Multimodal Biometric Systems for Personal Identification and Authentication using Machine and Deep Learning Classifiers

Thesis submitted to the University of Derby for the degree of
Doctor of Philosophy,
2022.

Sulaiman Alshebli

DoS: Mahmoud Shafik
First Supervisor: Fatih Kurugollu

Signature _____

Date _____ / _____ / _____

Consent

I declare that the thesis has been composed by myself and that the work has not be submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly authored publications has been included. My contribution and those of the other authors to this work have been explicitly indicated below. I confirm that appropriate credit has been given within this thesis where reference has been made to the work of others.

Abstract

Multimodal biometrics, using machine and deep learning, has recently gained interest over single biometric modalities. This interest stems from the fact that this technique improves recognition and, thus, provides more security. In fact, by combining the abilities of single biometrics, the fusion of two or more biometric modalities creates a robust recognition system that is resistant to the flaws of individual modalities. However, the excellent recognition of multimodal systems depends on multiple factors, such as the fusion scheme, fusion technique, feature extraction techniques, and classification method.

In machine learning, existing works generally use different algorithms for feature extraction of modalities, which makes the system more complex. On the other hand, deep learning, with its ability to extract features automatically, has made recognition more efficient and accurate. Studies deploying deep learning algorithms in multimodal biometric systems tried to find a good compromise between the false acceptance and the false rejection rates (FAR and FRR) to choose the threshold in the matching step. This manual choice is not optimal and depends on the expertise of the solution designer, hence the need to automatize this step. From this perspective, the second part of this thesis details an end-to-end CNN algorithm with an automatic matching mechanism.

This thesis has conducted two studies on face and iris multimodal biometric recognition. The first study proposes a new feature extraction technique for biometric systems based on machine learning. The iris and facial features extraction is per-

formed using the Discrete Wavelet Transform (DWT) combined with the Singular Value Decomposition (SVD). Merging the relevant characteristics of the two modalities is used to create a pattern for an individual in the dataset. The experimental results show the robustness of our proposed technique and the efficiency when using the same feature extraction technique for both modalities. The proposed method outperformed the state-of-the-art and gave an accuracy of 98.90%.

The second study proposes a deep learning approach using DensNet121 and FaceNet for iris and faces multimodal recognition using feature-level fusion and a new automatic matching technique. The proposed automatic matching approach does not use the threshold to ensure a better compromise between performance and FAR and FRR errors. However, it uses a trained multilayer perceptron (MLP) model that allows people's automatic classification into two classes: recognized and unrecognized. This platform ensures an accurate and fully automatic process of multimodal recognition. The results obtained by the DenseNet121-FaceNet model by adopting feature-level fusion and automatic matching are very satisfactory. The proposed deep learning models give 99.78% of accuracy, and 99.56% of precision, with 0.22% of FRR and without FAR errors.

The proposed and developed platform solutions in this thesis were tested and validated in two different case studies, the central pharmacy of Al-Asria Eye Clinic in Dubai and the Abu Dhabi Police General Headquarters (Police GHQ). The solution allows fast identification of the persons authorized to access the different rooms. It thus protects the pharmacy against any medication abuse and the red zone in the military zone against the unauthorized use of weapons.

Keywords: biometrics; multimodal biometric systems; iris recognition; face recognition; fusion; feature extraction, machine learning, deep learning, cyber security.

Acknowledgements

This was a challenging journey but well worth it. It would not have been possible without the guidance and encouragement of extraordinary and dear people who inspired and gave me the motivation to carry on.

I would especially like to express sincere gratitude to my first supervisor, Professor Fatih Kurugollu, and director of supervision, Professor Mahmoud Shafik, for their vision, commitment, and dedication throughout this Ph.D. journey. Their scientific expertise and supportive demeanor made this a pleasant experience. They were highly professional and diligent in this thesis's overview and strategic research direction.

Secondly, I would like to thank my parents and family for their patience. They knew their son was engaged in something critical for professional development, and the sacrifice they made by not being too demanding of my time all the while.

