpose of the pupil, which is purely optical, is to regulate the quantity of light that reaches the retina in order to produce clear images under all circumstances. This amount is proportionate to the pupil's area, or its diameter squared. The width of the pupil is controlled by the antagonistic actions of the constrictor and dilator muscles found in the iris. Its role is to maintain continuous retinal illumination by adjusting its diameter in accordance with the amount of light falling on the retina. The constrictor muscles contract in order to constrict the pupil, while the dilator muscles contract in order to dilate it. The constrictor muscle's cells are grouped in concentric rings around the pupil, whereas the dilator muscle is radially organised and innervated from the orthosympathetic. When the retina is stimulated by shining light into the eye, the pupil instinctively constricts (*miosis*), with the degree of constriction being proportional to the intensity of the light and the size of the illuminated retinal surface. The pupil instinctively dilates (*mydriasis*) as the environment changes from bright to dark or when the intensity of light decreases (Figure 5.2).

The term "mydriasis" is derived from the Greek word "amadros," which means "dark," and refers to pupil dilatation. Physiologically, the pupillary aperture expands momentarily for the eye's adaptation to darkness as well as during painful sensations and emotional mental arousals such as worry, surprise, and panic. Additionally, this may be suggestive of specific clinical disorders. Pathologic mydriasis, for instance, occurs seconds after a heart attack and can persist for hours after blood circulation has been restored. Acute glaucoma also produces pupil dilatation and a lack of pupillary response in the affected eye; this pathologic occurrence is an ophthalmologic emergency requiring immediate medical attention. Mydriasis is an indication that the third cranial nerve (oculomotor) is being compressed due to a stroke. Pupillary dilatation is also reported in the presence of ocular injuries, brain trauma, viral conditions, and toxic states. In addition, antihistamines, barbiturates, estrogens, antidepressants, etc. can induce it. In order for the ophthalmologist to examine the ocular fundus, atropine and other mydriatic medications, such as tropicamide and cyclopentolate, are administered to enlarge the pupils.

Miosis is derived from the Greek term meiosis, which means "reduction." It indicates that the diameter of the pupil is decreasing. Pupillary constriction happens physically during near vision or in reaction to very bright light stimulation, but it can also be observed when the eye accommodates for near vision, when the eyeballs converge inward, and during deep sleep. It is also observed in certain clinical situations, such as iridocyclitis, uveitis, corneal foreign bodies, and ocular or orbital trauma. Pupil constriction may potentially suggest the existence of a brain hemorrhage, encephalitis, or other neuropathological conditions. Drugs such as pilocarpine, timolol, and reserpine can induce miosis. Sometimes, punctiform pupils can indicate drug or chemical poisoning, such as that caused by heroin, codeine, or morphine.

From an anatomical point of view, a normal, nonpathological eye is known as an emmetropic eye and is, in general, studied very little compared with myopic and hyperopic eyes. Normal pupil size in adults ranges from 2 to 4 mm in diameter in bright light and from 4 to 8 mm in darkness. Results in the literature show that healthy emmetropic women have a larger pupil diameter than men.

The maximum pupil size varies significantly among different age groups: it changes as a function of an individual's age. Particularly at birth, pupil size is small (i.e., under 3 mm) and gradually increases during the first few years of life, presumably until puberty. With advancing age, the size gradually decreases [219], [220], [221]. Not only does size decrease, but also the speed of pupillary responses to a light stimulus decreases [222], [223]

[224]. With age, the muscles that control pupil size and reaction to light lose strength. This makes the pupil smaller and less sensitive to changes in the amount of light around it.

5.4.1 *Pupillary responses to emotional-cognitive stimuli*

It is commonly believed that the eyes are the window to the soul, but it is well-established that they are actually the window to the brain. Our pupils do much more than simply respond to light. The participation of the sympathetic nervous system physiologically explains how and why pupillary fluctuations occur. Clarifying why people have different pupillary reflexes in response to the same emotional stimulus or mental activities is more complex. The involvement of the sympathetic nervous system explains physiologically why and how pupillary variations occur. It is challenging to explain why people have distinct pupillary responses in response to the same emotional stimulus or mental activity.

Pupillometry, or the investigation of the psychological origins of pupillary reactions to external stimuli, dates back to an early pilot research conducted by Hess and Poss (1965 [225]). The two discovered a relationship between the emotional value of a visual stimulus and pupillary changes: when displaying different images to a group of subjects, men exhibited a greater pupillary response to images depicting female nudity, while women exhibited a greater pupillary dilation in response to images depicting male nudity, a child, or a mother with a child. This pupillary dilatation was regarded by Hess and Poss as a sign of attention in the visual stimuli. Other studies later report the same outcome (Qu & Guo [226]).

Regarding pupillary responses to various types of emotional stimuli, however, the literature has a number of conflicting findings. In further studies, Hess and Poss observed the expected pupil dilation in response to positively valued images as well as miosis in response to negatively rated images. This outcome was also discovered in subsequent studies. Ultimately, it appears that emotional and psychological reactions to external stimuli are mirrored by fluctuations in pupillary size. Even though each subject's response to such stimuli is unique, there are still patterns that can be seen based on the subject's demographics. As a result, pupillary responses to external stimuli should always be analyzed with demographic variables such as the subject's gender and age in mind.

6 STATE-OF-THE-ART

The study of periocular characteristics is an important source of information. There are various sectors that exploit its potential, from clinical and experimental psychology to the fields of security and biometrics. In 2009, Park et al. [227] conducted the earliest studies in periocular biometrics.

6.1 RECOGNITION SYSTEM

Periocular recognition has emerged as a particularly interesting area in difficult real-world situations, such as when there are uncooperative subjects, the presence of occlusions, and other similar situations, and yet it has demonstrated reliable performance. In recent years, deep learning techniques have been favourably embraced in computer vision and pattern recognition. Several studies have utilised the advantages of deep learning or convolutional neural networks to improve periocular biometric identification rates. Periocular area is represented by the sub-portion of the face near an eye with eyebrows, eyelids, and eye folds. In [228], the authors provided a framework for implementing a periocular recognition system that was robust to variations in image capture distances since the periocular region has recently emerged as a promising biometric modality for unconstrained human authentication. In [229], the authors offered a method that examined the dynamics of the entire face using a geometric descriptor organised in time series, with a special emphasis on the periocular region, and achieved interesting results with high identification accuracy.

Despite the fact that numerous studies in the literature use the periocular area as a sub-part of the face, there are relatively few that extract its soft properties in terms of pupil, fixations, movements, and blinks. This is rather inconsistent with the available scientific information on these features, both in terms of the cognitive domain and their potential to contribute to the recognition of a subject or group of people. For instance, there is widespread agreement that eye movement data contains features that can be utilised to uniquely identify individuals. One of the most popular solutions is to extract or model the characteristics of the oculomotor system that controls eye-movements [230].

Therefore, these soft periocular biometries are underutilised and understudied, especially for biometric reasons. One of the potential causes could be the widespread belief that in order to obtain exact and reliable data, specialised hardware such as high-end eye trackers with a high sample rate is required. However, in our work [231], we demonstrated that it was possible to investigate and derive information on these characteristics even using data collected without the use of specialised technology.

The eye tracker is a device that must be calibrated for each individual user before each experiment; this could be regarded as a limitation. In [232], the authors investigated the feasibility of identifying individuals based on the calibration input they provide to an eye tracker. Given the outputs of an uncalibrated eye tracker compared to genuine gaze points, the mistakes will be reproducible for the same individual; this was the focus of this research. They analyzed the data from three datasets, with an accuracy range of 49% to 71%.

Instead, a neuromorphic vision sensor with microsecond-level temporal resolution was utilised in the [233] research. Existing biometric identification methods based on explicit and static properties are susceptible to impersonation attacks. For these reasons, the authors presented a biometric authentication system based on transitory blink signals. The neuromorphic vision sensor only sends blink-induced local pixel-level changes as they occur, resulting in advantageous characteristics such as ultra-low response latency. Based on the microsecond temporal resolution of the event density, the examined information is a collection of effective biometric parameters describing the motion, speed, energy, and blink frequency signal. These were inputs for two distinct models. The trials demonstrated a high degree of accuracy in identifying and verifying the subjects' identities.

Eye-movements exhibit many of the favorable attributes frequently encountered in other behavioral approaches to biometrics, such as fundamental resistance to replication and the ability to discretely record the sample. [234] was a pioneering work that presented a study on the potential of eye-movement data for biometric purposes. The eye-movements of twelve participants were measured while standing still and viewing moving objects. The measured data includes pupil sizes and their dynamics, gaze speed, and the distances of the infrared reflections of the eyes. The best dynamic feature was found to be pupil size.

Using approaches from the field of speech recognition, [235] proposed a method for evaluating eye movement signals for person authentication in a visual task-independent scenario. In [236], subjects viewed images of faces while their eye-movements were monitored, providing insight into each participant's points of attention. The authors proposed a graph-based framework. It treated eye trajectories as 2D distributions of points on the image plane. The focus was on the identification task.

Work [237] also includes an objective evaluation of numerous biometric parameters based on eye-movements and their capacity to properly and precisely distinguish unique persons. Considered biometric candidates were associated with a number of basic eye-movements and their aggregate scanpath characteristics (acquired during the reading), including the number of fixations, average fixation duration, average saccade amplitude, average saccade velocity, average peak saccade velocity, velocity waveform, scanpath length, scanpath area, regions of interest, scanpath inflections, amplitude-duration relationship, main sequence relationship, and pairwise distance between fixations. To combine these metrics into a single identification algorithm, a fusion method was presented. With limited testing, a not particularly competitive error rate of 27% was achieved. Also in [238] a type of reading eye-movement biometric recognition technology has been proposed. The authors built a Reading Eye-Movement

Recognition computational model based on a multi-input deep neural network and identified human subjects by comparing predicted and actual fixation sequences. The experimental results showed that the fixation sequence similarity recognition algorithm obtained an equal error rate of 19.4% on the test set, and the model obtained an 86.5% Rank-1 recognition rate on the test set.

Cantoni et al. [239] presented a work on gaze analysis in which the eye movement model was constructed using fixation time and regression counts on different gaze points of different subjects, and the similarity between the two records was determined using the Frobenius norm of the density map. This article described a technique that made use of a graphical depiction of the fixation sites determined by an eye tracker during human-computer interaction. The primary purpose was to prove the hypothesis that the way an individual perceived an image could be a personal distinguishing feature, i.e., a soft biometric application.

Since several works in the field, especially in the past, have studied periocular features of the spectral domain or temporal dynamics (speed, acceleration), little research has been conducted using spatial features. For these reasons, [240] proposed a biometric recognition system based on the fixation density map. The idea was to develop a method capable of representing visual scanning in the stimulus plane and extracting idiosyncratic features for biometric identification of subjects under free-vision conditions.

6.2 DEMOGRAPHIC CLASSIFICATION

Demography is the study of the dynamics of populations. It involves the study of the number, structure, and distribution of populations, as well as the evolution of populations over time as a result of births, deaths, migrations, and ageing. The scope of a demographic analysis can range from entire societies to smaller subgroups based on factors such as education, religion, ethnicity, age, and gender.

Applications of biometric demographic analysis are fairly diverse. Often, advertising is targeted. If the knowledge of clients (such as age and gender) can be automatically assessed from their faces or voices, customised products and services can be suggested. Human-computer interface is another frequently studied application where automatic demographic analysis might enhance social competence in interaction.

In the context of demographic classification by age and gender, our eyes can also be a valuable source of data.

- ***Blinking.*** *Spontaneous blinks.* Gender and age affect blinking. Sun et al. demonstrated in their study that the mean amplitude and peak velocity of spontaneous blinks decrease with age [241]. This reduction could be partially ascribed to a peripheral phenomena, narrowing of the palpebral fissure width. Age also decreased the spontaneous blink-down phase main sequence slope. In contrast, neither the blink rate nor the synchronisation of eyelid movements (blink conjugacy) changed.

  Voluntary blinks. The gender-related differences in voluntary blinks were also highlighted in [242]'s studies. Fifty healthy volunteers—10 women and 15 men under 40 years of age (range: 22–38 years) and 20 women and 5 men above 60 years of age (range: 63–85 years)—were examined. Women tended to make deeper and faster voluntary blinks. Age-based differences were highlighted too. Voluntary blinks in the younger subjects were more frequent than in the older subjects. Additionally compatible with [241]'s findings. The authors of this research studied the eyelid kinematics of people with normal ageing in order to test the hypothesis that eyelid movements undergo age-related changes and that the blink abnormalities widespread among the elderly are the result of normal ageing processes. The researchers demonstrated that the average amplitude and peak blink rate of spontaneous and, to a lesser extent, voluntary blinks diminish with age. In [242], the authors discovered that the average amplitude and duration time of spontaneous blinks were greater in younger subjects than in older subjects. There was no difference in the frequency of spontaneous blinking between younger and older participants. In both age categories, women blink more frequently than men.

Reflex blinking. It is generally recognized that reflex action is impaired with increasing age and so influenced by the age of the sample [243]; in particular, complex reflexes tend to have longer delays [244]. In [245] the authors suggested that age and gender variation should be taken into consideration in the interpretation of the brainstem reflexes in basic and clinical studies. Taken in consideration healthy subjects, the results represented true age-related changes: the amplitude measures of the blink reflex component were lower in older than in younger and lower in females than in males.

- ***Eye-movement.*** Male and female brains have varied architecture, which may result in distinct eye movement patterns between the gender. Based on this finding in [246], the authors provided the findings of an eye tracking experiment in which they analysed the eye-movements of 25 male and 20 female volunteers while passively viewing images. In order to analyse gender variations in eye movement patterns during image viewing, several characteristics, such as scanpath length, number of saccades, spatial density, saccade breadth, etc., were evaluated. Eye movement patterns were shown to differ significantly by gender. In fact, an eye-movement study revealed that females exhibit more exploratory gaze behaviour, as evidenced by greater saccade amplitudes and longer scanning routes. Additionally, it is theorised that females inspect visuals more rapidly than males. In [247], saccadic eye-movements were measured in 168 normal human volunteers aged 5 to 79 years to determine any age-related alterations. Under varied fixation conditions, subjects were told to either stare toward (a pro-saccade task) or away from (an anti-saccade task) an eccentric object. Young children (5-8 years old) were found to have slow saccadic reaction times and a large intra-subject variance; young adults (20-30 years old) had the fastest saccadic reaction times and the lowest intra-subject variance; and finally, elderly subjects (60-79 years old) had slower saccadic reaction times and longer duration saccades than the other groups. These findings reveal very substantial age-related impacts on the performance of the participants, which may reflect various stages of normal growth and degeneration of the neurological system.

- ***Fixations.*** Fixations were investigated in the study [226] using 18 product photos as stimuli. This study examines the relationship between eye-movements and the emotional response of the user to the qualities of a certain product. The stimuli included harsh, neutral, and pleasurable visuals. Forty participants (20 males and 20 females, mean age = 35.6, SD = 6.38, range: 21-48 years) participated in an eye-tracking study in which they viewed randomly displayed product photographs. Participants' eye-movements while viewing product graphics were measured. Participants quickly rated their emotional response to the product photographs on a seven-point scale after viewing the stimulus. The results demonstrated that stimulus category and gender differences resulted in distinct changes in fixation number and duration. In terms of gender differences, the results revealed that females had higher scores for fixation count and duration. Age effects have not been fully explored, but the overall pattern suggests that differences in fixation durations and fixation probabilities are quantitative rather than qualitative [248].

- ***Pupils.*** Qu and Guo indicate that the pupillary size of women in response to negative stimuli is significantly smaller than that of men in response to the same type of stimulus. However, numerous other studies [249], [250], [251], [252], [253] have indicated that the pupil dilates in response to both positive and negative valence images, whereas neutral stimuli have no effect. In particular, Partala and Surakka (2003 [251]) observed that positive stimuli generated more pupillary dilation in women than in men, whereas negative stimuli caused larger dilation in males. Although the finding appears similar to that of Qu and Guo's study, a lower pupillary response to negative stimuli in women cannot be attributed to pupil constriction, as a smaller response to neutral stimuli was seen. In contrast, Yrttiaho et al. [253] conducted a study in which women reported increased pupillary dilatation in reaction to negative visual stimuli depicting children in apparent emotional distress. Current thought correlates pupil dilation more with the subject's emotional arousal than with the stimulus's positive or negative quality. Specifically, an individual's emotional response to an external stimulus depends heavily on whether or not the stimulus is perceived as stimulating. Age, of course, also influences the effects of the aforementioned emotional and cognitive factors: in particular, it has been shown that the combined effect of emotional involvement and the advanced age of a subject produces less pupillary dilation than the same level of arousal in a younger subject [254].

7 OUR CONTRIBUTION TO LITERATURE

Fixations, pupils, eye-movements, and blinks are periocular biometric characteristics that have been demonstrated in the literature to be significant biomarkers for detecting cognitive effort and emotional responses, as well as for getting useful information for subject recognition. Despite the fact that this relationship is well-established in the literature, it was noted that there was a paucity or absence of study into constructing a system that would learn the patterns of this data and then draw conclusions.

For this reason, we first investigated the use of the pupil as a single soft biometric trait for the purposes of demographic classification, particularly by age and gender, providing a critical comparison and extensive discussion of which ML technique was most appropriate to the problem among a large pool of classifiers and beyond. Subsequently, with the same purposes of demographic classification, this biometric was also included and evaluated in a broader context of fusion with two other biometric traits, fixations and blinks. The performance achieved encourages studies of the periocular area as a soft biometric to be detected when the lower part of the face is not visible. For the development of a web user identification model, features extracted from the periocular area related to pupils, blinks, and fixations were analyzed together with another behavioral biometric data source, touch dynamics. The results obtained demonstrated the promise of these two different biometric traits and, more importantly, their fusion. Finally, with the aim of analyzing changes in periocular features in order to classify various cognitive processes, two studies were conducted. In the first, a system capable of measuring audience attention was implemented based on blinks and fixations, among other features. In the second, however, to investigate the effects on the periocular region of visual stimuli that stimulate two different mental tasks: visual memory recall and understanding the semantic complexity of an image, information from blinks, pupils and gaze movements were extracted and analyzed.

Additional information on these studies is available below (Sections 7.1, 7.2, 7.3).

7.1 DEMOGRAPHIC CLASSIFICATION

In addition to referring to the measurable statistics of a population, the term "demography" also refers to qualities such as age, gender, ethnicity, and race that are commonly employed in population statistics. Since early papers in [255] and [256], the computer vision community has supported research on this class of soft biometrics.