Many others facilitated my research by being amenable to data collection and discussion, both at my workplace and regarding the managers of the corporations I was engaged with for this research. Special thanks go to the central pharmacy staff of the Al Asria Eye Clinic in - Abu Dhabi - the United Arab Emirates, and Abu Dhabi Police General Headquarters (Police GHQ). Thank you all for making me feel accomplished and making this dream come true!

List of Publications

- Sulaiman ALSHEBLI, Fatih KURUGOLLU, Mahmoud SHAFIK. Biometric Authentication based on Face-Iris Images and Deep Learning Classifiers. Journal of Neural Computing and Applications (NCAA). **Submitted on June 30, 2021**.

- Sulaiman ALSHEBLI, Fatih KURUGOLLU, Mahmoud SHAFIK. Multimodal biometric recognition using iris and face features. IEEE TRANSACTIONS ON BIOMETRICS, BEHAVIOR, AND IDENTITY SCIENCE (T-BIOM). **Submitted on May 25, 2021**.

- Sulaiman ALSHEBLI, Fatih KURUGOLLU, Mahmoud SHAFIK. Cyber Security Solution Based on the Facial and Fingerprint Recognition. 6th International Conference on Artificial Intelligence and Pattern Recognition (AIPR2019), 16 – 18 of September 2021, Conference hosted by Lodz University of Technology, Lodz, Poland.

- Sulaiman ALSHEBLI, Fatih KURUGOLLU, Mahmoud SHAFIK. Multimodal biometric recognition using iris and face features. 18th International Conference in Manufacturing Research, 7 – 10 of September 2021, Conference hosted by the University of Derby, UK.

- Sulaiman ALSHEBLI, Fatih KURUGOLLU, Mahmoud SHAFIK. The Cyber Security Biometric Authentication based on Liveness Face-Iris Images and Deep Learning Classifier. The 8th International Conference on Informatics and Ap-

plications (ICIA2019), August 2-4, 2019, Japan, pages 16-26.

- Sulaiman ALSHEBLI, Fatih KURUGOLLU, Mahmoud SHAFIK. Cyber Security Platform Solution Based on the Facial Image and Fingerprint. International Journal of Cyber-Security and Digital Forensics (IJCSDF), 2019, Vol. 8, No. 2, pages 166-176. http://dx.doi.org/10.17781/P002603

Contents

List of Tables

List of Figures

Abbreviations

ADLVQ Adaptive Deep Learning Vector Quantiza- tion.

AI Artificial Intelligence.

ANN Artificial Neural Networks.

BSA Backtracking Search Algorithm.

CASIA Chinese Academy of Science Institute of Automation.

CCA Canonical Correlation Analysis.

CCL Cosmetic Contact Lens.

CCR Correct Classification Rate.

CNN Convolutional neural networks.

CS Cosine similarity.

DBN Deep Belief Network.

DCA Discriminant Correlation Analysis.

DCT Discrete Coefficient Transform.

DL Deep Learning.

DWT Discrete Wavelet Transform.

ECG Electrocardiogram.

ED Euclidean distance.

EER Equal Error Rate.

FAR False Acceptance Rate.

FK-NN Fuzzy K-Nearest Neighbor.

FRR False Rejection Rate.

FTC Failure To Capture.

FTE Failure To Enroll.

GA Genetic Algorithms.

GRU Gated Recurrent Units.

HBO Home Box Office.

HT Hough Transform.

ICT Information and Communications Technology.

LGPV Local Gradient Pattern with Variance.

LSTM Long Short-Term Memory.

MD Manhattan distance.

ML Machine Learning.

MT Malakooti Transform.

NIG Normal Inverse Gaussian.

ORL Olivetti Research Laboratory.

PCA Principal Component Analysis.

PSO Particle Swarm Optimization.

ReLu Rectified Linear Unit.

RNN Recurrent neural network.

ROC Receiver Operator Characteristic's.

SSA Singular Spectrum Analysis.

SVD Singular Value Decomposition.

SVM Support Vector Machine.

Tanh Hyperbolic tangent function.

Chapter 1

Introduction

1.1 General introduction

In our daily life, with the technological advances and the explosion of computer networks, knowing how to determine a person's identity automatically remains essential. Recognizing users to grant them authorization to use or access specific resources is imperative. Biometrics is seen as an indispensable solution to the ongoing security, fraud, and terrorism problem and is seen by governments as an excellent security solution [Cathy, 2005].

Nowadays, thanks to the computing power of computers and data storage capabilities, combined with complex computer programs, biometrics is no longer limited to fingerprints, and its use is no longer reduced to a law enforcement framework [Mróz-Gorgoń et al., 2022]. Today, governments are undertaking a policy of strengthening the use of biometric technologies and the private sector to combat terrorism and fraud, given the colossal security and economic stakes involved [Leese, 2022].

Biometric features are an alternative solution to the old means of identity verifica-

tion. The advantage of these biometric characteristics is that they are universal, i.e., present in all persons to be identified. On the other hand, they are measurable and unique: no two people can have the same characteristic. They are also permanent, which means that they do not vary over time [Guennouni et al., 2020].
To be qualified as biometrics modalities, collected characteristics must be:

- universal (exist in all individuals),
- unique (to differentiate one individual from another),
- permanent (allowing for evolution over time),
- recordable (collect the characteristics of an individual with his or her agreement),
- measurable (allowing for future comparison) [Micheli-Tzanakou and Plataniotis, 2011].

The main interest of biometrics is, therefore, to recognize and identify automatically the identities of individuals using their physiological characteristics or behavior. Physiological features may include the face, iris, impressions, fingerprint, hand geometry, palm print, hand vein, and retina. Behavioral characteristics include voice, signature, gait, etc.

1.2 Research Background

While there are many potential applications for biometrics, the main ones can be divided into four categories [Salveggio et al., 2012]. The first is security systems or logical access systems. Their role is to monitor, restrict or authorize access to data or information. In this case, biometrics replaces or complements pin codes, passwords, and tokens. The volume and turnover of the financial and banking industry

(particularly through e-commerce) and the value of sensitive personal data conveyed and/or stored in networks and computers make biometrics for securing logical access a much more widely deployed industry than that of physical security.

The second biometric application category concerns facility or physical access systems [Rao and Nayak, 2014]. Their role is to monitor, restrict or authorize the movement of a person into or out of a specific area. In this case, biometrics replaces or complements access keys and cards, allowing authorized users access to secure areas. Physical access systems are often deployed in the main perimeters of public infrastructures, such as airports, busy public places (museums, etc.), and border facilities to monitor and limit the movement of unauthorized or suspicious persons. In addition to entry to secure areas, physical access systems applied in a commercial setting are a tool to assist human resources management through employee attendance control systems by combining access verification at a location with recording the time when authentication was produced.

The third category concerns biometric systems that ensure the uniqueness of individuals [Raghavendra et al., 2017]. These systems typically focus on preventing double enrolment in programs or applications. Their main use occurs in the public sector, particularly in social assistance programs or voting systems.

Finally, government applications are the last and most important category of biometric applications [Scott et al., 2005]. They mainly cover the military and homeland security and the judicial fields (criminology, control of penitentiary institutions). They also cover civilian sectors such as government services, education, transportation, and health care.

1.3 Research Gap and Challenges

While face and iris recognition techniques are well established, the automatic recognition of faces and irises captured by digital cameras in a real, unconstrained environment remains very difficult, as it involves significant variations in both acquisition conditions and facial expressions and poses changes.

The Iris image has very rich texture information, and this information is randomly oriented in all directions and has multiple frequency components that need to be considered. Traditional feature extraction and selection techniques are difficult and time-consuming due to a large number of features in the Iris texture, especially when these features are statistically independent. To obtain the high-resolution feature extraction, a combination of methods and algorithms are required to ensure that all intelligent information in the Iris is captured and extracted for recognition. Most researchers have used the Gabor filter, in which fixed numbers of filter masks are used with predetermined frequencies and bandwidths. The outputs of the Gabor filter banks are non-orthogonal, and they will degrade more Iris features in the region of noisy images.