As described in the preceding sections, the biometric properties of the eyes demonstrated a degree of discrimination for demographic classification purposes. Specifically, it was demonstrated that these characteristics could provide useful information regarding gender and age differences. Soft periocular features are of particular interest for this type of purpose both when considered physical (e.g., at the level of pupillary base size) and behavioral (blink frequency, scanpath, etc.) characteristics by reacting differently to stimuli depending on the category (male/female, young/adult/old). Despite this evidence, the literature on demographic classification work based on these parameters and also jointly using powerful ML principles has been relatively limited or nonexistent.

Because of this, we first did a comparison study to find out which periocular biometric trait seemed to be a good starting point for this kind of classification [257]. Specifically, taking into consideration the results obtained from a previous study [258] in which gaze behavior data were analyzed in terms of fixation and scanpath, replicating the same experimental conditions, we compared pupil indicators with the results obtained and observed how the mean pupil diameter has proved to be the best discriminating feature for both gender (male and female) and age (under 30 and over 30).

Based on these results, therefore, as detailed in more detail below, we decided to do a broad and deep analysis using the most famous and popular ML techniques to analyze pupillary size as a single feature, which was then fused and combined with information from fixations and blinks to obtain a gender (male and female) and age (under 30 and over 30) classification.

Table 7.1: Age range of participants in GANT dataset.

Age range	**# participants**
a (17–18)	11
b (21–30)	57
c (31–40)	9
d (41–50)	16
e (51–60)	8
f (61–70)	9
g (71–80)	1

The dataset used for these studies is GANT (Gaze ANalysis Technique), which is publicly available online and involves 112 subjects [239]. The participants are 72 males and 39 females with an age range between 17 and 80 years (see Table 7.1). It defines seven age groups: a (17–18), b (21–30), c (31–40), d (41–50), e (51–60), f (61–70), and g (71–80). The acquisitions have been made through the Tobii 1750 eye tracker (1280×1024 screen resolution, 50 Hz sampling frequency). Participants were shown 18 images as experimental stimuli (examples in Figure 7.1), 16 human face images (s1-s16), and 2 landscape ones (s17 and s18), interleaved with blank white screens with a cross at the center. All images were displayed for 10 s, while the empty ones 3 s, except the first one which was displayed for 5 s. The order in which all images were shown is random. The 16 images of the human face are of eight males and eight females. Among these images, 4 images of females and 4 ones of males are of famous people, while the others are of persons unfamiliar for the participants. The gray-level distributions are similar. The acquisition sessions are two: one was made in 2012, and the other in 2013. In both, there

Figure 7.1: Some examples of images shown to participants during the data acquisition process in the GANT dataset. The first column shows images of two landscapes. The last two columns, on the other hand, show images of women and men: in the first row there are images of unknown people while in the second row there are images of two famous actors.

are subjects that are acquired several times for a maximum of 3 per year. Each acquisition is carried out at a temporal distance from the previous one, in a time interval between five and nine days. The first session of 2012 involves 88 participants, 36 of whom were also involved in a second session and 16 in a third. A total of 140 acquisitions were made. In the second session of 2013, instead, 34 subjects were involved, 10 of whom also participated in the first session of 2012. 17 subjects participated in a second session (9 of which participated in 2012), and 13 out of 17 (6 of which participated in 2012) also performed the last acquisition session. In total, 64 acquisitions were made in the three sessions over a period of time ranging from a minimum of one day to a maximum of 21 days. In this paper have been taken into consideration all set of tests for gender and age classification. The acquisitions in the GANT dataset have been carried out by the Tobii 1750 eye tracker. This model assigns a validity code (0–4) to the acquisitions relating to the right and left eye, respectively. The lower the code, the more accurate the extraction of eye data is. In the first phase, we filtered out the data whose code was different from zero. If the data of one eye has been associated with a high confidence code, for example 0, and the data of the other eye has been associated with a low confidence code, for example 4, both have been deleted. As a result, only data deemed relevant for both eyes was saved.

7.1.1 *Pupil size for Age and Gender Classification*

The pupil size is a soft biometric trait, but in-depth studies to analyze it for biometric purposes are lacking in the literature, as are datasets focused on this field of research. On the basis of these observations, we presented in [11] an extensive study with the objective of demonstrating that pupil size and dilation over time can be potentially used to classify people by age and by gender. To do this, 14 supervised classifiers were applied to a dataset meant for gaze analysis. Measuring the right and left pupil individually and also simultaneously, the performances of the classifiers have been compared and the worst and best performing selected to support potential fusion strategies.

Given a training set of N instances $X = \{X_1, X_2, ..., X_N\}$ and a set of t possible classes $C = \{c_i, ...c_t\}$ associated to X, the corresponding class labels are $y_1, y_2, ..., y_N$ where $y_i \in C, (i = 1, ...N)$. The aim was to predict the correct class label for a new set of points X^*. In this work, we mainly focused on the following groups of ML algorithms:

- ***Decision Tree.*** Decision Tree Classifier (DTC) is based on a tree structure where the inner nodes represent the features, the leaves the outcomes and the branches the decision rules.

- ***Ensemble methods.*** The goal of ensemble methods is, instead, to combine the responses of several learning estimators in order to improve the predictive performance and robustness of a single algorithm. These kinds of methods can be divided into two families: averaging and boosting methods. The former are Random Forests (RF) and Bagging (BG) classifiers. The basic idea is to independently calculate the average of the predictions of different estimators. This ensures that the variance of a single base estimator is reduced. The second group of methods includes AdaBoost (ADA) and Gradient Boosting (GB) classifiers. The driving principle is to create a strong classifier from a combination of weak ones. The base estimators are used sequentially to correct the errors from the previous model.

- ***Instance-based learning strategy.*** Instance-based learning strategy is a decision-making problem based on instances seen in the training phase that are deemed important or representative of the problem. These algorithms, called winner-take-all methods, generate a database of sample data stored in memory and compare the new data for which you want to obtain a prediction through a similarity measure in order to find the best match. Among the most famous and widely used methods belonging to this class are K-Nearest Neighbor (KNN) and Support Vector Machines (SVM).

- ***Artificial neural networks.*** Artificial neural networks are models whose structure or function is inspired by biological neural networks. They are a type of pattern matching that is often used for regression and classification issues, but they are actually a vast subfield with hundreds of methods and variants for all sorts of problem types. The two most famous algorithms for artificial neural networks are Stochastic Gradient Descent Classifier (SGD) and Multilayer Perception (MLP). SGD is an optimization algorithm that implements a simple learning routine for the descent of the stochastic gradient. MLP is based on one or more layers of perceptrons. The former receives the input, and the latter expects the class. Among these, there may be an arbitrary number of hidden nodes that allow you to model the correlation between input and output. During the training phase, the various parameters are adjusted with a backpropagation approach with the aim of minimizing the error.

- ***Probabilistic methods.*** Bayesian methods are those that explicitly apply Bayes' Theorem for problems such as classification and regression. Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Gaussian Naive Bayes (GNB) are special instances of the Bayes classifier;

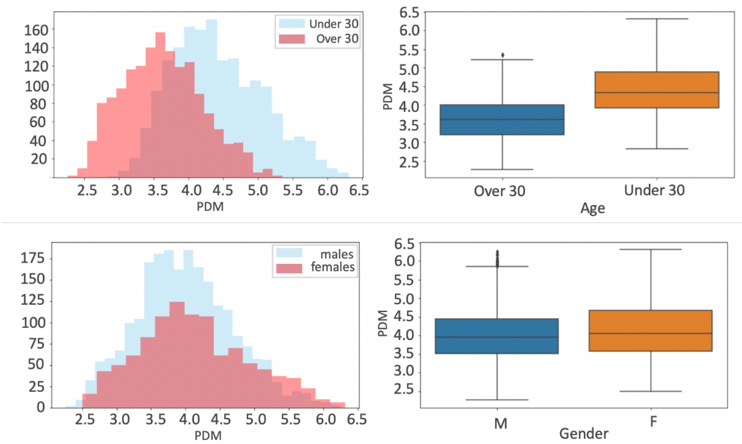


Figure 7.2: (left) Distribution of PDM for under 30 and over 30. (right) Distribution of PDM for males and females.

they all deal with continuous Gaussian predictors, and their assumptions on the relationships among predictors and across classes vary (i.e., the way they specify the covariance matrices). Gaussian Processes (GP) is a generalization of the Gaussian probability distribution and can be used as the basis for sophisticated non-parametric ML algorithms for classification and regression.

Table 2 of the [11] document provides more details on the above classifiers, together with the parameters used to set up an exhaustive search.

The pupil diameter mean (PDM) was calculated by exploiting the data for each image and for each eye of the GANT dataset (details in Section 7.1). Figure 7.2(top) shows the distribution of PDM for the participants belonging to the under-30 and over-30 classes. It can be observed that the distributions are such that it is feasible to properly separate the acquisitions into two classes. On the bottom of the Figure 7.2 it is also evident how much the distribution of males and females overlaps each other. This, in turn, suggested that a binary classification of gender would be more difficult to achieve. These average values were then appropriately labeled with respect to both gender and age. The analyzed samples are organized as follows:

- average left pupil size (PL)
- average right pupil size (PR)
- average left and right pupil size (PLR)
- average left and right pupil size + code of the relative image (PLRI)

Then, the feature vectors were scaled into a more appropriate representation for estimators that maintains the original distribution.

PLRI has not been analyzed for both classification activities. In fact, it is known in literature that the content of the images could influence the pupillary response of men and women in different ways. Therefore, for a complete analysis, we decided to integrate this information content into our analysis just for gender classification. This contribution was not relevant to the age classification.

The experiments have been performed on an empirical splitting of the dataset, consisting in a portion of the data for the train-set and the remaining one composing the test-set. More combinations have been considered. From the observation of the boxplots in Figure 7.2 it was evident that setting a simple threshold to separate both categories was not possible. Indeed, as regards the gender classification, it was not possible to define a separation line that allowed us to profitably divide the values of the two categories under analysis (males and females). Instead, for the age classification, it seemed possible to define a threshold. However, looking carefully at the distribution of the data in the various quartiles, we realized that, by setting a hypothetical threshold, more than 25% of subjects would have been classified incorrectly. The good separation of the two classes were such to require just a little amount of samples to build the discriminating model.

Table 7.2: Comparative analysis among the classifiers in terms of age classification. The results are presented in terms of the best configuration for each classifier achieved through an exhaustive search of its hyperparameters. In bold, the best-performing classifiers are highlighted, while the worst are in italics (underlined).

	PUPIL - LEFT PL				PUPIL - RIGHT PR				PUPIL - BOTH PLR			
CLASSIFIER	ACCUR.	PRECIS.	RECALL	F1	ACCUR.	PRECIS.	RECALL	F1	ACCUR.	PRECIS.	RECALL	F1
MLP	0.8018	0.7277	0.9145	0.8105	**0.8369**	**0.8362**	*0.8059*	0.8208	**0.8277**	**0.7616**	*0.9145*	**0.8311**
DTC	0.7988	0.756	*0.8355*	*0.7938*	0.8079	0.7214	0.9539	0.8215	0.8079	0.7214	0.9539	0.8215
GNB	*0.7957*	0.7361	0.8717	0.7982	0.7851	0.69	0.9737	0.8076	0.8049	0.728	0.9243	0.8145
ADA	0.8018	0.7364	0.8914	0.8065	0.7851	0.69	0.9737	0.8076	0.8034	0.7238	0.9309	0.8144
BG	0.8018	**0.7605**	*0.8355*	0.7962	0.7851	0.69	0.9737	0.8076	0.8003	0.7125	0.9539	0.8158
QDA	0.8018	0.7266	0.9178	0.811	0.7851	0.69	0.9737	0.8076	0.7927	0.7069	0.9441	0.8085
LDA	0.7988	0.7287	0.9013	0.8059	0.7759	0.6796	0.977	0.8017	0.7912	0.7072	0.9375	0.8062
LinSVC	**0.8034**	0.7285	0.9178	**0.8122**	0.779	0.6828	0.977	0.8038	0.7881	0.696	0.9638	0.8083
KNN	0.7988	0.7228	0.9178	0.8087	*0.7652*	*0.6667*	**0.9868**	*0.7958*	0.7851	0.6945	0.9572	0.805
GP	**0.8034**	0.7285	0.9178	**0.8122**	0.7698	0.6734	0.977	0.7973	0.7851	0.6909	0.9704	0.8071
SGD	0.8003	0.7294	0.9046	0.8076	0.8201	0.7397	0.9441	**0.8295**	0.7835	0.6893	0.9704	0.806
SVM	0.8018	0.7266	0.9178	0.811	0.7805	0.6852	0.9737	0.8043	0.7713	0.6742	**0.9803**	0.7989
GB	0.8003	0.7224	0.9243	0.811	*0.7683*	0.6743	0.9671	*0.7946*	0.7652	*0.6682*	**0.9803**	0.7947
RF	*0.7957*	*0.7146*	**0.9309**	0.8086	0.8064	0.7185	0.9572	0.8209	*0.7591*	*0.6629*	0.977	*0.7899*

The experimental design has been mainly organised in three stages:

- ***Dataset split by acquisitions.*** The first round of experimentation was performed on the dataset split into train:test proportions of 80:20, 70:30, and 10:90, respectively, by the acquisitions. The contribution of this experiment was limited to the fact that when training and testing shared different acquisitions from the same subjects, the classifiers separated the samples into classes with the support of biometric recognition of the users (who are seen at training time). In fact, the overall mean accuracy achieved in all these experiments, even when the train test was significantly less populated than the test set, was about 75$\pm$2%. This result was also motivated by the fact that the binary nature of age classification in this study, combined with the simple representation of the data, required a small number of samples at training time to build up the discriminating model. On the other hand, it was more interesting to consider the performance of the classifiers when unseen subjects are acquired. The ability of the classifier to discriminate the subject by age or gender gained more significance.

- ***Dataset split by subjects.*** Splitting the dataset in terms of acquisitions is not a sufficient condition to assure that the testing samples were from subjects that do not belong to the train set too. For this reason, to prevent that samples in the test-set would be used at training time, we split the dataset also by subjects, ensuring that if at least one acquisition of a subject is in the train-set then different acquisitions by the same subject cannot be in the test-set. This partitioning of the dataset has been performed with a proportion of 80:20 between the training and test sets, respectively. By removing the biometric contribution from the dataset splitting, the goal was to effectively demonstrate the capability of classifiers to separate the acquisitions into classes.

Age classification. The results achieved in terms of binary age classification are summarised at Table 7.2. The Table presents the results achieved with all classifiers and techniques separated by observed data. That is, the first part presents the performance for left pupil dilation only, and the middle part for right pupil dilation. The last part of the Table considers the contributions of both pupils. It can be observed that the MLP achieves the best accuracy on the right pupil and when both pupils are considered. But, on the contrary, on left pupil Linear Support Vector Machines (LinSVC) and GP outperform the others. This result helped us to support the observation that pupils dilation did not show the same behaviour in both eyes for all individuals. In some cases, it may happen that personal attitude or poor eyesight can affect the speed and trend of dilation of the pupils in a different way per eye. Therefore, with this consideration in mind, the results reported at Table 7.2 did not suggest that MLP did not succeed with left eyes but, rather, that the variability of dynamics in the dilation of pupils in the dataset was such as to determine a (very limited) decrease in performance. In fact, it can be observed that MLP is also the most precise on average over the experimental setting but, conversely, reports the lowest recall rate. Moving the attention on the more compact result provided by F1-score, it can be observed that performances exhibited by the classifiers varied significantly from each other. But, we could again notice that MLP was overall the one that achieved the more stable result among the treatments. In addition, even though the boxplot at Figure 7.2 suggested a good separations in terms of age, linear approaches like LinSVC, SVM with linear kernels or LDA did not provide in general a good classification.

Gender classification. Concerning gender classification, the performances were quite less promising. Such a result was anticipated by the boxplot in Figure 7.2(bottom) and the linked overlapping among the distributions of the two classes. In fact, if we consider the results summarized at Table 7.3, we can observe that none of the classifiers achieved significantly high levels of accuracy. Table reports the best accuracy achieved, but all classifiers showed comparable levels of performance. This result was motivated by the

Table 7.3: Comparative analysis among the classifiers in terms of gender classification. The results are presented in terms of the best accuracy for each sample achieved through an exhaustive search of their hyperparameters.

	CLASSIF.	ACCUR.	PRECIS.	RECALL	F1
GENDER PR	MLP	0.5156	0.5310	0.2679	0.3561
GENDER PL	MLP	0.4978	0.4989	0.9955	0.6647
GENDER PLR	SGD	**0.5848**	0.6067	0.4821	0.5373
GENDER PLRI	KNN	0.5379	0.5281	0.7143	0.6072

fact that, being the pupil dynamics a behavioural biometric trait, aimed acquisitions are necessary to provide those stimuli that allow to stress the different behavioural patterns among males and females. The experiments PLRI, i.e., the one which considers a specific image provided during the acquisition session of GANT dataset, did not result such to extract this kind of contribution from the data.

- ***Score-level fusion strategies.*** To also take into consideration the computational demand of the classifiers, we performed some tests to verify if score-level strategies of fusion could improve the overall performance achieved. Not surprisingly, MLP with a high number of nodes and layers can be significantly more time-consuming than linear classifiers or simpler classifiers. These last ones, on the contrary, achieved lower performances in terms of accuracy, but they have a very reduced training time as well as the ability to run in real-time. On these premises, the third round of experiments focused attention on a fusion strategy that could get the best from the classifiers, thus achieving a higher overall classification rate. To perform the fusion at score-level, the behaviour of each classifier has been carefully inspected. The goal was to look for a combination of classifiers whose estimates were correct when aggregated. To achieve such a goal, a fusion strategy based on the weighted sum (see equation 7.2) of the classifiers' responses has been implemented. Proportionally to best performing algorithms, let

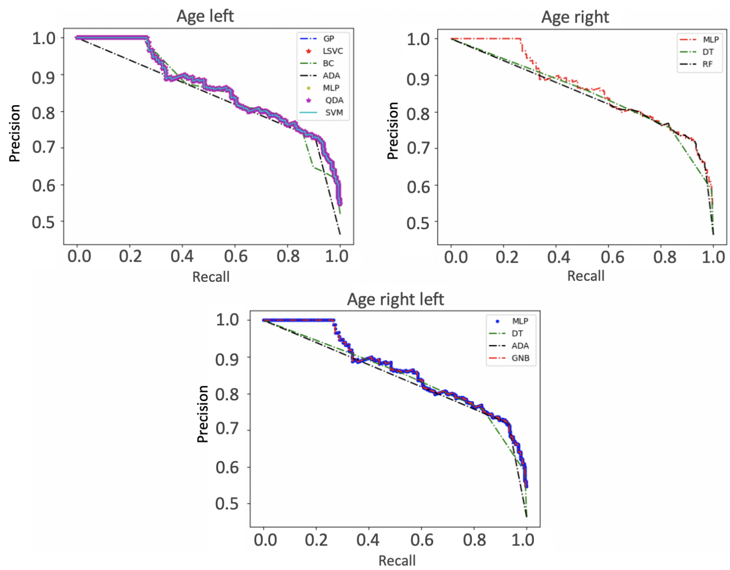


Figure 7.3: The precision and recall measurements for the top three best accuracy values.

m be the recognition method of a fusion of M classifiers and a_m its accuracy, the weight w_m was calculated as:

$$w_m = \frac{a_m}{\sum_{m=1}^{M} a_m} \tag{7.1}$$

where $0 \leq w_m \leq 1$ and $\Sigma_{m=1}^{M} w_m = 1$. Once the weights have been computed, the weighting fused score f was obtained as:

$$f = \sum_{m=1}^{M} w_m s_m \tag{7.2}$$

where s_m is original the score of $m - th$ classifier.

Different combinations of classifiers have been considered, according to the level of performance achieved in terms of accuracy. In particular, they consisted of four different fusions of classifiers: (1) the best classifiers, (2) the worst classifiers, (3) the worst and the best classifiers, and (4) the least computationally intensive classifiers (QDA - LDA - GNB). The first two combinations, the best and the worst

Table 7.4: The classifiers are combined according to the achieved accuracy.

	Worst	Best	Worst + Best	QDA - LDA GNB
AGE PL	0.7957	0.8003	0.8003	0.7988
AGE PR	0.7698	0.8079	0.8079	0.7835
AGE PLR	0.7393	0.8003	0.7973	0.7988
GENDER PL	0.4531	0.4353	0.4643	0.4710
GENDER PR	0.4420	0.4754	0.4576	0.4844
GENDER PLR	0.4442	0.4688	0.4777	0.4755
GENDER PLRI	0.4442	0.4799	0.4688	0.4821

classifiers, have been chosen according to the accuracy results achieved in Table 7.2. The best-performing classifiers have also been represented in terms of their curves of recall and precision in Figure 7.3. In the case of equal accuracy, all classifiers have been considered. The results of the combination are shown in Table 7.4. It can be observed that nor of the fusion improved the accuracy achieved by best classifiers taken alone. Consequently, none of the fusions really introduced a benefit deriving from the combination of the responses of the single classifiers, suggesting that the higher computational demand required to compute the fused score was even counterproductive.

7.1.2 *Periocular Data Fusion*

A unimodal biometric system that is based on a single biometric characteristic has several problems and limitations due to a lack of data, the poor quality of the information collected, or, as in the case of soft biometrics, a low discriminatory power. To overcome these issues, a multi-biometric system, i.e., a system that merges different biometric features, can help to improve performance and consolidate the information obtained. Using only the information provided by the periocular area, we investigated how effectively fusion approaches that combine pupils, fixations, and blinks can estimate the gender and age group of users [259]. The implemented data fusion strategy is shown in Figure 7.4. First,

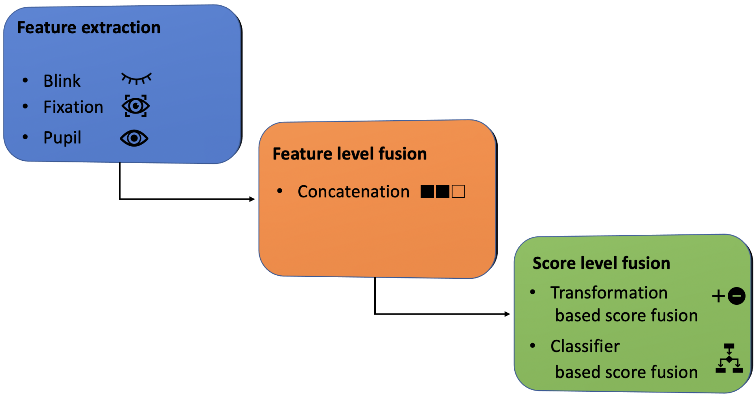


Figure 7.4: The workflow of the proposed fusion strategy.

we adopted a concatenation scheme for fusion at the feature level, while at the score level, we applied and then evaluated the performance of transformation and classifier-based score fusion methods. More details on these approaches are given in Section 1.1. This work employed the following ML algorithms: DTC, RF, BG, ADA, GB, KNN, SVM, and SGD.

According to our knowledge, this was the first paper to combine numerical data derived from the periocular area for the purposes of demographic recognition. For this reason, an ad hoc experiment designed to evaluate the combination of the two investigated fusion techniques was conducted. Extensive tests on the optimal configuration of the classifiers to be selected are also reported. The dataset used is GANT (details in section 7.1).

For each image, the characteristics extracted from the pupil were studied in terms of their diameter measurement, while those relating to fixations and blinks were studied in terms of duration and number. Specifically, for blinks, a counter of "fast" blinks is also increased, i.e., those for which it is not possible to actually calculate the duration. Since the samples obtained were gender-unbalanced, a randomly selected subset was analyzed. Feature vectors obtained were combined through the feature level fusion by a concatenation strategy. This strategy was preferred to a summarizing one that could reduce dimensionality because,

Table 7.5: Spearman's correlation coefficients with respect to the pairs of the three modalities.

	Spearman's correlation coefficients
Blink-pupil	0.1439
Blink-fixation	-0.0992
Fixation-pupil	0.0644

in this case, compared to the three biometric modalities (pupil, fixation, and blink), we had no redundant information. In fact, applying the Spearman non-parametric test between the characteristics of the three modality pairs (blink-pupil, fixation-pupil, and blink-fixation), we observed that there is poor correlation between them. For example, when we studied the pupil-blink pair, we applied this test to all possible pairs of characteristics that could be obtained by comparing these two biometric traits. We reported in Table 7.5 the average of these values for each pair analyzed. So, it was evident how the association between these modalities was very poor. Then the feature vectors were transformed by scaling each feature over a given interval. The rationale behind this choice is that this scaling included robustness to very small standard deviations of features and the retention of zero entries in the sparse data.

After these considerations and operations, the experimentation was conducted by partitioning the available data into random training and test subsets. To avoid strongly tying the results obtained to the training-test random selection data and address the issue of the impact of parameters on the final performance of classification models, we used an exhaustive search. In particular, we tested different parameter values, specific for each estimator, with a k-fold cross-validation procedure with a variable "k" between 2 and 10. The basic idea behind this procedure is to divide the training dataset into k parts and train the model on a subset consisting only of $k-1$. The resulting part was used to validate the model with a performance measure such as accuracy. The measure of the performance obtained is the average of the $k-1$ values derived individually. For each model trained, a confidence score associated with each element of the test set was generated with respect to each label (male/female or under 30/over 30), i.e., the probability that that particular subject in the test belongs to a category rather than to another. As we said, we focused on two approaches: transformation-based score fusion and classifier-based score fusion.

Transformation-based score fusion. For the first strategy, only those classifier scores reporting an accuracy greater than an empirically chosen threshold were selected. The transformation techniques analyzed were:

- weighted sum: see equation 7.2;
- weighted product: the weighed product is obtained from a variant of the classic arithmetic product. Let $(s_1, ..., s_n)$ score vectors obtained from n algorithms, its formula is:

  $$S = \prod_{i=1}^{n} s_i^{w_i}$$

  where w_i is the weight of the algorithm i. It can be observed how, in this case, the scores influence each other more than the weighted sum. For example, if one of the scores is close to 0, the score obtained from the fusion will also be close to 0.

- Bayes fusion rule: Bayes' rule is one of the fundamental pillars in probabilistic theory. The definition of this rule for the event x and y is:

  $$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

  If we have a score matrix M and different classes i to which our observations can belong, the Bayes' rule can be rewritten as [260]:

  $$p(i|M) = \frac{p(M|i)p(i)}{p(M)}$$

where $p(i)$ is marginal probability of i, $p(M|i)$ is the conditional probability, $p(M) = \sum_{i=1}^{n} p(M|i)p(i)$ is the evidence and n is the number of classes. Assuming that the scores are conditionally independent given the classes, $p(M|i)$ can be rewritten as:

$$p(M|i) = \prod_{j=1}^{n} p_j(s_j|i)$$

where $p_j(s_j|i)$ is the score of j-th algorithm related to the i-th class. So, let t_a and t_b the scores obtained by two algorithms for a binary classification problem, their fusion through the Bayes function is given by [261][262]:

$$S = \frac{t_a * t_b}{(1 - t_a)(1 - t_b) + (t_a * t_b)}$$

Also in this case, to give greater importance to one biometric over another, we have assigned a weight (w_i), to be multiplied to the vector of the score which must have a lower value for the final decision. This weight tends to reflect the accuracy achieved by the algorithms involved. If one algorithm tends to have a better performance than another, the prediction of the first should have a greater weight when making a choice.

These have been applied to the scores of the selected classifiers by matching them in all possible combinations. The best weights were calculated with brute-force combinations of weights between 1 and 10 [259].

Classifier-based score fusion. The models chosen are the same ones used to obtain the scores. These classifiers were trained on a new training data set obtained by concatenating the scores relating to the two classifiers that reported the highest accuracy in the first experimentation with concatenation only. The rate train:test chosen was 70:30. Also in this case, an exhaustive research of the best parameters was applied and different k-fold strategies for the models were tested in order to obtain the best results.

The results of our tests showed that using a fusion strategy improved the overall performance of the system. Score fusion based on transformation was found to be preferable to that based on classification. Moreover, it was evident that the fusion of these features resulted in a significant improvement in terms of classification compared with the values obtained from the individual analyses. For gender and age, the best accuracy was found with both a sum and a weighted product (84.62% and 84.45%, respectively). Based on the collected data, it was observed that concatenation increased the accuracy of classification by about 4% for age and a bit more for gender. The fusion technique based on the transformation of scores versus sum and weighted product achieved the best performance for both classification tasks while adjusting the number of algorithms involved. To obtain the best performance in gender classification, it was sufficient to examine only two classifiers and rely on the weighted product. For figuring out a person's gender, the proposed fusion technique, which used feature-level concatenation and score-based fusion, did better than the current best method by more than 25%. For age classification, however, there was a slight increase.

7.2 BIOMETRIC RECOGNITION

The advancement of information technology has caused a data explosion. Companies collect and preserve more consumer information than ever before. Similarly, the amount of recorded data associated with individual users has exploded during the past decade. This is especially true in the online environment. Online registrations and surveys are used to acquire user information. In addition, information about an online user's activities is frequently collected covertly, even as the person surfs the web. An essential difficulty is efficiently summarising user-level information so that it can be utilised successfully in electronic commerce.

User profiles can assist in summarising the vast quantities of information accessible from a user and achieving objectives such as product recommendations and customised information delivery. A user profile can include information explicitly submitted by individuals through registration and surveys. The name, telephone number, and address of a user, as well as information about their hobbies, are frequently provided by users. Explicit information also includes simple facts about the activity or transactions of the user. Such information may include, for instance, the frequency with which a user visits a website, the average amount spent per purchase transaction, and the most popular product category. In addition to explicit information, a user profile may also contain implicit information derived from evaluating the user's activities, typically using more complex statistical or data mining techniques.

Since smartphone usage has surpassed that of personal computers, academics are focusing on mobile devices. In fact, modern smartphones are equipped with a variety of sensors that may be used to deduce the actions of the device's owner and infer relevant information such as position, rotation, acceleration, magnetic field, etc. All of this information may also be relevant to the analysis of a user's behaviour. In addition, nearly all smartphones feature touchscreens as a means of user engagement. It is feasible to evaluate *touch dynamics* to examine how people interact with a touchscreen. These dynamics take into account a variety of factors, such as the time between keystrokes, the pressure of each touch, the movement of a user's touch, the screen, etc. The last ones are difficult to imitate and can therefore be considered a distinguishing characteristic for identifying users.The benefit of mining and utilising such biometric data is that it does not interfere with the user's activities (although it may be considered "privacy mining") because it is fully transparent to the device's owner. Therefore, the usage of smartphones enables access to a huge number of behavioural and physiological characteristics, which is the optimum solution for a biometric technology-based system.

Users touch their smartphones, but they also use their eyes to interact with such devices. Together with the dynamics of touch, it may be particularly fascinating to examine the eyes as an additional biometric factor.

For this reason, the driving idea of our work [263] was to merge together the eye data and the touch dynamics to further improve the performance of the recognition systems by develop-

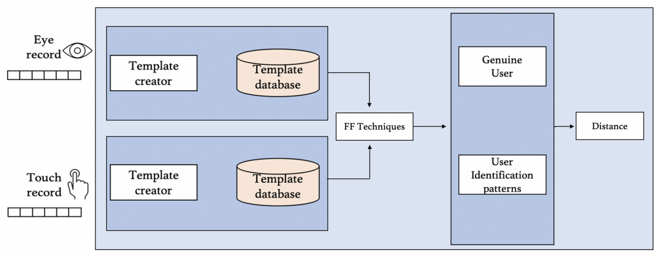


Figure 7.5: Architecture of the proposed multi-modal biometric system. After feature extraction, Feature-level fusion (FF) techniques were applied.

ing efficient and effective methodologies that used both types of biometrics. The approach adopted to improve the overall performance of the multimodal biometric recognition system was based on a fusion at the feature level, to which different distance measurement techniques (Euclidean, Bray-Curtis, Manhattan, Canberra, Chebyshev, Cosine) were applied to determine if the test sample belonged to the target subject. To further improve the system performance, we applied multi-data processing methods such as Canonical Correlation Analysis (CCA) and Principal Component Analysis (PCA). So, this approach also gave an additional advantage: it allowed to gather data originating from various sources and transform it into a unique representation.

The proposed work required a data set encompassing the aforementioned biometric characteristics, namely a benchmark of keystroke dynamics obtained using a touch screen phone and eye data patterns such as pupil size, blinking, and fixation points. A new one was formed because, to the best of our knowledge, no database existed that simultaneously gave these two characteristics of a human. The RHU KeyStroke Dynamics Benchmark dataset [264] and the GANT dataset were integrated to create a multi-modal database. Based on their age and gender, 19 individuals were picked from the combined datasets. Each subject had eighteen acquisitions.

The architecture of the proposed multi-modal biometric system is shown in Figure 7.5. The input of each module was represented by the data belonging to the corresponding dataset. Then, the Feature-level fusion techniques were applied. Subsequently, 70% of the resulting dataset was used to extract the identification pattern for each user, and the remaining samples made up the test set, ensuring that at least one acquisition per subject was included. Finally, to evaluate the dissimilarity or similarity score between the test samples and the training samples of the 19 subjects, several distance measurement techniques were calculated.

In order to examine the performance of the system, a large comparison experiment was conducted, using a variety of statistical indices and combining different methodologies for fusing characteristics, as well as analysing the application of CCA and PCA. The performance was remarkable, and the results exceeded 90%.

7.3 COGNITIVE PROCESSES

Cognitive processes are chemical and electrical signals in the brain that help people understand and learn. Neurons emit substances that create electrical signals in adjacent neurons, which are translated into conscious and unconscious thoughts. Cognition includes five-sense interpretation, procedural knowledge, and emotional reactions. Sensation, attention, perception, memory, learning, language use, and problem solving are cognitive activities.

7.3.1 *Attention measure*

Attention is the cognitive process through which certain aspects of the world-environment are perceived, made present to consciousness, and evaluated in their details and meanings in preference to other aspects that are likewise present in the perceptual field. It operates as a filter that, on the one hand, ignores certain stimuli and, on the other, enables the selection and organization of information to be analyzed in order to implement an appropriate response; it also turns out to be the basis of recognition and perception. It is believed that attention is not a unitary process but rather different information-processing mechanisms that operate at different levels, mediating different but complementary attentional aspects, concur in it. It is not a linear process; it is characterized by slow, involuntary, physiologically regulated changes that are influenced by various factors.

Blinking is a good indicator of attention status. When attention levels are high, people tend to blink less so as not to lose eye contact with the object of interest. An increase in the blink rate, in contrast, is associated with fatigue. This can best be understood as a cessation of attentional inhibition of blinking. The gaze can detect how people acquire information, playing an important role in human communication by reflecting cognitive processes and also allowing one to recognize areas of interest. If the gaze concentrates on a particular area for a prolonged period of time, the interpretation is twofold: either a cognitive difficulty in understanding information or greater interest.

In our work [231] we proposed a software module that analyzed these features jointly with information from another mostly involuntary reflex, the yawning. The goal was to develop a system that could support the teacher during a distance lecture.

During a lesson, the teacher requires continuous feedback from students to know if the audience is getting the topic and the explanations provided. In other words, if the lesson is understood and if it is held at an adequate pace. If this is a desirable requirement in classroom education, in case of distance teaching modalities it becomes crucial. The answers to these questions are usually deduced by students by examining their expressions and attitudes, listening to their questions, or simply asking them directly. In the recent years, the COVID-19 pandemic outbreak forced every level of educational institutes to adopt distance learning as their first choice. In distance learning, the absence of visual feedback makes the teacher unaware of the potential effectiveness of their lesson. The full and wide adoption of this learning modality has underlined the need for tools to reduce the distance between teacher and students. With our system, we wanted to offer a measurement of attention by tracing the blinks, gaze, and student expressions. The tool aggregated information

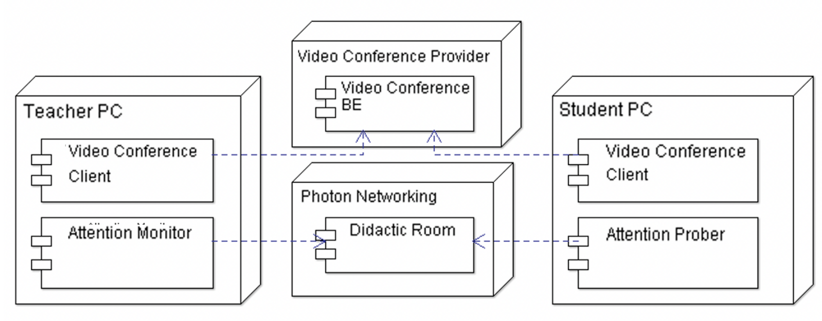


Figure 7.6: The deployment diagram of the proposed architecture.

on how the attendees' visual attention was being distributed on the projected slides in the form of heat-maps and provided other aggregated information on the classroom. The teacher was so helped by a useful tool to overcome the lack of visual feedback in the classroom from the audience and was also able to understand if a specific topic was confusing the entire audience or a significantly high part of the class by observing the visual patterns that tended to be spread out.