The most important problem in the Iris image recognition is the effects of the Cosmetic Contact Lens (CCL) on the captured image, the CCL with printed pattern image, or the printout of the Iris image. The researchers are familiar with this problem which badly has degraded the captured Iris image, and have applied both hardware and software techniques to reduce the effect of CCL. Since the effect of clear soft prescription lenses on recognition accuracy has been understated until recently, the research focus has been on textured contact lens detection, and many hardware and software-based approaches (and a combination of both) have been proposed in the literature. Although the hardware-based solution provides an efficient and gen-

eralized means for presentation attack detection due to its additional interaction and limitation, it is not the most optimal, and a robust software-based approach that can only work based on the standard Iris sensor would be the best solution for Iris recognition. Also, the noisy environment can degrade the cybersecurity recognition system because the effect of noise on the captured Iris image will change the extracted features, which are required to be compared with that feature in the database.

1.4 Aim and Objectives of the thesis

This thesis aims to explore machine and deep learning to propose a new multimodal biometric solutions. Multimodal systems enable improved recognition performance by combining several sources of information. They also solve the problem of non-universality of certain biometrics and offer a high degree of flexibility since biometric traits that are unusable or not preferred in some individuals can be compensated for by other biometric modalities. They limit the possibilities of fraud since they provide additional protection, making it more difficult to obtain and reproduce several characteristics simultaneously [Ross et al., 2019][Oloyede and Hancke, 2016]. For these reasons, multimodal biometric systems have been the subject of much research [Akulwar and Vijapur, 2019].

Multimodal systems enable improved recognition performance by combining several sources of information. They also solve the problem of non-universality of certain biometrics and offer a high degree of flexibility since biometric traits that are unusable or not preferred in some individuals can be compensated for by other biometric modalities. They limit the possibilities of fraud since they provide additional protection, making it more difficult to obtain and reproduce several characteristics simultaneously [Oloyede and Hancke, 2016]. For these reasons, multimodal biometric systems have been the subject of much research.

One of the objectives of this thesis is to find the most suitable architecture using machine learning or deep learning. To achieve this, we tested and compared both types. Machine learning employs statistical learning algorithms to identify patterns in the already available data, make predictions, and categorize new data. On the other hand, deep learning uses many layers in neural network models to complete challenging tasks.

The development of a multimodal biometric identification system based on two biometrics, namely the iris and the face, is performedfollowing these steps :

- Creation of a new multimodal face-iris dataset by combining two existing datasets.
- Development of a recognition system based on the Face and Iris images where features are extracted using Machine Learning.
- Introduction of deep learning framework for face and iris recognition.
- Proposal of a new automatic matching technique for the identification/verification process.
- Comparison of the results of the different proposals with the state-of-the-art.

1.5 Thesis Key Contributions

This thesis contributes to the field of multimodal biometrics. More specifically, it introduces new ideas and techniques in cyber security using face and iris traits. The major contributions of this thesis lie in developing two multimodal systems to achieve the final goal of improving the efficiency and rate of authentication. The first

proposed system is based on classical Machine Learning (ML) and the second one is based on Deep Learning (DL).

The first contribution concerns a cyber security platform based on machine learning models. The proposed algorithm uses the same feature extraction technique for both modalities (face and iris) based on two levels of data compression performed by the Discrete Wavelet Transform (DWT) and the Singular Value Decomposition (SVD) respectively, as well as three levels of securities by applying DWT on both images followed by SVD on the third level of DWT to obtain the singular values of the DWT coefficients. The singular values of the Face and Iris images will be extracted and then saved into two vectors corresponding to the face image features and iris image features. Finally, the contents of the two vectors will be merged, and the matching will be performed using the Euclidean distance. The results were promising compared with other recent biometric authentication algorithms in the state of the art.