Figure 7.6 shows the architecture of the proposed didactic platform. In the proposed setting, as it is possible to notice, both the teacher and the students adopted a regular Video Conference system for synchronous distance lectures (e.g., Microsoft Teams, Zoom or others). The teacher shared the PowerPoint presentation of the lecture, and the students just attended the video lecture. Before starting the lecture, the students launched the "Attention Prober" application, and the teacher his/her "Attention Monitor". The "Attention Monitor" aggregated information for the teacher and exposed the heat-map made at run-time with students' gaze directions and the distributions of classified expressions. In particular, the user expressions, shown in the center-low side of the teacher GUI (Figure 7.7), were organised in an histogram labeled with expression emoticons. This information quickly signaled to the teacher useful hints on the flow of distance didactic action. The applications were connected via Photon[1] networking engine, which, designed as a chatting and multiplayer facility

1 https://www.photonengine.com/pun

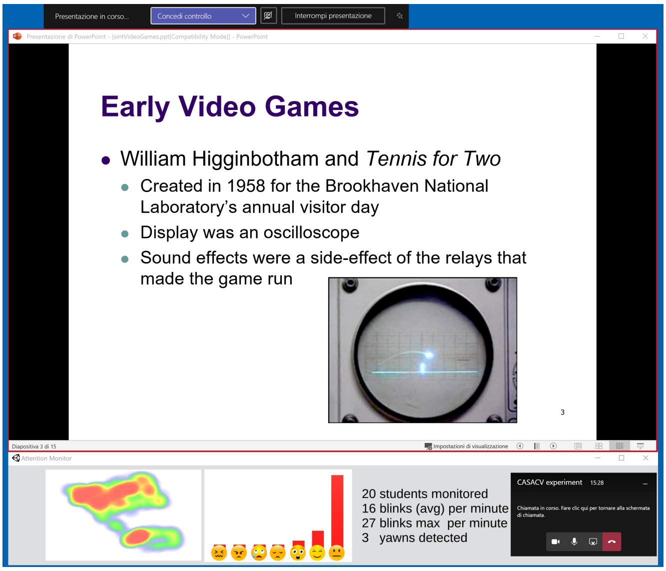


Figure 7.7: The Attention Monitor GUI, set by side of teacher Video-lecture software.

for gaming, supports distance lectures by ideally shifting the "room" metaphor to the "classroom" one (Didactic Room). The teacher arranged the applications on its main screen or put the game "Attention Monitor" on its second monitor, if available. During the controlled experiment, Microsoft Teams was adopted as a distance lecture environment, also according to our teaching habits. The lecture has been driven by PowerPoint, with the presentation software set in reading mode. We did not use typical presentation or speaker modalities because these required a full-screen presentation that covered the "Attention Monitor" application. As it is possible to notice in Figure 7.7, the Teacher GUI consisted of a predominant upper portion of the screen that was reserved for the content of the lecture, while the bottom was intended to provide instant feedback about the audience attention in the form of the cumulative heat-map, a histogram of user expressions, and measurements of blink and yawn rates. From the student's side, no GUI was required since the Attention Prober worked in the background, accessing the webcam and analysing the video-frames. Moreover, a GUI on student's side could only have been a source of distraction from the lesson.

The support provided by this proposed approach to synchronous distance learning was evaluated in two controlled experiments: one performed with the voluntary participation of 20 students attending the "Context Aware Security Analytics in Computer Vision" (Computer Science Master Degree, Internet of Things Curriculum of the University of Salerno, Italy, and the other involving 30 computer scientists and engineers of Kineton, an engineering company specialised in the automotive, media, entertainment, and telecommunications sectors. It is important to point out that the topics adopted for this assessment were not part of the student course program (as well as not related to working practices for Kineton employees) and that neither the students nor the employees were evaluated during the experiment. In addition, as it is possible to deduce from the system description, the tool was specifically formulated for Distance Education actions.

From the student's point of view, the proposed software module was considered non-intrusive, and they reported a sense of trust due to the chance of not sharing personal recordings. From the teacher's side, instead, the proposed software was considered a valid support tool to control the flow of the lecture. The teacher was solicited to ask questions when indicators suggested a decrease of attention in the audience, thus filling that gap in the communication that was often sensibly driven by body language and natural attitudes or the experience.

After the tuning of the applications, the effectiveness of the system was evaluated in two sessions: a *Didactic session* with students and *Industrial session* with employees. Both sessions were focused on the application of the proposed analysis to the video captured during two oral presentations performed by the teacher in a distance learning mode. In particular, the same two lectures were adopted for both sessions and focused on the "History of Video Games" and the "History of Typography". The idea behind the choice of the two topics was to propose two lectures that would evidently receive different attention from the participants. Indeed, because they were computer scientists in both the Didactic and the Industrial sessions, the involved volunteers clearly should prefer video games with respect to typography, and we were expecting to detect this preference in the evaluation results. The rationale behind this choice was that we expected more interest in topics about the video games from an audience of computer scientists, which should have reflected in their behaviours during the lessons in terms of gaze, blinks, and yawns. We adopted a fully balanced design for the experiment in both the evaluation sessions: half of the participants started with a video game lecture, and the other subjects with a typography lecture. This prevented from fatigue or boredom to bias the results.

Before the controlled experiment, a pre-evaluation questionnaire was submitted to participants to gauge their degree of interest in the presented topics. The questionnaire assessed the attitudes and practices of the participants by asking them directly 12 questions (6 per topic), but also in terms of purchase intentions. The "Pre-Experiment questionnaire" established a classification of user attitudes toward Video Games and Typography in general. The twelve questions were organised as follows: 6 questions from were focused on Video Games while their counterparts for Typography. According to their nature, the questions required different categorical answers.

After every presentation, the participants were asked to fill out the "Post-Experiment Questionnaire" with answers anchored on the 5-point Likert scale going from "Strongly Agree" (anchored at 1) to "Strongly Disagree" (anchored at 5). The questionnaire aimed at subjectively evaluating the expected interest toward the two subjects and the concentration applied during the attended distance lectures.

Both the controlled experiments were articulated in the same two phases:

- the preliminary ethnographic survey aiming at assessing the characteristics of the participants to the experiment (Pre-Experiment Questionnaire);

- the on-the-field evaluation performed between the two different degrees of participant's interest perceived and measured during the two distance lectures (video games and typography).

Both the presented lectures adopted in the last phases were made of 14 content slides and one header slide: the oral presentations performed by the teacher lasted both around 14 minutes, one minute per slide. The contents were simple and minimalist: just black text on white background.

The controlled experiment checked the participant's interest in the two topics proposed by asking him or her to answer a questionnaire before starting the two didactic sessions (the first phase) and by directly asking their opinions after both presentations. Objective measurement of user attention was performed, off-line, by measuring gaze fixations expressed as a percentage of the slide duration.

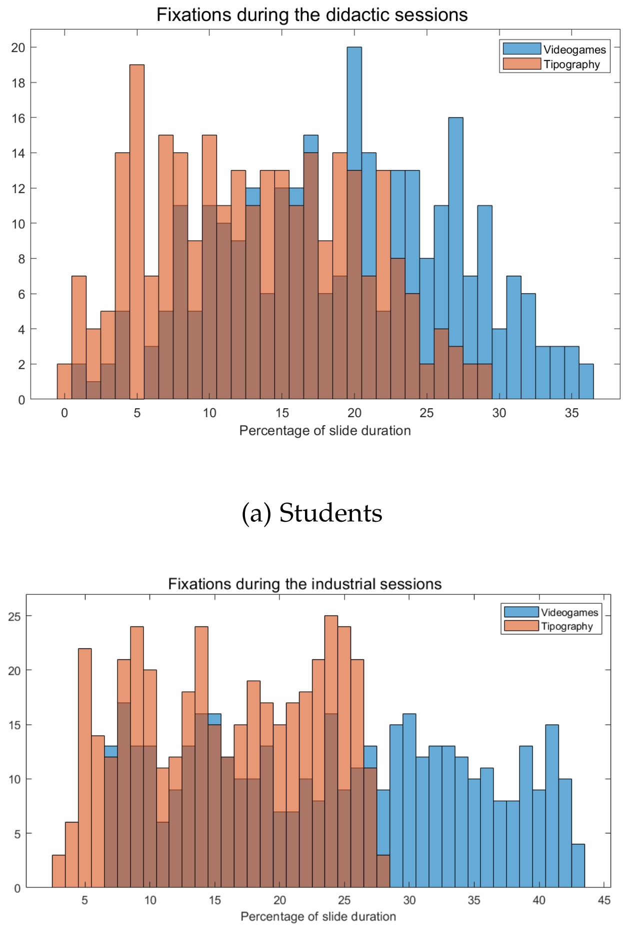
The controlled experiment provided two classes of results, with respect to didactic and industrial users:

- a subjective evaluation of participants' interest (before and after the presentations) toward the two lectures, collected via questionnaires.
- objective measurements were collected by analysing students' gaze direction (also in terms of fixations), their expressions, and counting detected blinks and yawns.

For the purposes of this evaluation, only gaze direction was taken into consideration, utilising fixation times as a rough estimation of student attention. The idea was to record participants' gaze fixations expressed as a percentage of the entire time the slide was presented.

The experimental results on volunteers reported positive feedback, both in terms of gaze tracking and evaluation questionnaire.

The Pre-Experiment Questionnaire was answered by all the participants (Didactic and Industrial) before starting the two lecture sessions. Computer Science students participating in the



(a) Students

(b) Employees

Figure 7.8: The fixations of participants' gazes expressed as the percentage of slides' duration.

experimental sessions confirmed the assumption that they were more interested in video games than in a lecture on the history of typography. The expectation about the propensity towards the proposed topics was confirmed also for the employees participating in the industrial experimental sessions: because of their technical skill and of their attitudes, also the second group of participants appeared more interested in video games than to the history of typography. This was even more evident, compared to the Didactic session participants.

After the Pre-Experiment questionnaire, we were expecting to detect the same differences in the answers given to the Post-Experiment questionnaire. In particular, we aimed at showing the same differences underlined by the previous phase of the evaluation, on objective indicators (fixation time) obtained by Deep Learning techniques applied to the student videos.

Figure 7.8 depicts the histograms of collected fixation times aggregated by session for students and employees. For the students, the average fixation time collected during the Video Games lecture was 19.31% and the one related to Typography was 13.1%. As it is possible to graphically notice (Figure 7.8(a)), the difference in fixation times between the two topics presented was evident. For the employees, the average fixation times collected were 24.31% for Video Games lecture and and 16.1% for the one related to Typography. As it is possible to graphically notice, the difference in detected fixation times between the two topics is evident in Figure 7.8(b) and it is higher than the one reported in the Didactic Session. We believe that the students were more accustomed to being concentrated on a presentation they were not interested in than the employees were.

We adopted the non-parametric Wilcoxon Signed Rank and Rank Sum tests[2] to obtain the statistical analysis of the results obtained from the questionnaires and fixations. We complied with all requirements providing good statistical power for the validity of the results.

The objective evaluation performed on the ML "raw" indicator of participants' attention (fixation time) has confirmed the results obtained by the questionnaires answered by the participants after the experiment. We believed this assonance as a good indicator

2 The Wilcoxon rank-sum test is used to compare two independent samples, whereas the Wilcoxon signed-rank test is used to compare two related samples, matched samples, or to conduct a paired difference test of repeated measurements on a single sample to determine whether their population mean ranks differ.

of the efficacy of the proposed approach.

7.3.2 *Perception and Memory*

Perception can be defined as the ability to capture, process, and make sense of information that reaches our senses. It is the mental process that lets us figure out what's going on around us based on the information we get from our senses. Memory, on the contrary, is the cognitive ability that permits us to encode, store, and retrieve past information.

Numerous studies have shown that the eyes are a key source of information about cognitive and emotional states. By assessing the periocular area, such as pupils, blinks, and eye-movements, it is possible to try to decipher the cognitive picture in which the subject is placed. Based on the evidence in the literature and using the data collected from these characteristics, we posed the following questions: can we classify a subject based on periocular characteristics when he or she sees an image for the first time but is already familiar with the visual stimulus? Can we determine from the study of a subject's periocular features if he is viewing a picture with obvious semantic information or not?

We tried to answer the following questions in our work [265]. The purposes are set out below:

- ***Memory task.*** Classify whether a subject had previously seen a particular image in terms of mnemonic processing;
- ***Perception task.*** Classify whether an image has clear semantic content (such as images of natural and urban environments) or unclear semantic content (such as noisy and geometric images).

Creating an inference system on the interaction between input from the periocular region and cognitive processes was the objective of this pioneering study.

Our research utilised the Memory I dataset [266]. The head-mounted Eyelink II eye-tracking system with monocular sampling at 500 Hz was used to gather gaze coordinates. 45 people (aged 18 to 48, with a mean age of 21.68) participated in the trial. They independently viewed 48 photos in a random order five times. The subject was shown each image for six seconds. The photographs spanned four distinct categories: natural, urban, fractal, and pink noise. The first two categories (Natural and Urban) were labelled as clear, immediately recognisable images. The other two kinds of visual stimuli were classified as uncertain because their content is either noisy (Pink Noise) or completely geometric (Fractal).

We deleted all samples having a negative timestamp, which matched the calibration procedure preceding the presentation of each image. In addition, because the pupil size reduced in the region of a blink, we deleted the samples below a particular threshold value (700) while keeping the blink indicator values at 0. For the memory task, we extracted just the data pertinent to the current investigation, namely those pertaining to the initial and final iterations. This was because we intended to investigate the differences in eye gazing behaviour induced by new images presented in iteration 1 versus familiar images resubmitted in iteration 5.

Memory task-Approach 1. In the scientific literature, the pupil has been discussed in a handful of studies, such as [267], which concluded that pupil diameter increases in response to previously witnessed stimuli. Since the paper did not specify the features to be extracted, we arbitrarily selected the statistical indices to use. We developed the same statistical indices for blink duration, which was seen to decrease with repeated stimulus presentation [268], for the same reason. Since a number of studies in the literature evaluated the duration of the saccade, we opted to calculate the information that performed the best: variance, as provided in the article [269], mean, total, and standard deviation, as stated in [270]. Moreover, the number of saccades cited in [271] and the relative and absolute angles (in terms of mean, total, and standard deviation) calculated by [270] appeared to be discriminatory.

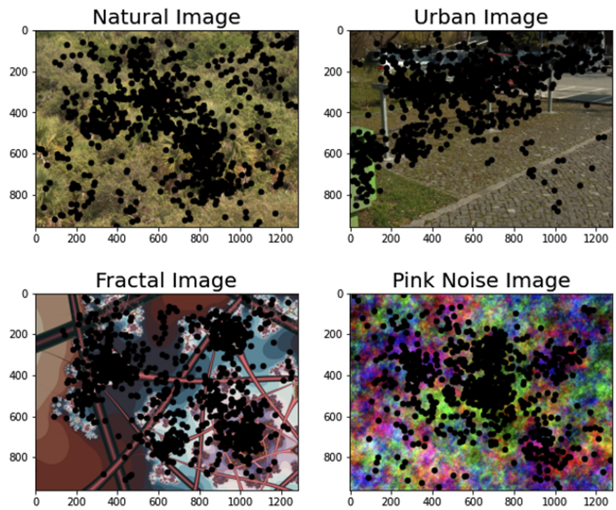


Figure 7.9: Analysis of the fixation point for each of the four image categories. A selection of images depicting natural, urban, fractal, and pink-noise settings, from top to bottom and left to right.

Perception task. For this classification problem, the computation was performed by coupling the elements of the subdivided macro-categories: Natural *vs* Fractal; Natural *vs* Pink Noise; Urban *vs* Fractal; Urban *vs* Pink Noise. We utilised descriptive statistical approaches (e.g. boxplot) and strategies of graphical visualisation of features (as illustrated in Figure 7.9) in an early stage of our research due to the paucity of relevant literature for this sort of classification. Then, we chose the most promising characteristics for the fixations: standard deviation on the x-axis, number of clusters (we used DBSCAN), and standard deviation, mean, 30-th and 70-th percentiles for their length. We have also estimated the standard deviation, mean, 30th and 70th percentiles, as well as the number and length of saccades.

Memory task-Approach 2. After the considerations made and the results obtained with respect to both *Memory task-Approach 1* and *Perception task*, we decided to make a further study for *Memory task* but this time taking into account different characteristics: information on the semantic content of the image (with respect to the division in *Perception task*, i.e. clear and unclear labels), num-

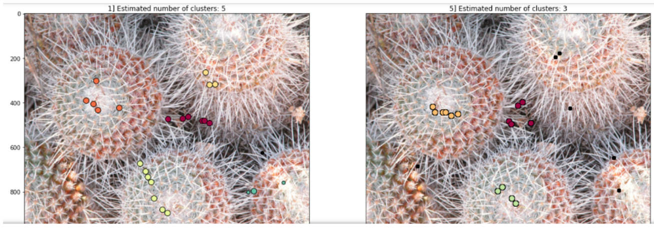


Figure 7.10: Number of fixation clusters on a natural image that is displayed for the first and fifth time.

ber and mean duration of fixations, number of clusters (we used DBSCAN, an example in the Figure 7.10), number of noise points, silhouette score [272]. Therefore, this study was conducted solely by utilising information regarding each participant's fixations.

After the extraction of the characteristics, we proceeded with the application of different ML techniques. Although the results were encouraging, it was clear that further investigation was required to determine which features should be explored for each task and which should not be included because they just served as a source of noise for the model. Some considerations were possible: for the *Memory task*, comparison of the two approaches revealed, however, that integrating the information of the other periocular features did not improve the system's overall performance; on the other hand, for the *Perception task*, it was observed that the Fractal class displayed similar behaviour to that of images we labeled as "clear." This we assumed was due to the type of images that allow the observer to have reference lines to follow in space that guide him or her in scanning the image, as may be the case with the lines delineating a stream or a road in the images we labeled as "clear." In contrast, when the individual was presented with a noisy image with no landmarks or lines to follow, he tended to focus his attention primarily on the image's center (as shown in Figure 7.10).

Part IV

CONCLUSIONS

8 FINAL DISCUSSION

Pensino ora i miei venticinque lettori che impressione dovesse fare sull'animo del poveretto, quello che s'è raccontato.

— Alessandro Manzoni, I promessi sposi [273]

In this thesis, we describe our idea in terms of soft biometric applications developed over the past three years. Based on our experience in this field, we can assert that soft biometric features should not only be viewed as a substitute for hard biometrics in circumstances where their usage is problematic but also as a valuable source of additional or self-consistent information. Soft biometrics is so flexible that it can be applied in contexts that may also be completely unrelated to the classical goal of recognizing a user's identity. The two main alternatives to the recognition task for applying this biometric are to provide a window into cognitive processes and to gather information about users' emotional responses to stimuli.

Then, we examined the integration of various soft, physical, behavioural, and physiological biometrics traits into a multi-biometric system designed to enable human interaction with HSRs as effective and realistic as possible. Particularly, in the presented investigations, they were mostly used as a single source of information for detecting suspicious behaviour or the subject's health or emotional state.

In general, therefore, it has been shown that these characteristics can be particularly interesting but also necessary for certain purposes. The same feature can be employed in various ways to gather diverse types of information. Being a rather vast area of study, it has been found that the majority of studies have concentrated on a small subset of traits while assessing the effectiveness of others just in passing. For this reason, in this study, we have decided to vertically focus on soft periocular characteristics in terms of the information that may be derived from pupils, blinks, eye movements, and fixations. After a thorough study of the literature, we found a variety of evidence of their potential in many research areas, including biometric recognition, demographic classification, and cognitive process detection. Nonetheless, we discovered a noticeable paucity of research in all three areas with regard to systems capable of making inferences from acquired data. Due to this lack, we decided to investigate the study of these traits through preliminary exploratory experiments that have demonstrated their potential.

Attribute correlation, distance, permanence or stability, discrimination, and feature or modality level fusion directly impact the performance of any soft biometrics recognition or retrieval system. Below is an overview of the most relevant issues and challenges, open issues, and future directions.

8.1 OPEN PROBLEMS

Several unresolved problems must be addressed before the use of soft biometrics can be successfully incorporated and become an increasingly researched study area. In a real and articulated application setting, such as the development of a robotics system to make Human-Robot Interaction natural and effective, there are numerous aspects that may be collected and handled in order to optimise the user experience.

The precise extraction of such traits is one of the foremost obstacles in designing systems based on soft biometrics. Automatic and reliable extraction of soft biometric information in a non-intrusive way without causing any inconvenience to users is a crucial point. In applications where these soft features are employed in a first stage as a rough filter to perform a first sorting in big databases or as a tool to alter the parameters of a biometric system, it is also acceptable for the characteristics of the data extraction module to be less than 100 percent accurate. Consequently, subsequent actions will serve as corrective steps. In contrast, soft biometric features employed as unique characteristics or merged in recognition systems with a primary biometric identifier to increase the system's overall accuracy must be extracted with extreme precision. This becomes more challenging in surveillance scenarios. Similar to other vision-based recognition systems, distance affects the accuracy of estimating various soft biometrics. In an open recognition environment, it becomes a bigger challenge. Obviously, distance affects recognition, therefore it would be interesting to have a distance-based sensitivity measurement for each soft biometric.

To develop an autonomous recognition system or soft biometric retrieval, it is necessary to identify a collection of permanent and distinctive traits. In general, it is prudent for any recognition system to employ a small but highly relevant collection of features. Obviously, the same holds true for soft biometrics. In this instance, the preferred traits are those with a higher permanence score and more discriminatory power. Various mathematical and statistical methodologies are used in a number of trials to calculate the claimed features of a given soft set. This investigation must be refined.

Benchmark datasets play a vital role in driving the goals of ML communities and tracking progress within the field. The lack of rigorous, standardized and shared datasets contributes to the lack of full development of the paradigm of soft biometrics and, in particular, of periocular ones.

In addition to the collection of biometric data, the secure storage of biometric data has arisen as a major concern. The majority of robotic equipment that uses biometrics relies on cloud-based storage to maintain biometric data records. These databases can be compromised, and biometric information can be stolen. When this occurs, the repercussions are significantly more severe than when a conventional account or password is compromised.

8.1.1 *Challenges*

The most difficult aspect of acquiring different types of soft biometric data is determining how to combine them.

It is not easy to create an information fusion system that improves overall recognition accuracy despite imperfect extraction of soft biometric parameters. In practice, there are a number of factors to consider, both in terms of combining strategies and finding the optimal balance between the various traits.

When deciding to apply fusion techniques to improve the accuracy of a system, it would be a good protocol to first investigate the correlation between soft biometric data; only in this way can meaningful input be obtained. Depending on the objective, it is obvious that some correlations between two or more soft biometric data are more significant than others. For example, if the goal is to develop a system that discriminates on the basis of ethnicity, those with darker skin tend to have darker hair, whereas gender may not help much for more accurate recognition. Associations also represent actual worldviews. Finding the correlation between soft features will not only increase the efficiency of the input, but will also reduce the size of the feature set, which has significant computational benefits.

It is certain that both robotics and biometrics will have a growing influence in the future. Discovering how these two paradigms can coexist, optimising their unique characteristics, will open the door for a more extensive body of research. Obviously, identifying and respecting the line between what is public and what is effectively private will be essential to the success of their joint implementation.

8.2 FUTURE DIRECTIONS

Regarding a particular classification goal, which soft biometric modalities are the most discriminating? How precisely can a soft biometric characteristic be derived automatically in an uncontrolled environment? How much can system efficiency be enhanced by implementing fusion strategies? These are only a few of the difficult and ambitious issues we hope to address in the future.

Even with regard to the outstanding problems and challenges, it is clear and undeniable that there is still much to be done in terms of research and development in soft biometrics and that their future involvement will be of increasing interest in a wide range of areas.

The main future contribution will be to develop an efficient, scalable, robust, and, above all, adaptive system depending on the situation under consideration for recognition purposes in an IoT, real-world, and wild context such as industry.

In such a scenario, then, data can be collected from various devices and under various conditions. Using a humanoid social robot both as the fulcrum of this IoT system in an uncontrolled environment and as a data acquisition tool, we will also analyse the effect of distance on performance and determine which biometric trait to favor based on the situation.

Therefore, first we will establish our case study and the biometrics to be extracted and, based on this, the acquisition devices that will be able to compete for the data acquisition. Then, we will proceed with the creation of our large and heterogeneous dataset, which will involve multiple acquisitions over an extended period of time.

With the database ready, we will implement an adaptive system based on multimodal biometric fusion techniques by evaluating the best approach and configuration.

So that the combination of these paradigms can be used to its fullest extent, it is first necessary to study the best techniques, but also to extract the data by taking into account the hardware and staging the most likely situations, while only looking at one stream of information at a time to get the best performance. To address the question of whether anti-cooperative behaviour is related to a prior or subsequent detrimental action, for instance, we intend to combine the idea of cooperativeness with action recognition.

Especially for periocular biometry, it will be important to do extensive preliminary research using a consistent database with samples extracted with respect to different conditions and from heterogeneous sources. We have observed that cognitive processes have an effect on performance. It would be interesting to conduct a cross-sectional study in which the same individual performs a variety of activities, from reading to daily tasks. We believe further work is needed to crystallize the accuracy performance currently found in the literature, especially when there is a significant time interval between training and testing sessions. How much does the time gap affect the results? There is no doubt that the different lighting conditions, emotional state, and cognitive condition a subject is in have an impact on the acquisitions. For this reason, it would also be of great interest to create a specific dataset that is as complete and varied as possible.

Thus, the initial stage of the study will involve expressing the full potential of each characteristic examined before taking it into account in the group of those that the system, of which the robot is the core, can acquire. Preliminary research will also focus on expanding the understanding of and application of brain activity-based features to close the gap with physical and behavioural features.

HSRs will play a key role. For this reason, it is essential to safeguard the data saved by in-house service robots, which, days, months, or years later, will be aware of every user's behaviour. In this way, we should devise a method for efficiently storing and protecting such data, possibly by integrating biometrics and encryption techniques studied in other contexts.

PUBLICATIONS

[1] Andrea F. Abate, Carmen Bisogni, Lucia Cascone, Aniello Castiglione, Gerardo Costabile, and Ilenia Mercuri. "Social Robot Interactions for Social Engineering: Opportunities and Open Issues." In: *2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*. 2020, pp. 539–547.

[2] Lucia Cascone, Michele Nappi, Fabio Narducci, and Ignazio Passero. "DTPAAL: Digital twinning pepper and ambient assisted living." In: *IEEE Transactions on Industrial Informatics* 18.2 (2021), pp. 1397–1404.

[3] Carmen Bisogni, Lucia Cascone, Aniello Castiglione, and Ignazio Passero. "Deep learning for emotion driven user experiences." In: *Pattern Recognition Letters* 152 (2021), pp. 115–121.

[4] Lucia Cascone, Aniello Castiglione, Michele Nappi, Fabio Narducci, and Ignazio Passero. "Waiting for Tactile: Robotic and Virtual Experiences in the Fog." In: *ACM Transactions on Internet Technology (TOIT)* 21.3 (2021), pp. 1–19.

[5] Andrea F Abate, Paola Barra, Carmen Bisogni, Lucia Cascone, and Ignazio Passero. "Contextual trust model with a humanoid robot defense for attacks to smart eco-systems." In: *IEEE access* 8 (2020), pp. 207404–207414.

[6] Lucia Cascone, Carlo Medaglia, Michele Nappi, and Fabio Narducci. "Pupil size as a soft biometrics for age and gender classification." In: *Pattern Recognition Letters* 140 (2020), pp. 238–244.

[7] Virginio Cantoni, Lucia Cascone, Michele Nappi, and Marco Porta. "Demographic classification through pupil analysis." In: *Image and Vision Computing* 102 (2020).

[8] Carmen Bisogni, Lucia Cascone, and Fabio Narducci. "Periocular Data Fusion for Age and Gender Classification." In: *Journal of Imaging* 8.11 (2022), p. 307.

[9] Lucia Cascone. "Periocular features as a window on cognitive processing: from memory to understanding the semantic complexity of an image." In: *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*. 2021, pp. 1–4.

[10] Andrea F Abate, Lucia Cascone, Michele Nappi, Fabio Narducci, and Ignazio Passero. "Attention monitoring for synchronous distance learning." In: *Future Generation Computer Systems* 125 (2021), pp. 774–784.

[11] Andrea Casanova, Lucia Cascone, Aniello Castiglione, Wei-zhi Meng, and Chiara Pero. "User recognition based on periocular biometrics and touch dynamics." In: *Pattern Recognition Letters* 148 (2021), pp. 114–120.

[12] Andrea Casanova, Lucia Cascone, Aniello Castiglione, Michele Nappi, and Chiara Pero. "Eye-movement and touch dynamics: A proposed approach for activity recognition of a web user." In: *2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE. 2019, pp. 719–724.

[13] Carmen Bisogni, Lucia Cascone, Jean-Luc Dugelay, and Chiara Pero. "Adversarial attacks through architectures and spectra in face recognition." In: *Pattern Recognition Letters* 147 (2021), pp. 55–62.

[14] Lucia Cascone, Michele Nappi, Fabio Narducci, and Chiara Pero. "Touch keystroke dynamics for demographic classification." In: *Pattern Recognition Letters* 158 (2022), pp. 63–70.

[15] Lucia Cascone, Piercarlo Dondi, Luca Lombardi, and Fabio Narducci. "Automatic Classification of Fresco Fragments: A Machine and Deep Learning Study." In: *International Conference on Image Analysis and Processing*. Springer. 2022, pp. 701–712.

[16] Lucia Cascone, Michele Nappi, Fabio Narducci, and Sara Linda Russo. "Classification of fragments: recognition of artistic style." In: *Journal of Ambient Intelligence and Humanized Computing* (2022), pp. 1–11.

BIBLIOGRAPHY

[1] Jong-Hwan Kim, Seung-Hwan Choi, In-Won Park, and Sheir Afgen Zaheer. "Intelligence Technology for Robots That Think [Application Notes]." In: *IEEE Computational Intelligence Magazine* 8.3 (2013), pp. 70–84. DOI: 10.1109/MCI.2013.2264573.

[2] Kim Baraka, Patrícia Alves-Oliveira, and Tiago Ribeiro. "An extended framework for characterizing social robots." In: *Human-robot interaction*. Springer. 2020, pp. 21–64.

[3] *Scopus*. https://www.scopus.com/.

[4] Thomas B Sheridan. "Teleoperation, telerobotics and telepresence: A progress report." In: *Control Engineering Practice* 3.2 (1995), pp. 205–214.

[5] B. Siciliano and O. Khatib. *Springer Handbook of Robotics*. Springer Handbook of Robotics. Springer Berlin Heidelberg, 2008. ISBN: 9783540239574. URL: https://books.google.it/books?id=Xpgi5gSuBxsC.

[6] Andrea F. Abate, Carmen Bisogni, Lucia Cascone, Aniello Castiglione, Gerardo Costabile, and Ilenia Mercuri. "Social Robot Interactions for Social Engineering: Opportunities and Open Issues." In: *2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*. 2020, pp. 539–547.

[7] Columbia Eye Clinic. *Anatomy of the Eye*. URL: https://columbiaeyeclinic.com/anatomy-eye/. Access: 14/11/22.

[8] Takeo Kanade, Jeffrey Cohn, and Ying-Li Tian. "Comprehensive Database for Facial Expression Analysis." In: *Proceedings of 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG '00)*. 2000, pp. 46 –53.

[9] Punam Kumari and Seeja K.R. "Periocular biometrics: A survey." In: *Journal of King Saud University - Computer and Information Sciences* 34 (June 2019). DOI: 10.1016/j.jksuci.2019.06.003.

[10] Ajita Rattani and Reza Derakhshani. "A Survey Of mobile face biometrics." In: *Computers & Electrical Engineering* 72 (2018), pp. 39–52. DOI: https://doi.org/10.1016/j.compeleceng.2018.09.005.

[11] Lucia Cascone, Carlo Medaglia, Michele Nappi, and Fabio Narducci. "Pupil size as a soft biometrics for age and gender classification." In: *Pattern Recognition Letters* 140 (2020), pp. 238–244.

[12] Muhammad Wasimuddin, Khaled Elleithy, Abdel-Shakour Abuzneid, Miad Faezipour, and Omar Abuzaghleh. "Stages-Based ECG Signal Analysis From Traditional Signal Processing to Machine Learning Approaches: A Survey." In: *IEEE Access* 8 (2020), pp. 177782–177803. DOI: 10.1109/ACCESS.2020.3026968.

[13] Anil Jain, Ruud Bolle, and Sharath Pankanti. *Biometrics: personal identification in networked society*. Vol. 479. Springer Science & Business Media, 1999.

[14] Arun A Ross, Karthik Nandakumar, and Anil K Jain. *Handbook of multibiometrics*. Vol. 6. Springer Science & Business Media, 2006.

[15] Charles Wilson, AR Hicklin, Mike Bone, Harold Korves, Patrick Grother, Bradford Ulery, Ross Micheals, Melissa Zoepfl, Steve Otto, and Craig Watson. "Fingerprint vendor technology evaluation 2003: Summary of results and analysis report." In: *NIST Technical Report NISTIR* 7123 (2004).

[16] Davide Maltoni, Dario Maio, Anil K Jain, and Salil Prabhakar. *Handbook of fingerprint recognition*. Springer Science & Business Media, 2009.

[17] Samarth Bharadwaj, Mayank Vatsa, and Richa Singh. "Biometric quality: a review of fingerprint, iris, and face." In: *EURASIP journal on Image and Video Processing* 2014.1 (2014), pp. 1–28.

[18] Tapalina Bhattasali, Khalid Saeed, Nabendu Chaki, and Rituparna Chaki. "A survey of security and privacy issues for biometrics based remote authentication in cloud." In: *IFIP International Conference on Computer Information Systems and Industrial Management*. Springer. 2015, pp. 112–121.

[19] P. A. Johnson, B. Tan, and S. Schuckers. "Multimodal fusion vulnerability to non-zero effort (spoof) imposters." In: *2010 IEEE International Workshop on Information Forensics and Security*. 2010, pp. 1–5. DOI: 10.1109/WIFS.2010.5711469.

[20] Carmen Bisogni, Lucia Cascone, Jean-Luc Dugelay, and Chiara Pero. "Adversarial attacks through architectures and spectra in face recognition." In: *Pattern Recognition Letters* 147 (2021), pp. 55–62.

[21] Maneet Singh, Richa Singh, and Arun Ross. "A comprehensive overview of biometric fusion." In: *Information Fusion* 52 (2019), pp. 187–205.

[22] E.G. Llano, M.S. García Vázquez, J.M.C. Vargas, L.M.Z. Fuentes, and A.A. Ramírez Acosta. "Optimized robust multi-sensor scheme for simultaneous video and image iris recognition." In: *Pattern Recognition Letters* 101 (2018), pp. 44–51.

[23] M. Gomez-Barrero, J. Galbally, A. Morales, and J. Fierrez. "Privacy-Preserving Comparison of Variable-Length Data with Application to Biometric Template Protection." In: *IEEE Access* 5 (2017), pp. 8606–8619.

[24] Christian Rathgeb and Christoph Busch. "Cancelable Multi-Biometrics: Mixing Iris-Codes based on Adaptive Bloom Filters." In: *Computers & Security* 42 (May 2014). DOI: 10.1016/j.cose.2013.12.005.

[25] Tanuja Sudhakar and Marina Gavrilova. "Multi-instance cancelable biometric system using convolutional neural network." In: *2019 International Conference on Cyberworlds (CW)*. IEEE. 2019, pp. 287–294.

[26] G. Goswami, M. Vatsa, and R. Singh. "Face Verification via Learned Representation on Feature-Rich Video Frames." In: *IEEE Transactions on Information Forensics and Security* 12.7 (2017), pp. 1686–1698. DOI: 10.1109/TIFS.2017.2668221.

[27] S.D. Jamdar and Y. Golhar. "Implementation of unimodal to multimodal biometrie feature level fusion of combining face iris and ear in multi-modal biometric system." In: vol. 2018-January. 2018, pp. 625–629.

[28] G.U. Bokade and R.D. Kanphade. "Secure Multimodal Biometric Authentication Using Face, Palmprint and Ear: A Feature Level Fusion Approach." In: 2019.

[29] Richa Singh, Mayank Vatsa, Arun Ross, and Afzel Noore. "A mosaicing scheme for pose-invariant face recognition." In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 37.5 (2007), pp. 1212–1225.

[30] Samarth Bharadwaj, Mayank Vatsa, and Richa Singh. "Aiding face recognition with social context association rule based re-ranking." In: *IEEE International Joint Conference on Biometrics*. IEEE. 2014, pp. 1–8.

[31] Kalyan Veeramachaneni, Lisa Osadciw, Arun Ross, and Nisha Srinivas. "Decision-level fusion strategies for correlated biometric classifiers." In: *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. IEEE. 2008, pp. 1–6.

[32] Antitza Dantcheva, Petros Elia, and Arun Ross. "What else does your biometric data reveal? A survey on soft biometrics." In: *IEEE Transactions on Information Forensics and Security* 11.3 (2015), pp. 441–467.

[33] Aditya Sundararajan, Arif I Sarwat, and Alexander Pons. "A survey on modality characteristics, performance evaluation metrics, and security for traditional and wearable biometric systems." In: *ACM Computing Surveys (CSUR)* 52.2 (2019), pp. 1–36.