The second proposed framework is based on Deep Learning. In this study, different deep architectures have been used to correctly classify the face and iris. A new technique for automatic matching has been proposed instead of choosing a threshold. The proposed approaches were compared with other recent biometric authentication algorithms. The obtained results prove the efficiency of the proposed techniques.

These contributions have resulted in three international conferences, one published article, and two others under review.

1.6 Thesis outlines and structure

The flowchart in Figure 1.1 illustrates the structure of the thesis.

The first chapter is devoted to the introduction. It describes the research context and background information, the aim, objectives, problem, and contribution of the

research.

The second chapter presents the basic notions of biometrics, i.e., the history of biometric systems, the different modalities, characteristics, and architecture of a biometric system, and the limitations encountered by the single-mode biometric system and state of the art on existing methods. This chapter is also devoted to the presentation of multimodal systems by detailing the advantages, design, and architecture, the different forms of information sources, and the different fusion levels. Finally, a state of the art of multimodal system combining iris and face will be presented.

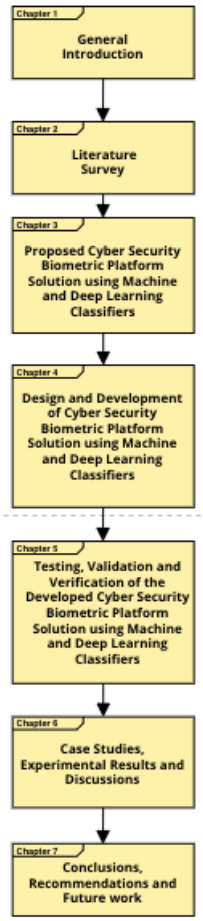


Figure 1.1: Structure of the thesis

The first part of the third chapter presents the first algorithm, based on Machine Learning, where DWT and SVD perform two levels of data compression. Details

will be given on the approaches used for feature extraction (DWT, SVD, etc.) and matching methods. The notion of cybersecurity and the encryption/decryption of feature vectors will also be addressed.
The second part of the third chapter presents the second proposed multimodal biometric Face-Iris system based on Deep Learning, detailing each of the architectures used and the design of the multimodal system that combines them.

The fourth chapter presents the proposed solutions' design and development using machine and deep learning.

In the fifth chapter, a complete study of the different methods proposed and a comparison with state of the art will be presented. A detailed discussion will highlight the advantages and disadvantages of each approach.

The sixth chapter presents two case studies with their results and discussions.

Chapter seven is a conclusion, in which the main contributions of this thesis will be summarized before outlining the considered Feature works.

Chapter 2

Literature Survey

2.1 Introduction

A biometric system is essentially a pattern recognition system that works in four steps. The first is the acquisition of the user's biometric data. Then comes the step of extracting characteristics from the acquired data, possibly preceded by a pre-processing phase. The third step is the comparison of the extracted characteristics against the model in the database to generate similarity measurements. Finally, a decision stage is used to conclude the user's identity (see Figure 2.1). The starting point for the biometric system is the enrollment phase. In this phase, a user's biometric data is initially collected and processed in a template: a form in which it is then stored for permanent use. Templates are not raw data or digitized images of a biometric sample but are a mathematical representation of distinctive features extracted by the biometric system [Jain et al., 2006].