[34] Antitza Dantcheva, Carmelo Velardo, Angela D'angelo, and Jean-Luc Dugelay. "Bag of soft biometrics for person identification." In: *Multimedia Tools and Applications* 51.2 (2011), pp. 739–777.

[35] Simon Denman, Michael Halstead, Clinton Fookes, and Sridha Sridharan. "Searching for people using semantic soft biometric descriptions." In: *Pattern Recognition Letters* 68 (2015), pp. 306–315.

[36] Gian Luca Marcialis, Fabio Roli, and Daniele Muntoni. "Group-specific face verification using soft biometrics." In: *Journal of Visual Languages & Computing* 20.2 (2009), pp. 101–109.

[37] Douglas D Heckathorn, Robert S Broadhead, and Boris Sergeyev. "A methodology for reducing respondent duplication and impersonation in samples of hidden populations." In: *Journal of Drug Issues* 31.2 (2001), pp. 543–564.

[38] Carmen Bisogni, Lucia Cascone, Aniello Castiglione, and Ignazio Passero. "Deep learning for emotion driven user experiences." In: *Pattern Recognition Letters* 152 (2021), pp. 115–121.

[39] Joshua C Klontz and Anil K Jain. "A case study on unconstrained facial recognition using the boston marathon bombings suspects." In: *Michigan State University, Tech. Rep* 119.120 (2013), p. 1.

[40] Oskar Palinko, Francesco Rea, Giulio Sandini, and Alessandra Sciutti. "Robot reading human gaze: Why eye tracking is better than head tracking for human-robot collaboration." In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2016, pp. 5048–5054.

[41] James D Naughton. "Futurology and robots: Karel Čaper's RUR." In: *Culture, Theory and Critique* 28.1 (1984), pp. 72–86.

[42] William Morris. "American heritage dictionary of the English language." In: (1969).

[43] Emily E Auger. "Gods and Robots: Myths, Machines, and Ancient Dreams of Technology by Adrienne Mayor." In: *Mythlore: A Journal of JRR Tolkien, CS Lewis, Charles Williams, and Mythopoeic Literature* 38.1 (2019), p. 20.

[44] Karel Čapek and Iosif Kallinikov. *Rossum's Universal Robots*. Fr. Borový, 1940.

[45] Masato Hirose and Kenichi Ogawa. "Honda humanoid robots development." In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365.1850 (2007), pp. 11–19.

[46] Alessandro Gasparetto and Lorenzo Scalera. "A Brief History of Industrial Robotics in the 20th Century." In: *Advances in Historical Studies* 08 (Jan. 2019), pp. 24–35. DOI: 10.4236/ahs.2019.81002.

[47] Jong-Hwan Kim, Yong-Duk Kim, and Kang-Hee Lee. "The third generation of robotics: Ubiquitous robot." In: (Jan. 2004).

[48] Kuk-Hyun Han, Yong-Jae Kim, Jong-Hwan Kim, and Steve Hsia. "Internet control of personal robot between kaist and uc davis." In: *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*. Vol. 2. IEEE. 2002, pp. 2184–2189.

[49] Michael Brady. "Artificial intelligence and robotics." In: *Robotics and Artificial Intelligence*. Springer, 1984, pp. 47–63.

[50] Jong-Hwan Kim, In-Bae Jeong, In-Won Park, and Kang-Hee Lee. "Multi-layer architecture of ubiquitous robot system for integrated services." In: *International Journal of Social Robotics* 1.1 (2009), pp. 19–28.

[51] "Ubiquitous robotics: Recent challenges and future trends." In: *Robotics and Autonomous Systems* 61.11 (2013). Ubiquitous Robotics, pp. 1162–1172. ISSN: 0921-8890.

[52] Samuel B Kesner and Robert D Howe. "Position control of motion compensation cardiac catheters." In: *IEEE Transactions on Robotics* 27.6 (2011), pp. 1045–1055.

[53] Jun Ueda, Ding Ming, Vijaya Krishnamoorthy, Minoru Shinohara, and Tsukasa Ogasawara. "Individual muscle control using an exoskeleton robot for muscle function testing." In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 18.4 (2010), pp. 339–350.

[54] Thilina Dulantha Lalitharatne, Kenbu Teramoto, Yoshiaki Hayashi, and Kazuo Kiguchi. "Towards hybrid EEG-EMG-based control approaches to be used in bio-robotics applications: Current status, challenges and future directions." In: *Paladyn, Journal of Behavioral Robotics* 4.2 (2013), pp. 147–154.

[55] Lina Zhong, Liyu Yang, Jia Rong, and Xiaonan Li. "A complexity analysis of user interaction with hotel robots." In: *Complexity* 2020 (2020).

[56] Matteo Luperto, Javier Monroy, Jennifer Renoux, Francesca Lunardini, Nicola Basilico, Maria Bulgheroni, Angelo Cangelosi, Matteo Cesari, Manuel Cid, Aladar Ianes, et al. "Integrating Social Assistive Robots, IoT, Virtual Communities and Smart Objects to Assist at-Home Independently Living Elders: the MoveCare Project." In: *International Journal of Social Robotics* (2022), pp. 1–31.

[57] Lucy Anderson-Bashan, Benny Megidish, Hadas Erel, Iddo Wald, Andrey Grishko, Guy Hoffman, and Oren Zuckerman. "The Greeting Machine: An Abstract Robotic Object for Opening Encounters." In: June 2018. DOI: 10.1109/ROMAN.2018.8525516.

[58] Francisco Rubio, Francisco Valero, and Carlos Llopis-Albert. "A review of mobile robots: Concepts, methods, theoretical framework, and applications." In: *International Journal of Advanced Robotic Systems* 16.2 (2019).

[59] Abdelfetah Hentout, Mustapha Aouache, Abderraouf Maoudj, and Isma Akli. "Human–robot interaction in industrial collaborative robotics: a literature review of the decade 2008–2017." In: *Advanced Robotics* 33.15-16 (2019), pp. 764–799.

[60] Panagiota Tsarouchi, Sotiris Makris, and George Chryssolouris. "Human–robot interaction review and challenges on task planning and programming." In: *International Journal of Computer Integrated Manufacturing* 29.8 (2016), pp. 916–931.

[61] Milos Vasic and Aude Billard. "Safety issues in human-robot interactions." In: *2013 IEEE International Conference on Robotics and Automation*. 2013, pp. 197–204. DOI: 10.1109/ICRA.2013.6630576.

[62] Sharath Akkaladevi, Matthias Propst, Michael Hofmann, Leopold Hiesmair, Markus Ikeda, Naresh Chitturi, and Andreas Pichler. "Programming-Free Approaches for Human–Robot Collaboration in Assembly Tasks." In: June 2021, pp. 283–317. ISBN: 978-3-030-69177-6. DOI: 10.1007/978-3-030-69178-3_12.

[63] Zahra Rezaei Khavas, S Reza Ahmadzadeh, and Paul Robinette. "Modeling trust in human-robot interaction: A survey." In: *International Conference on Social Robotics*. Springer. 2020, pp. 529–541.

[64] Mary Ellen Foster. "Natural language generation for social robotics: opportunities and challenges." In: *Philosophical Transactions of the Royal Society B* 374.1771 (2019), p. 20180027.

[65] Thomas B Sheridan. "Human–robot interaction: status and challenges." In: *Human factors* 58.4 (2016), pp. 525–532.

[66] Gaaitzen J. de Vries, Elisabetta Gentile, Sébastien Miroudot, and Konstantin M. Wacker. "The rise of robots and the fall of routine jobs." In: *Labour Economics* 66 (2020), p. 101885.

[67] Murphy Wonsick, Philip Long, Aykut Onol, Maozhen Wang, and Taskin Padir. "A Holistic Approach to Human-Supervised Humanoid Robot Operations in Extreme Environments." In: *Frontiers in Robotics and AI* 8 (June 2021). DOI: 10.3389/frobt.2021.550644.

[68] Lorenzo Sabattini et al. “The PAN-Robots Project: Advanced Automated Guided Vehicle Systems for Industrial Logistics.” In: *IEEE Robotics & Automation Magazine* PP (Nov. 2017), pp. 1–1.

[69] Anastasia K. Ostrowski, Christina N. Harrington, Cynthia Breazeal, and Hae Won Park. “Personal Narratives in Technology Design: The Value of Sharing Older Adults’ Stories in the Design of Social Robots.” In: *Frontiers in Robotics and AI* 8 (2021).

[70] Ben Tribelhorn and Zachary Dodds. “Evaluating the Roomba: A low-cost, ubiquitous platform for robotics research and education.” In: *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE. 2007, pp. 1393–1399.

[71] Ryan Weber. *Seeing like a Rover: How Robots, Teams, and Images Craft Knowledge of Mars*. 2018.

[72] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. “Computers are social actors.” In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1994, pp. 72–78.

[73] Riccardo Campa. “The rise of social robots: a review of the recent literature.” In: *Journal of Evolution and Technology* 26.1 (2016).

[74] Maartje MA de Graaf, Soumaya Ben Allouch, and Jan AGM Van Dijk. “What makes robots social?: A user’s perspective on characteristics for social human-robot interaction.” In: *International Conference on Social Robotics*. Springer. 2015, pp. 184–193.

[75] Anca D Dragan, Kenton CT Lee, and Siddhartha S Srinivasa. “Legibility and predictability of robot motion.” In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE. 2013, pp. 301–308.

[76] Henny Admoni and Brian Scassellati. “Social eye gaze in human-robot interaction: a review.” In: *Journal of Human-Robot Interaction* 6.1 (2017), pp. 25–63.

[77] Kim Baraka and Manuela M Veloso. "Mobile service robot state revealing through expressive lights: formalism, design, and evaluation." In: *International Journal of Social Robotics* 10.1 (2018), pp. 65–92.

[78] Cindy L Bethel and Robin R Murphy. "Auditory and Other Non-verbal Expressions of Affect for Robots." In: *AAAI Fall Symposium: Aurally Informed Performance*. 2006, pp. 1–5.

[79] Diana Löffler, Nina Schmidt, and Robert Tscharn. "Multimodal expression of artificial emotion in social robots using color, motion and sound." In: *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE. 2018, pp. 334–343.

[80] Nikolaos Mavridis. "A review of verbal and non-verbal human–robot interactive communication." In: *Robotics and Autonomous Systems* 63 (2015), pp. 22–35.

[81] Ana Paiva, Iolanda Leite, Hana Boukricha, and Ipke Wachsmuth. "Empathy in virtual agents and robots: A survey." In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7.3 (2017), pp. 1–40.

[82] Silvia Rossi, Francois Ferland, and Adriana Tapus. "User profiling and behavioral adaptation for HRI: A survey." In: *Pattern Recognition Letters* 99 (2017), pp. 3–12.

[83] Jesse Fox and Andrew Gambino. "Relationship Development with Humanoid Social Robots: Applying Interpersonal Theories to Human–Robot Interaction." In: *Cyberpsychology, Behavior, and Social Networking* 24.5 (2021). PMID: 33434097, pp. 294–299. DOI: 10.1089/cyber.2020.0181.

[84] Maila Pentucci. "Form@re -Open Journal per la formazione in rete." In: 2 (Aug. 2016), pp. 271–289. DOI: 10.13128/formare-18319.

[85] Fiorella Operto. "Ethics in advanced robotics." In: *IEEE robotics & automation magazine* 18.1 (2011), pp. 72–78.

[86] Peter M Asaro. "What should we want from a robot ethic?" In: *Machine Ethics and Robot Ethics*. Routledge, 2020, pp. 87–94.

[87] Laurel Riek and Don Howard. "A code of ethics for the human-robot interaction profession." In: *Proceedings of we robot* (2014).

[88] David J Gunkel. *Robot rights*. mit Press, 2018.

[89] Wendell Wallach and Colin Allen. *Moral machines: Teaching robots right from wrong*. Oxford University Press, 2008.

[90] Przemyslaw A Lasota, Terrence Fong, Julie A Shah, et al. "A survey of methods for safe human-robot interaction." In: *Foundations and Trends® in Robotics* 5.4 (2017), pp. 261–349.

[91] Pieter Wolfert, Jorre Deschuyteneer, Djamari Oetringer, Nicole Robinson, and Tony Belpaeme. "Security Risks of Social Robots Used to Persuade and Manipulate: A Proof of Concept Study." In: Mar. 2020, pp. 523–525. DOI: 10.1145/3371382.3378341.

[92] Solomon E Asch. "Opinions and social pressure." In: *Scientific American* 193.5 (1955), pp. 31–35.

[93] Anna-Lisa Vollmer, Robin Read, Dries Trippas, and Tony Belpaeme. "Children conform, adults resist: A robot group induced peer pressure on normative social conformity." In: *Science robotics* 3.21 (2018), eaat7111.

[94] Berardina De Carolis, Nicola Macchiarulo, and Giuseppe Palestra. "Soft biometrics for social adaptive robots." In: *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer. 2019, pp. 687–699.

[95] Ronald Leenes, Erica Palmerini, Bert-Jaap Koops, Andrea Bertolini, Pericle Salvini, and Federica Lucivero. "Regulatory challenges of robotics: Some guidelines for addressing legal and ethical issues." In: *Law, Innovation and Technology* 9.1 (May 2017), pp. 1–44. ISSN: 1757-9961.

[96] Andreas Matthias. "Robot lies in health care: When is deception morally permissible?" In: *Kennedy Institute of Ethics Journal* 25.2 (2015), pp. 169–162.

[97] Serena Booth, James Tompkin, Hanspeter Pfister, Jim Waldo, Krzysztof Gajos, and Radhika Nagpal. "Piggybacking Robots: Human-Robot Overtrust in University Dormitory Security." In: *HRI '17: ACM/IEEE*. HRI '17. Vienna, Austria, 2017, pp. 426–434. ISBN: 978-1-4503-4336-7.

[98] Bertram F Malle, Matthias Scheutz, Thomas Arnold, John Voiklis, and Corey Cusimano. "Sacrifice one for the good of many? People apply different moral norms to human and robot agents." In: *HRI 2015*. IEEE. 2015, pp. 117–124.

[99] C. Hadnagy and P. Wilson. *Social Engineering: The Art of Human Hacking*. Wiley, 2010. ISBN: 9781118029718.

[100] Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. "Would You Trust a (Faulty) Robot? Effects of Error, Task Type and Personality on Human-Robot Cooperation and Trust." In: *2015 10th IEEE International Conference on Human-Robot Interaction (HRI)*. 2015, pp. 1–8.

[101] Denise Y Geiskkovitch, Derek Cormier, Stela H Seo, and James E Young. "Please continue, we need more data: an exploration of obedience to robots." In: *Journal of Human-Robot Interaction* 5.1 (2016), pp. 82–99.

[102] Arash Ajoudani, Andrea Maria Zanchettin, Serena Ivaldi, Alin Albu-Schäffer, Kazuhiro Kosuge, and Oussama Khatib. "Progress and prospects of the human–robot collaboration." In: *Autonomous Robots* 42.5 (2018), pp. 957–975.

[103] Miki Watanabe, Kohei Ogawa, and Hiroshi Ishiguro. "Can Androids Be Salespeople in the Real World?" In: *CHI '15*. ACM, 2015, 781–788.

[104] Shih Huan Tseng, Ju-Hsuan Hua, Shao-Po Ma, and Li-ehen Fu. "Human awareness based robot performance learning in a social environment." In: *ICRA 2013*. IEEE. 2013, pp. 4291–4296.

[105] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[106] Anna-Lisa Vollmer, Robin Read, Dries Trippas, and Tony Belpaeme. "Children conform, adults resist: A robot group induced peer pressure on normative social conformity." In: *Science Robotics* 3.21 (2018).

[107] Matthew Rueben, Alexander Mois Aroyo, Christoph Lutz, Johannes Schmölz, Pieter Van Cleynenbreugel, Andrea Corti, Siddharth Agrawal, and William D Smart. "Themes and research directions in privacy-sensitive robotics." In: *ARSO 2018*. IEEE. 2018, pp. 77–84.

[108] Belpaeme Tony, Deschuyteneer Jorre, Oetringer Djamari, and Pieter Wolfertt. *The potential of social robots for persuasion and manipulation: a proof of concept study*. https://securelist.com/robots-social-impact/94431/. Kaspersky, 2019.

[109] Alexander Mois Aroyo, Francesco Rea, Giulio Sandini, and Alessandra Sciutti. "Trust and social engineering in human robot interaction: Will a robot make you disclose sensitive information, conform to its recommendations or gamble?" In: *IEEE Robotics and Automation Letters* 3.4 (2018), pp. 3701–3708.

[110] Andrea F Abate, Paola Barra, Carmen Bisogni, Lucia Cascone, and Ignazio Passero. "Contextual trust model with a humanoid robot defense for attacks to smart eco-systems." In: *IEEE access* 8 (2020), pp. 207404–207414.

[111] Manuel Díaz, Cristian Martín, and Bartolomé Rubio. "State-of-the-art, challenges, and open issues in the integration of Internet of things and cloud computing." In: *Journal of Network and Computer applications* 67 (2016), pp. 99–117.

[112] Fan Wu, Taiyang Wu, and Mehmet Rasit Yuce. "Design and implementation of a wearable sensor network system for IoT-connected safety and health applications." In: *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*. IEEE. 2019, pp. 87–90.

[113] Aqeel Kazmi, Elias Tragos, and Martin Serrano. "Underpinning IoT for road traffic noise management in smart cities." In: *2018 IEEE international conference on pervasive computing and communications workshops (PerCom Workshops)*. IEEE. 2018, pp. 765–769.

[114] M Bala Krishna and Anudit Verma. "A framework of smart homes connected devices using Internet of Things." In: *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*. IEEE. 2016, pp. 810–815.

[115] MR Rashmi and C Vidya Raj. "A review on trust models of social Internet of Things." In: *Emerging Research in Electronics, Computer Science and Technology* (2019), pp. 203–209.

[116] Fenye Bao and Ray Chen. "Trust management for the internet of things and its application to service composition." In: *2012 IEEE international symposium on a world of wireless, mobile and multimedia networks (WoWMoM)*. IEEE. 2012, pp. 1–6.

[117] Pasquale De Meo, Katarzyna Musial-Gabrys, Domenico Rosaci, Giuseppe ML Sarne, and Lora Aroyo. "Using centrality measures to predict helpfulness-based reputation in trust networks." In: *ACM Transactions on Internet Technology (TOIT)* 17.1 (2017), pp. 1–20.

[118] Hannan Xiao, Nitin Sidhu, and Bruce Christianson. "Guarantor and reputation based trust model for social internet of things." In: *2015 International wireless communications and mobile computing conference (IWCMC)*. IEEE. 2015, pp. 600–605.

[119] Scott Rose, Oliver Borchert, Stu Mitchell, and Sean Connelly. *Zero trust architecture*. Tech. rep. National Institute of Standards and Technology, 2020.

[120] Lucia Cascone, Michele Nappi, Fabio Narducci, and Ignazio Passero. "DTPAAL: Digital twinning pepper and ambient assisted living." In: *IEEE Transactions on Industrial Informatics* 18.2 (2021), pp. 1397–1404.

[121] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee. "Digital Twin in Industry: State-of-the-Art." In: *IEEE Transactions on Industrial Informatics* 15.4 (2019), pp. 2405–2415.

[122] Lei Ren, Zihao Meng, Xiaokang Wang, Lin Zhang, and Laurence T Yang. "A Data-driven Approach of Product Quality Prediction for Complex Production Systems." In: *IEEE Transactions on Industrial Informatics* (2020).

[123] M. Schluse, M. Priggemeyer, L. Atorf, and J. Rossmann. "Experimentable Digital Twins—Streamlining Simulation-Based Systems Engineering for Industry 4.0." In: *IEEE Transactions on Industrial Informatics* 14.4 (2018), pp. 1722–1731.

[124] Niki Kousi, Christos Gkournelos, Sotiris Aivaliotis, Christos Giannoulis, George Michalos, and Sotiris Makris. "Digital twin for adaptation of robots' behavior in flexible robotic assembly lines." In: *Procedia Manufacturing* 28 (2019). 7th International conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV2018), pp. 121 –126. ISSN: 2351-9789. DOI: https://doi.org/10.1016/j.promfg.2018.12.020.

[125] Klaus Dröder, Paul Bobka, Tomas Germann, Felix Gabriel, and Franz Dietrich. "A machine learning-enhanced digital twin approach for human-robot-collaboration." In: *Procedia Cirp* 76 (2018), pp. 187–192.

[126] A. Bécue, Y. Fourastier, I. Praça, A. Savarit, C. Baron, B. Gradussofs, E. Pouille, and C. Thomas. "CyberFactory#1 — Securing the industry 4.0 with cyber-ranges and digital twins." In: *2018 14th IEEE International Workshop on Factory Communication Systems (WFCS)*. 2018, pp. 1–4.

[127] F. Biesinger, D. Meike, B. Kraß, and M. Weyrich. "A Case Study for a Digital Twin of Body-in-White Production Systems General Concept for Automated Updating of Planning Projects in the Digital Factory." In: *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*. Vol. 1. 2018, pp. 19–26.

[128] M Mazhar Rathore, Awais Ahmad, Anand Paul, Jiafu Wan, and Daqiang Zhang. "Real-time medical emergency response system: exploiting IoT and big data for public health." In: *Journal of medical systems* 40.12 (2016), p. 283.

[129] H. Laaki, Y. Miche, and K. Tammi. "Prototyping a Digital Twin for Real Time Remote Control Over Mobile Networks: Application of Remote Surgery." In: *IEEE Access* 7 (2019), pp. 20325–20336.

[130] Aitor Ardanza, Aitor Moreno, Álvaro Segura, Mikel de la Cruz, and Daniel Aguinaga. "Sustainable and flexible industrial human machine interfaces to support adaptable applications in the Industry 4.0 paradigm." In: *International Journal of Production Research* 57.12 (2019), pp. 4045–4059.

[131] M. Lacomblez, B. Jeanne, V. Havard, and D. Baudry. "Co-simulation architecture between a digital twin and a virtual reality environment in an industrial context." In: vol. 8. 2018, pp. 429–434.

[132] Antonio R Damasio. "Emotions and feelings." In: *Feelings and emotions: The Amsterdam symposium*. Cambridge University Press Cambridge. 2004, pp. 49–57.

[133] Paul Ekman, Wallace V. Friesen, and Phoebe Ellsworth. *Emotion in the human face: Guidelines for research and an integration of findings*. Pergamon Press, 1972.

[134] Matteo Spezialetti, Giuseppe Placidi, and Silvia Rossi. "Emotion recognition for human-robot interaction: Recent advances and future perspectives." In: *Frontiers in Robotics and AI* 7 (2020), p. 532279.

[135] Lola Canamero. "Embodied robot models for interdisciplinary emotion research." In: *IEEE Transactions on Affective Computing* 12.2 (2019), pp. 340–351.

[136] Silvia Rossi and Martina Ruocco. "Better alone than in bad company: effects of incoherent non-verbal emotional cues for a humanoid robot." In: *Interaction Studies* 20.3 (2019), pp. 487–508.

[137] Marc Weber Tobias. "Kate Middleton's Medical Information: The Larger Issue." In: *Forbes* (2012). URL: https://www.forbes.com/sites/marcwebertobias/2012/12/13/kate-middletons-medical-information-the-larger-issue/\#376b268d6f21X.

[138] Hedda Lausberg. *Understanding Body Movement: A Guide to Empirical Research on Nonverbal Behaviour-With an Introduction to the NEUROGES Coding System*. Peter Lang International Academic Publishers, 2013.

[139] Mark Coulson. "Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence." In: *Journal of nonverbal behavior* 28.2 (2004), pp. 117–139.

[140] Ferdous Ahmed, A. S. M. Hossain Bari, and Marina L. Gavrilova. "Emotion Recognition From Body Movement." In: *IEEE Access* 8 (2020), pp. 11761–11781. DOI: 10.1109/ACCESS.2019.2963113.

[141] F. Noroozi, D. Kaminska, C. Corneanu, T. Sapinski, S. Escalera, and G. Anbarjafari. "Survey on Emotional Body Gesture Recognition." In: *IEEE Trans. on Affective Computing* (2018), pp. 1–1. ISSN: 2371-9850.

[142] F. Ahmed, A. S. M. H. Bari, and M. L. Gavrilova. "Emotion Recognition From Body Movement." In: *IEEE Access* 8 (2020), pp. 11761–11781. ISSN: 2169-3536.

[143] Nourhan Elfaramawy, Pablo Barros, German I. Parisi, and Stefan Wermter. "Emotion Recognition from Body Expressions with a Neural Network Architecture." In: *HAI 2017*. ACM, 2017, 143–149. ISBN: 9781450351133.

[144] Hyisung Hwang and David Matsumoto. "Evidence for the Universality of Facial Expressions of Emotion." In: *Understanding Facial Expressions in Communication: Cross-cultural and Multidisciplinary Perspectives*. Ed. by Manas K. Mandal and Avinash Awasthi. Springer India, 2015, pp. 41–56.

[145] I Michael Revina and WR Sam Emmanuel. "A survey on human face expression recognition techniques." In: *Journal of King Saud University-Computer and Information Sciences* 33.6 (2021), pp. 619–628.

[146] Deepak Kumar Jain, Pourya Shamsolmoali, and Paramjit Sehdev. "Extended deep neural network for facial emotion recognition." In: *Pattern Recognition Letters* 120 (2019), pp. 69 –74. ISSN: 0167-8655.

[147] J. Deng, G. Pang, Z. Zhang, Z. Pang, H. Yang, and G. Yang. "cGAN Based Facial Expression Recognition for Human-Robot Interaction." In: *IEEE Access* 7 (2019), pp. 9848–9859. ISSN: 2169-3536.

[148] L. Chen, M. Li, W. Su, M. Wu, K. Hirota, and W. Pedrycz. "Adaptive Feature Selection-Based AdaBoost-KNN With Direct Optimization for Dynamic Emotion Recognition in Human–Robot Interaction." In: *IEEE Trans. on Emerging Topics in Computational Intelligence* (2019), pp. 1–9. ISSN: 2471-285X.

[149] Archana Verma and Lokesh Kumar Sharma. "A comprehensive survey on human facial expression detection." In: *International Journal of Image Processing (IJIP)* 7.2 (2013), p. 171.

[150] Sergey Tulyakov, Xavier Alameda-Pineda, Elisa Ricci, Lijun Yin, Jeffrey F. Cohn, and Nicu Sebe. "Self-Adaptive Matrix Completion for Heart Rate Estimation From Face Videos Under Realistic Conditions." In: *CVPR 2016*. 2016.

[151] Gerald Bieber, Niklas Antony, and Marian Haescher. "Touchless Heart Rate Recognition by Robots to Support Natural Human-Robot Communication." In: *PETRA '18*. ACM, 2018, 415–420. ISBN: 9781450363907.

[152] Moataz El Ayadi, Mohamed S Kamel, and Fakhri Karray. "Survey on speech emotion recognition: Features, classification schemes, and databases." In: *Pattern recognition* 44.3 (2011), pp. 572–587.

[153] Joy Nicholson, Kazuhiko Takahashi, and Ryohei Nakatsu. "Emotion recognition in speech using neural networks." In: *Neural computing & applications* 9.4 (2000), pp. 290–296.

[154] Rainer Banse and Klaus R Scherer. "Acoustic profiles in vocal emotion expression." In: *Journal of personality and social psychology* 70.3 (1996), p. 614.

[155] Carl E Williams and Kenneth N Stevens. "Vocal correlates of emotional states." In: *Speech evaluation in psychiatry* (1981), pp. 221–240.

[156] Janet E Cahn. "The generation of affect in synthesized speech." In: *Journal of the American voice I/O society* 8.1 (1990), pp. 1–1.

[157] Jackson J Liscombe. *Prosody and speaker state: paralinguistics, pragmatics, and proficiency*. Columbia University, 2007.

[158] Y. Li, C. T. Ishi, N. Ward, K. Inoue, S. Nakamura, K. Takanashi, and T. Kawahara. "Emotion recognition by combining prosody and sentiment analysis for expressing reactive emotion by humanoid robot." In: *APSIPA ASC 2017*. 2017, pp. 1356–1359.

[159] Carlos Ramos, Juan Carlos Augusto, and Daniel Shapiro. "Ambient intelligence—the next step for artificial intelligence." In: *IEEE Intelligent Systems* 23.2 (2008), pp. 15–18.

[160] Jan Vanus, Jana Belesova, Radek Martinek, Jan Nedoma, Marcel Fajkus, Petr Bilik, and Jan Zidek. "Monitoring of the daily living activities in smart home care." In: *Human-centric Computing and Information Sciences* 7.1 (2017), pp. 1–34.

[161] Berardina De Carolis, Stefano Ferilli, and Nicola Macchiarulo. "Ambient Assisted Living and Social Robots: Towards Learning Relations between User's Daily Routines and Mood." In: *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. 2022, pp. 123–129.

[162] R. Minerva, G. M. Lee, and N. Crespi. "Digital Twin in the IoT Context: A Survey on Technical Features, Scenarios, and Architectural Models." In: *Proceedings of the IEEE* (2020), pp. 1–40.

[163] Louise Wright and Stuart Davidson. "How to tell the difference between a model and a digital twin." In: *Advanced Modeling and Simulation in Engineering Sciences* 7.1 (2020), pp. 1–13.

[164] Lucia Cascone, Aniello Castiglione, Michele Nappi, Fabio Narducci, and Ignazio Passero. "Waiting for Tactile: Robotic and Virtual Experiences in the Fog." In: *ACM Transactions on Internet Technology (TOIT)* 21.3 (2021), pp. 1–19.

[165] *Proximal Policy Optimization*. https://openai.com/blog/openai-baselines-ppo/. Accessed: 2020-12-10.

[166] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. "Proximal policy optimization algorithms." In: *arXiv preprint arXiv:1707.06347* (2017).

[167] Saad Khan, Simon Parkinson, Liam Grant, Na Liu, and Stephen Mcguire. "Biometric systems utilising health data from wearable devices: applications and future challenges in computer security." In: *ACM Computing Surveys (CSUR)* 53.4 (2020), pp. 1–29.

[168] Ruggero Donida Labati, Vincenzo Piuri, Francesco Rundo, and Fabio Scotti. "Photoplethysmographic biometrics: A comprehensive survey." In: *Pattern Recognition Letters* (2022).

[169] Mario Mekhael, Chan Ho Lim, Abdel Hadi El Hajjar, Charbel Noujaim, Christopher Pottle, Noor Makan, Lilas Dagher, Yichi Zhang, Nour Chouman, Dan L Li, et al. "Studying the Effect of Long COVID-19 Infection on Sleep Quality Using Wearable Health Devices: Observational Study." In: *Journal of Medical Internet Research* 24.7 (2022), e38000.

[170] Jesse Fine, Kimberly L Branan, Andres J Rodriguez, Tananant Boonya-Ananta, Jessica C Ramella-Roman, Michael J McShane, and Gerard L Coté. "Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring." In: *Biosensors* 11.4 (2021), p. 126.

[171] Binghui Lin, Zhouchen Ma, Mohamed Atef, Liang Ying, and Guoxing Wang. "Low-power high-sensitivity photoplethysmography sensor for wearable health monitoring system." In: *IEEE Sensors Journal* 21.14 (2021), pp. 16141–16151.

[172] Jorge Sancho, Alvaro Alesanco, and Jose Garca. "Photoplethysmographic authentication in long-term scenarios: A preliminary assessment." In: *EMBEC & NBC 2017*. Springer, 2017, pp. 1085–1088.

[173] Jason Leake, Nigel Harris, Edmund Keogh, and Christopher Eccleston. "Is photoplethysmography-derived pulse shape useful for fall detection?" In: *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*. 2014, pp. 1–8.

[174] Sophini Subramaniam, Abu Ilius Faisal, and M Jamal Deen. "Wearable sensor systems for fall risk assessment: a review." In: *Frontiers in digital health* 4 (2022).

[175] Kimberley S Van Schooten, Mirjam Pijnappels, Sietse M Rispens, Petra JM Elders, Paul Lips, and Jaap H van Dieën. "Ambulatory fall-risk assessment: amount and quality of daily-life gait predict falls in older adults." In: *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* 70.5 (2015), pp. 608–615.

[176] Marianna Amboni, Paolo Barone, and Jeffrey M Hausdorff. "Cognitive contributions to gait and falls: evidence and implications." In: *Movement disorders* 28.11 (2013), pp. 1520–1533.

[177] Neamah Al-Naffakh, Nathan Clarke, and Fudong Li. "Continuous user authentication using smartwatch motion sensor data." In: *IFIP International Conference on Trust Management*. Springer. 2018, pp. 15–28.

[178] Alessandra Rossi, Patrick Holthaus, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L Walters. "Getting to know Pepper: Effects of people's awareness of a robot's capabilities on their trust in the robot." In: *Proceedings of the 6th international conference on human-agent interaction*. 2018, pp. 246–252.

[179] Daniel Ullman and Bertram F Malle. "What does it mean to trust a robot? Steps toward a multidimensional measure of trust." In: *Companion of the 2018 acm/ieee international conference on human-robot interaction*. 2018, pp. 263–264.

[180] Harold Soh, Yaqi Xie, Min Chen, and David Hsu. "Multitask trust transfer for human–robot interaction." In: *The International Journal of Robotics Research* 39.2-3 (2020), pp. 233–249.

[181] Samuele Vinanzi, Massimiliano Patacchiola, Antonio Chella, and Angelo Cangelosi. "Would a robot trust you? Developmental robotics model of trust and theory of mind." In: *Philosophical Transactions of the Royal Society B* 374.1771 (2019), p. 20180032.

[182] Farag Azzedin. "Taxonomy of reputation assessment in peer-to-peer systems and analysis of their data retrieval." In: *The Knowledge Engineering Review* 29.4 (2014), pp. 463–483.

[183] Olasimbo Ayodeji Arigbabu, Sharifah Mumtazah Syed Ahmad, Wan Azizun Wan Adnan, and Salman Yussof. "Recent advances in facial soft biometrics." In: *The Visual Computer* 31.5 (2015), pp. 513–525.

[184] Andrea F Abate, Paola Barra, Silvio Barra, Cristiano Molinari, Michele Nappi, and Fabio Narducci. "Clustering facial attributes: Narrowing the path from soft to hard biometrics." In: *IEEE Access* 8 (2019), pp. 9037–9045.

[185] Changsheng Wan, Li Wang, and Vir V Phoha. "A survey on gait recognition." In: *ACM Computing Surveys (CSUR)* 51.5 (2018), pp. 1–35.

[186] Kun Ha Suh and Eui Chul Lee. "Face liveness detection for face recognition based on cardiac features of skin color image." In: *First International Workshop on Pattern Recognition*. Vol. 10011. SPIE. 2016, pp. 62–66.

[187] John G. Lawrenson, Rosalind Birhah, and Paul J. Murphy. "Tear-film lipid layer morphology and corneal sensation in the development of blinking in neonates and infants." In: (2005).

[188] Jacob Klenø Nøjgaard, Karl Bang Christensen, and Peder Wolkoff. "The effect on human eye blink frequency of exposure to limonene oxidation products and methacrolein." In: (2005).

[189] von Cramon D. and Schuri U. "Blink frequency and speech motor activity." In: *Neuropsychologia* (1980). DOI: https://doi.org/10.1016/0028-3932(80)90164-5.

[190] F. Yamada. "Frontal midline theta rhythm and eyeblinking activity during a VDT task and a video game: useful tools for psychophysiology in ergonomics." In: *Ergonomics* 41.5 (1998), pp. 678–688. DOI: 10.1080/001401398186847.

[191] J. Hori, K. Sakano, and Y. Saitoh. "Development of a communication support device controlled by eye movements and voluntary eye blink." In: *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2004), pp. 4302–4305. DOI: 10.1109/IEMBS.2004.1404198.

[192] Craig N. Karson, Karen Faith Berman, Edward F. Donnelly, Wallace B. Mendelson, Joel E. Kleinman, and Richard Jed Wyatt. "Speaking, thinking, and blinking." In: *Psychiatry research* 5.3 (1981), pp. 243–246. DOI: 10.1016/0165-1781(81)90070-6.

[193] Saroj KL Lal and Ashley Craig. "Driver fatigue: electroencephalography and psychological assessment." In: *Psychophysiology* 39.3 (2002), pp. 313–321.

[194] PW Gray, RP Allen, ML Thomas, DR Thorne, HC Sing, and DH Quigg. "Eye blink rates during a simulated driving task before and after sleep deprivation." In: *Sleep Res* 24 (1995), p. 445.

[195] Lindsey K McIntire, R Andy McKinley, Chuck Goodyear, and John P McIntire. "Detection of vigilance performance using eye blinks." In: *Applied ergonomics* 45.2 (2014), pp. 354–362.

[196] Matthew Luckiesh and Frank K Moss. "The eyelid reflex as a criterion of ocular fatigue." In: *Journal of experimental Psychology* 20.6 (1937), p. 589.

[197] Sylvia Peißl, Christopher D Wickens, and Rithi Baruah. "Eye-tracking measures in aviation: A selective literature review." In: *The International Journal of Aerospace Psychology* 28.3-4 (2018), pp. 98–112.