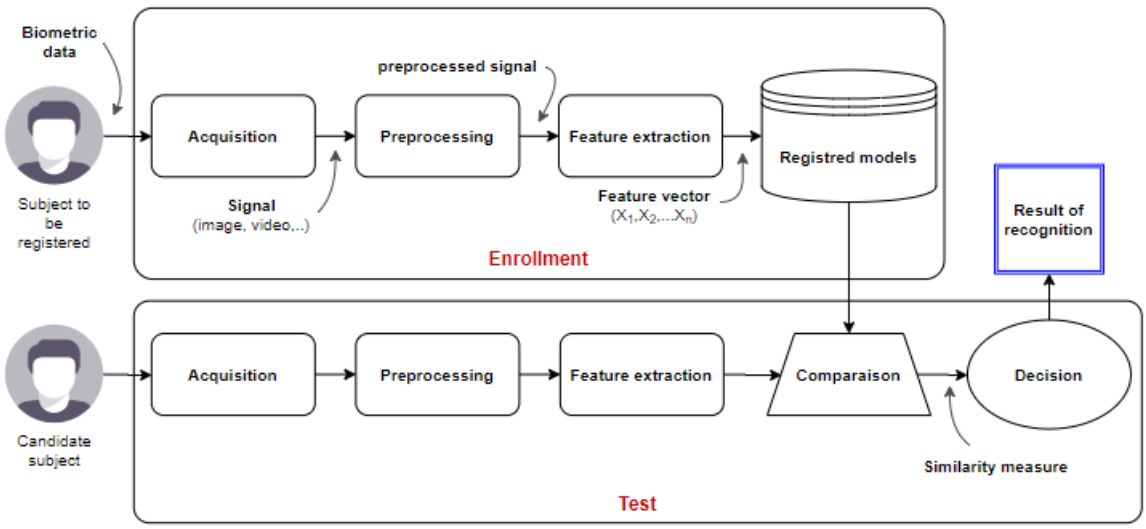


Figure 2.1: The main operations of a biometric system [Damer, 2018]

The architecture of a biometric system consists of five main modules:

- **acquisition or capture module:** it is a biometric sensor that can be of contact or non-contact type to acquire a specific modality of the person, for example, a camera in the case of the iris or face.
- **Signal processing module:** the acquired raw data is first pre-processed to improve its quality. After that, some relevant discriminatory characteristics are extracted to generate a compact representation called "Model or Template".
- **Storage Module:** contains the biometric templates of the system's users. The storage system can be a simple file in a single smartcard or a large database managed by a database management system.
- **Matching module (comparison):** compares the biometric data extracted by the feature extraction module to one or more previously stored templates (model). This module thus determines the degree of similarity (or divergence) between two biometric vectors.
- **The decision module:** usually, the result of the comparison is a score that represents the degree of similarity between 0 (total difference) and 1 (perfect

match) that allows the system to make the appropriate decision according to the application requirements.

2.2 Biometric system modes of operation

Biometric systems can be designed according to different modes of operation. There are two possible modes for a biometric system: the **identification mode** that answers the question "**Who am I**" and the **verification (or authentication)** mode that answers the question "**Am I the person I say I am?** [Jain et al., 2006].

In the identification mode, also known as "one-to-many" recognition, the comparison is made between the candidate's biometric template and the templates of all users stored in the enrollment base. Two contexts are possible. The first is to work within a closed group, i.e., one is sure that the candidate belongs to the database of authorized persons, and the problem is determining which of the identities best matches this candidate. Typically, this closed group context is adopted in research work. The second context is to work in an open group context, i.e., the applicant may be an impostor that should be rejected. Thus, the system can generate two decisions: rejection or acceptance with a determination of the identity of the candidate.

In the verification mode, also called authentication, it is always an open group context, i.e., one is not at all sure that the system knows the candidate's identity. In practice, the candidate claims the identity of one of the individuals registered in the database. The comparison is then made only between the candidate's biometric template and the templates of the advertised individual. Thus, it is a "one-to-one" recognition.

2.3 Biometric characteristics

A biometric characteristic is a measurable physical or behavioral characteristic of an individual that is unique and distinguishable. It determines how an individual will be recognized. Each biometric modality has its strengths and weaknesses; the choice generally depends on the field of application and sometimes on the population to be identified.

[Ross et al., 2006] have identified some requirements that a typical biometric system must meet:

- **Universality:** everyone accessing the system should have the features, for example, we cannot use the iris as a feature to identify blind people.
- **Uniqueness:** to enable differentiation between one individual and another.
- **Permanence (stability):** biometric characteristics should be resistant to change over time.
- **Measurability:** biometric characteristics must be quantitatively measurable and then processed and used to compare two individuals.
- **Performance:** a practical biometric system must have acceptable accuracy and reasonable recognition speed for the resources required.
- **Acceptability:** the degree to which the persons intended to use the application agree to present their biometric features to the system.
- **Circumvention:** reflects how easy it is to deceive the system by fraudulent methods.

Based on these criteria, a first comparison of the leading biometric technologies is proposed in Table 2.1.

Table 2.1: The qualities of some biometric modalities [Maltoni et al., 2009]

Modality	Universality	Uniqueness	Permanence	Measurability	Performance	Acceptability	Circumvention
Face	High	Weak	Medium	High	Weak	High	Weak
Finger-print	Medium	High	High	Medium	High	Medium	Weak
Hand geo-metry	Medium	Medium	Medium	High	Medium	Medium	Medium
Iris	High	High	High	Medium	High	Weak	High
Retina	High	High	Medium	Weak	High	Weak	High
Signa-ture	Weak	Weak	Weak	High	Weak	High	Weak
Voice	Medium	Weak	Weak	Medium	Weak	High	Weak
Facial thermo-gram	High	High	Weak	High	Medium	High	High

Many biometric modalities have been proposed and are used in various applications. Physiological modalities are based on morphological or biological characteristics and include the face, ear, iris, retina, finger and palm prints, hand geometry, venous network, DNA, etc. Behavioral modalities use a personal trait of behavior such as voice, signature dynamics, gait, typing dynamics, or lip movement. Some of these biometrics have a long history and can be considered mature technologies, while others are still young research arenas [Dargan and Kumar, 2020].

A practical biometric system should have acceptable accuracy and speed of recognition with reasonable resources, be accepted by the target population, and be robust enough for various fraudulent attacks [Maltoni et al., 2009].

Among the most mature techniques, we distinguish face, fingerprint, hand geometry, iris, and retina, which contain good characteristics. However, none of them is perfect. Each technique has advantages and disadvantages, acceptable or unacceptable depending on the application in terms of the level of security and/or ease of use, etc.

We are therefore tempted to say that these five biometric solutions are not systematically in competition with each other. That said, retinal recognition, which requires sophisticated and expensive acquisition equipment, can already be discarded as too intrusive for general public applications.

For the most reliable results, it is best to use two or more of these biometrics (like the face and iris) together for important tasks [Ammour et al., 2020]. So, when these two methods are used together, the face can get around the limits of the iris, and vice versa. Indeed, the face will compensate for the low acceptability of the iris, and the high uniqueness of the iris will counterbalance the low uniqueness of the face.

2.4 Evaluation of a biometric system

Despite the many advantages of biometric authentication systems, their implementation carries some risks. Even the most accurate biometric system is not perfect, and errors will be generated.

In this thesis, performance indicators that compare different algorithms such as Correct Classification Rate (CCR), False Acceptance Rate (FAR), False Rejection Rate (FRR) will be used. These metrics are widely used in biometrics as they judge the system on the rigor in granting access to authorized persons on one side and the non-indulgence with unauthorized persons on the other side.

A biometric authentication system can produce two possible decisions: acceptance or rejection of an applicant. Two errors are then likely to occur:

- **False acceptances:** when the system accepts impostors by wrongly consider-

ing them to be authorized persons,

- **False rejections:** when the system rejects genuine people by wrongly considering them as imposters.