[198] A Carpenter. "The rate of blinking during prolonged visual search." In: *Journal of Experimental Psychology* 38.5 (1948), p. 587.

[199] Tsugunosuke Sakai, Harunya Tamaki, Yosuke Ota, Ryohei Egusa, Shigenori Imagaki, Fusako Kusunoki, Masanori Sugimoto, and Hiroshi Mizoguchi. "Eda-based estimation of visual attention by observation of eye blink frequency."

In: *International Journal on Smart Sensing and Intelligent Systems* 10.2 (2017), pp. 1–12.

[200] Stefanie E Kuchinsky, Kenneth I Vaden Jr, Jayne B Ahlstrom, Stephanie L Cute, Larry E Humes, Judy R Dubno, and Mark A Eckert. "Task-related vigilance during word recognition in noise for older adults with hearing loss." In: *Experimental aging research* 42.1 (2016), pp. 50–66.

[201] Nanda NJ Rommelse, Stefan Van der Stigchel, and Joseph A Sergeant. "A review on eye movement studies in childhood and adolescent psychiatry." In: *Brain and cognition* 68.3 (2008), pp. 391–414.

[202] Sarah Uzzaman and Steve Joordens. "The eyes know what you are thinking: Eye movements as an objective measure of mind wandering." In: *Consciousness and cognition* 20.4 (2011), pp. 1882–1886.

[203] Erik J Sirevaag and John A Stern. "Ocular measures of fatigue and cognitive factors." In: *Engineering psychophysiology: Issues and applications* (2000), pp. 269–287.

[204] Brian R Cornwell, Sven C Mueller, Raphael Kaplan, Christian Grillon, and Monique Ernst. "Anxiety, a benefit and detriment to cognition: behavioral and magnetoencephalographic evidence from a mixed-saccade task." In: *Brain and cognition* 78.3 (2012), pp. 257–267.

[205] MW Schurgin, J Nelson, S Iida, H Ohira, JY Chiao, and SL Franconeri. "Eye movements during emotion recognition in faces." In: *Journal of vision* 14.13 (2014), pp. 14–14.

[206] Susan B Perlman, James P Morris, Brent C Vander Wyk, Steven R Green, Jaime L Doyle, and Kevin A Pelphrey. "Individual differences in personality predict how people look at faces." In: *PloS one* 4.6 (2009), e5952.

[207] Kai Preuss, Christopher Hilton, Klaus Gramann, and Nele Russwinkel. "Cognitive Processing Stages During Mental Folding Are Reflected in Eye Movements." In: *ACM Symposium on Eye Tracking Research and Applications*. 2020, pp. 1–5.

[208] Samuel Salvaggio, Nicolas Masson, Alexandre Zénon, and Michael Andres. "The predictive role of eye movements in mental arithmetic." In: *Experimental Brain Research* 240.5 (2022), pp. 1331–1340.

[209] Liping Pang, Yurong Fan, Ye Deng, Xin Wang, and Tianbo Wang. "Mental Workload Classification By Eye Movements In Visual Search Tasks." In: *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. 2020, pp. 29–33. DOI: 10.1109/CISP-BMEI51763.2020.9263668.

[210] Jia Zheng Lim, James Mountstephens, and Jason Teo. "Eye-Tracking Feature Extraction for Biometric Machine Learning." In: *Frontiers in Neurorobotics* 15 (2021).

[211] Sandeep Vidyapu, Vijaya Saradhi Vedula, and Samit Bhattacharya. "Attention prediction on webpage images using multilabel classification." In: *2019 IEEE international conference on systems, man and cybernetics (SMC)*. IEEE. 2019, pp. 1975–1980.

[212] Anup Kumar Roy, Md Nadeem Akhtar, Manjunatha Mahadevappa, Rajlakshmi Guha, and Jayanta Mukherjee. "A novel technique to develop cognitive models for ambiguous image identification using eye tracker." In: *IEEE Transactions on Affective Computing* 11.1 (2017), pp. 63–77.

[213] Antes James and Kristjanson Arlinda. "Discriminating Artists from Nonartists by Their Eye-Fixation Patterns." In: *Perceptual and motor skills* 73 (Jan. 1992), pp. 893–4. DOI: 10.2466/PMS.73.7.893-894.

[214] Just Marcel and Carpenter Patricia. "A theory of reading: From eye fixations to comprehension." In: *Psychological review* 87 (Aug. 1980), pp. 329–54. DOI: 10.1037/0033-295X.87.4.329.

[215] Millen Ailsa, Hope Lorraine, Hillstrom Anne, and Vrij Aldert. "Tracking the truth: The effect of face familiarity on eye fixations during deception." In: *The Quarterly Journal of Experimental Psychology* 70 (Feb. 2016). DOI: 10.1080/17470218.2016.1172093.

[216] Moutinho Raquel, Castro São, and Silva Susana. "Look at me: The relation between empathy and fixation on the emotional eye-region in low vs. high social anxiety." In: *Journal of Behavior Therapy and Experimental Psychiatry* 70 (Aug. 2020), p. 101610. DOI: 10.1016/j.jbtep.2020.101610.

[217] Mock Julia, Huber Stefan, Klein Elise, and Moeller Korbinian. "Insights into numerical cognition: considering eye-fixations in number processing and arithmetic." In: *Psychological research* 80 (Feb. 2016). DOI: 10.1007/s00426-015-0739-9.

[218] Sekiguchi Takahiro. "Individual differences in face memory and eye fixation patterns during face learning." In: *Acta psychologica* 137 (Feb. 2011), pp. 1–9. DOI: 10.1016/j.actpsy.2011.01.014.

[219] A. W. Eriksson, J. Fellman, H. Nieminen, and H. Forsius. "Influence Of Age On The Position And Size Of The Iris Frill And The Pupil." In: *Acta Ophthalmologica* 43.5 (1965), 629–641. DOI: 10.1111/j.1755-3768.1965.tb00334.x.

[220] Sourabh Sharma, Mani Baskaran, Annadata V. Rukmini, Monisha E. Nongpiur, HlaMyint Htoon, Ching-Yu Cheng, Shamira A. Perera, Joshua J. Gooley, Tin Aung, and Dan Milea. "Factors influencing the pupillary light reflex in healthy individuals." In: *Graefe's Archive for Clinical and Experimental Ophthalmology* 254 (7 July 2016). DOI: 10.1007/s00417-016-3311-4.

[221] Michel Guillon, Kathryn Dumbleton, Panagiotis Theodoratos, Marine Gobbe, C. Benjamin Wooley, and Kurt Moody. "The Effects of Age, Refractive Status, and Luminance on Pupil Size." In: *Optometry and Vision Science* 93.9 (2016), 1093–1100. DOI: 10.1097/opx.0000000000000893.

[222] Richard Feinberg and Edward Podolak. "Latency of pupillary reflex to light stimulation and its relationship to aging." In: *Behavior, Aging and The Nervous System* (1965).

[223] Barry Winn, David Whitaker, David Elliott, and N.J. Phillips. "Factors affecting light-adapted pupil size in normal

human subjects." In: *Investigative ophthalmology & visual science* 35 (Mar. 1994), pp. 1132–7.

[224] D Fotiou, Catherine Brozou, Dimitrios Tsiptsios, A Fotiou, A Kabitsi, M Nakou, Ch Giantselidis, and A. Goulas. "Effect of age on pupillary light reflex: Evaluation of pupil mobility for clinical practice and research." In: *Electromyography and clinical neurophysiology* 47 (July 2015), pp. 11–22.

[225] Eckhard H Hess. "Attitude and pupil size." In: *Scientific american* 212.4 (1965), pp. 46–55.

[226] Qing-Xing Qu and Fu Guo. "Can eye movements be effectively measured to assess product design?: Gender differences should be considered." In: *International Journal of Industrial Ergonomics* 72 (2019), 281–289. DOI: 10.1016/j.ergon.2019.06.006.

[227] Unsang Park, Raghavender Reddy Jillela, Arun Ross, and Anil K Jain. "Periocular biometrics in the visible spectrum." In: *IEEE Transactions on Information Forensics and Security* 6.1 (2010), pp. 96–106.

[228] V.M. Ipe and T. Thomas. "Periocular Recognition Under Unconstrained Image Capture Distances." In: *Communications in Computer and Information Science* 1365 (2021), pp. 175–186.

[229] Andrea F. Abate, Lucia Cimmino, Fabio Narducci, and Chiara Pero. "Biometric Face Recognition Based on Landmark Dynamics." In: *2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*. 2020, pp. 601–605. DOI: 10.1109/DASC-PICom-CBDCom-CyberSciTech49142.2020.00105.

[230] Oleg V Komogortsev, Sampath Jayarathna, Cecilia R Aragon, and Mechehoul Mahmoud. "Biometric identification via an oculomotor plant mathematical model." In: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. 2010, pp. 57–60.

[231] Andrea F Abate, Lucia Cascone, Michele Nappi, Fabio Narducci, and Ignazio Passero. "Attention monitoring for synchronous distance learning." In: *Future Generation Computer Systems* 125 (2021), pp. 774–784.

[232] Pawel Kasprowski. "Identifying users based on their eye tracker calibration data." In: Apr. 2020. DOI: 10.1145/3379157.3391419.

[233] Guang Chen, Fa Wang, Xiaoding Yuan, Zhijun Li, Zichen Liang, and Alois Knoll. "NeuroBiometric: an eye blink based biometric authentication system using an event-based neuromorphic vision sensor." In: *IEEE/CAA Journal of Automatica Sinica* 8.1 (2020), pp. 206–218.

[234] Pawel Kasprowski and Józef Ober. "Eye movements in biometrics." In: *International Workshop on Biometric Authentication*. Springer. 2004, pp. 248–258.

[235] Tomi Kinnunen, Filip Sedlak, and Roman Bednarik. "Towards task-independent person authentication using eye movement signals." In: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. 2010, pp. 187–190.

[236] Ioannis Rigas, George Economou, and Spiros Fotopoulos. "Biometric identification based on the eye movements and graph matching techniques." In: *Pattern Recognition Letters* 33.6 (2012), pp. 786–792.

[237] Corey Holland and Oleg V. Komogortsev. "Biometric identification via eye movement scanpaths in reading." In: *2011 International Joint Conference on Biometrics (IJCB)*. 2011, pp. 1–8. DOI: 10.1109/IJCB.2011.6117536.

[238] Xiaoming Wang, Xinbo Zhao, and Yanning Zhang. "Deep-learning-based reading eye-movement analysis for aiding biometric recognition." In: *Neurocomputing* 444 (2021), pp. 390–398.

[239] V. Cantoni, C. Galdi, M. Porta, D. Riccio, L. De Maio, R. Di Stasi, and M. Nappi. "Gaze recording - Free observation of human faces." In: 2020. DOI: http://dx.doi.org/10.17632/3kn4jdd4kf.2.

[240] Ioannis Rigas and Oleg V Komogortsev. "Biometric recognition via fixation density maps." In: *Biometric and Surveillance Technology for Human and Activity Identification XI*. Vol. 9075. SPIE. 2014, pp. 154–163.

[241] Wensi S Sun, Robert S Baker, John C Chuke, Bahman R Rouholiman, S Akbar Hasan, Wendy Gaza, Michael W Stava, and John D Porter. "Age-related changes in human blinks. Passive and active changes in eyelid kinematics." In: *Investigative ophthalmology & visual science* 38.1 (1997), pp. 92–99.

[242] Naoko Kimura, Akihide Watanabe, Kazutaka Suzuki, Haruyoshi Toyoda, Naotoshi Hakamata, Yoshiko Nakamura, and Shigeru Kinoshita. "Measurement of age-related changes in human blinks using a high-speed blink analysis system." In: *Nippon Ganka Gakkai Zasshi* 116.9 (2012), pp. 862–868.

[243] BW Ongerboer De Visser and C Goor. "Electromyographic and reflex study in idiopathic and symptomatic trigeminal neuralgias: latency of the jaw and blink reflexes." In: *Journal of Neurology, Neurosurgery & Psychiatry* 37.11 (1974), pp. 1225–1230.

[244] DM Corden and OC Lippold. "Age-related impaired reflex sensitivity in a human hand muscle." In: *Journal of neurophysiology* 76.4 (1996), pp. 2701–2706.

[245] Anitha Peddireddy, Kelun Wang, Peter Svensson, and Lars Arendt-Nielsen. "Influence of age and gender on the jaw-stretch and blink reflexes." In: *Experimental brain research* 171.4 (2006), pp. 530–540.

[246] Bahman Abdi Sargezeh, Niloofar Tavakoli, and Mohammad Reza Daliri. "Gender-based eye movement differences in passive indoor picture viewing: An eye-tracking study." In: *Physiology & behavior* 206 (2019), pp. 43–50.

[247] DP Munoz, JR Broughton, JE Goldring, and IT Armstrong. "Age-related performance of human subjects on saccadic eye movement tasks." In: *Experimental brain research* 121.4 (1998), pp. 391–400.

[248] Jochen Laubrock, Reinhold Kliegl, and Ralf Engbert. "SWIFT explorations of age differences in eye movements during reading." In: *Neuroscience & Biobehavioral Reviews* 30.6 (2006), pp. 872–884.

[249] Robert M. Stelmack and Nathan Mandelzys. "Extraversion and Pupillary Response to Affective and Taboo Words." In: *Psychophysiology* 12.5 (1975), 536–540. DOI: 10.1111/j.1469-8986.1975.tb00042.x.

[250] Timo Partala, Maria Jokiniemi, and Veikko Surakka. "Pupillary responses to emotionally provocative stimuli." In: *Proceedings of the symposium on Eye tracking research & applications - ETRA '00* (2000). DOI: 10.1145/355017.355042.

[251] Timo Partala and Veikko Surakka. "Pupil size variation as an indication of affective processing." In: *International Journal of Human-Computer Studies* 59.1-2 (2003), 185–198. DOI: 10.1016/s1071-5819(03)00017-x.

[252] Elena Geangu, Petra Hauf, Rishi Bhardwaj, and Wolfram Bentz. "Infant Pupil Diameter Changes in Response to Others' Positive and Negative Emotions." In: *PLoS ONE* 6.11 (2011). DOI: 10.1371/journal.pone.0027132.

[253] Santeri Yrttiaho, Dana Niehaus, Eileen Thomas, and Jukka M. Leppänen. "Mothers' pupillary responses to infant facial expressions." In: *Behavioral and Brain Functions* 13.1 (2017). DOI: 10.1186/s12993-017-0120-9.

[254] Patrick Gomez, Armin Von Gunten, and Brigitta Danuser. "Autonomic nervous system reactivity within the valence–arousal affective space: Modulation by sex and age." In: *International Journal of Psychophysiology* 109 (2016), 51–62.

[255] George Mather and Linda Murdoch. "Gender discrimination in biological motion displays based on dynamic cues." In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 258.1353 (1994), pp. 273–279.

[256] K Sudo, J Yamato, and A Tomono. "Determining gender of walking people using multiple sensors." In: *1996 IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems*. IEEE. 1996, pp. 641–646.

[257] Virginio Cantoni, Lucia Cascone, Michele Nappi, and Marco Porta. "Demographic classification through pupil analysis." In: *Image and Vision Computing* 102 (2020).

[258] Chiara Galdi, Harry Wechsler, Virginio Cantoni, Marco Porta, and Michele Nappi. "Towards demographic categorization using gaze analysis." In: *Pattern Recognition Letters* 82 (2016), pp. 226–231.

[259] Carmen Bisogni, Lucia Cascone, and Fabio Narducci. "Periocular Data Fusion for Age and Gender Classification." In: *Journal of Imaging* 8.11 (2022), p. 307.

[260] Marco F Huber, Andreas Merentitis, Roel Heremans, Maria Niessen, Christian Debes, and Nikolaos Frangiadakis. "Bayesian Score Level Fusion for Facial Recognition." In: ().

[261] Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli. "Evaluation of multimodal biometric score fusion rules under spoof attacks." In: *2012 5th IAPR International Conference on Biometrics (ICB)*. 2012, pp. 402–407. DOI: 10.1109/ICB.2012.6199784.

[262] W. Kim, J.M. Song, and K.R. Park. "Multimodal biometric recognition based on convolutional neural network by the fusion of finger-vein and finger shape using near-infrared (NIR) camera sensor." In: *Sensors (Switzerland)* 18.7 (2018).

[263] Andrea Casanova, Lucia Cascone, Aniello Castiglione, Wei-zhi Meng, and Chiara Pero. "User recognition based on periocular biometrics and touch dynamics." In: *Pattern Recognition Letters* 148 (2021), pp. 114–120.

[264] Mohamad El-Abed, Mostafa Dafer, and Ramzi El Khayat. "RHU Keystroke: A mobile-based benchmark for keystroke dynamics systems." In: *2014 International Carnahan conference on security technology (ICCST)*. IEEE. 2014, pp. 1–4.

[265] Lucia Cascone. "Periocular features as a window on cognitive processing: from memory to understanding the semantic complexity of an image." In: *2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART)*. 2021, pp. 1–4.

[266] Niklas Wilming, Selim Onat, José P. Ossandón, Alper Açik, Tim Christian Kietzmann, Kai Kaspar, Ricardo Ramos Gameiro, Alexandra Vormberg, and Peter König. "An extensive dataset of eye movements during viewing of complex images." In: *Scientific Data* 4 (2017).

[267] Becky Heaver and Sam Hutton. "Keeping an eye on the truth? Pupil size changes associated with recognition memory." In: *Memory (Hove, England)* 19 (May 2011), pp. 398–405. DOI: 10.1080/09658211.2011.575788.

[268] Anwesha Banerjee, Shreyasi Datta, Amit Konar, DN Tibarewala, and R Janarthanan. "Electrooculogram based detection of visual memory recall process." In: *2014 International Conference on Communication and Signal Processing*. IEEE. 2014, pp. 147–151.

[269] Andreas Bulling and Daniel Roggen. "Recognition of visual memory recall processes using eye movement analysis." In: *Proceedings of the 13th international conference on Ubiquitous computing*. 2011, pp. 455–464.

[270] Florian Marchal, Sylvain Castagnos, and Anne Boyer. "First Attempt to Predict User Memory from Gaze Data." In: *International Journal on Artificial Intelligence Tools* 27 (Sept. 2018), p. 1850029. DOI: 10.1142/S021821301850029X.

[271] Kai Kaspar and Peter König. "Overt attention and context factors: the impact of repeated presentations, image type, and individual motivation." In: *PloS one* 6.7 (2011), e21719.

[272] Ketan Rajshekhar Shahapure and Charles Nicholas. "Cluster Quality Analysis Using Silhouette Score." In: *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE. 2020, pp. 747–748.

[273] Alessandro Manzoni. *I promessi sposi*. Vol. 29. Newton Compton Editori, 2010.