These error rates are then measured according to equations 2.1 and 2.2 with the decision threshold τ. The False Acceptance Rate (FAR) is calculated from the ratio of the number of false acceptances (FA) (i.e., the number of misclassified imposters) over the total number of imposter accesses in the base (N_{Imp}). Similarly, the False Rejection Rate (FRR) is calculated by dividing the number of false rejects (FR) (number of misclassified authentic persons) by the total number of authentic accesses (N_{Aut}).

$$FAR(\tau) = \frac{FA(\tau)}{N_{Imp}} \tag{2.1}$$

$$FRR(\tau) = \frac{FR(\tau)}{N_{Aut}} \tag{2.2}$$

The evolution of these error rates according to the decision threshold can be visualized through the distributions of Authentic and Impostor scores as shown in Figure2.2. Ideally, the choice of decision threshold should correspond to the value of τ, which minimizes both the FAR and the FRR. However, these two errors have opposite monotonies since FAR is increasing while FRR is decreasing. The choice of the threshold will then have to be the subject of a compromise between safety and comfort, which will depend on the application's needs. Some applications require a very low FAR (or almost zero, for example, when accessing confidential documents in a military context). Other applications do not tolerate a high FAR false rejection rate (such as when accessing a cell phone).

A particular operating point is the Equal Error Rate (EER) which is achieved when FAR and FRR are equal. This point is often used because it is neutral and indepen-

dent of the type of application.

To evaluate our models, we will also calculate the Correct Classification Rate (CCR), which represents the number of well-classified individuals (WCI) relative to the total number of individuals (N_{Ind}) (equation 2.3). A person is considered a WCI in two cases: first, if it appears in the database and is accepted and recognized as being that person, or second if a person does not exist in the database and has been rejected.

$$CCR = \frac{WCI}{N_{Ind}} \tag{2.3}$$

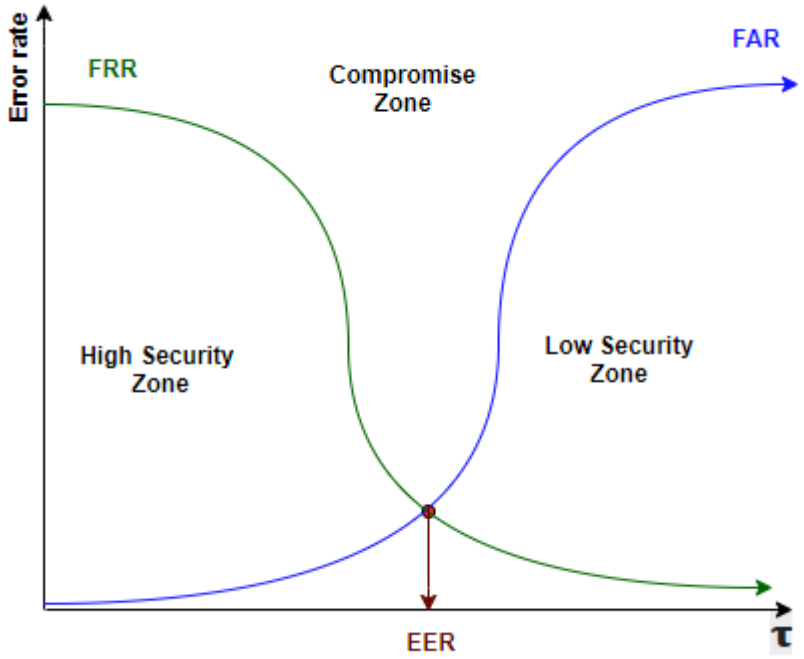


Figure 2.2: Illustration of the FAR and the FRR through their variation according to the decision threshold

2.5 Limitations and challenges of monomodal biometric systems

Although small- and medium-scale biometric applications (a few hundred users) can still use single-mode systems (called single-mode or unimodal systems), they suffer from several limitations that hamper their use when it comes to building a high-precision recognition system for large-scale applications [Jain and Ross, 2004].
Among these limitations are :

- **Noise in collected data:** examples of noisy data are a fingerprint with a scar or a voice altered by a cold. But noisy data is usually due to defective or poorly maintained sensors (e.g., dust accumulation on a fingerprint sensor, bad camera focus causing blurring) or unfavorable environmental conditions (e.g., bad lighting). The noisy biometric data may then be mismatched with the templates in the database resulting in the incorrect rejection of an authorized user.

- **Non-universality:** the use of a unimodal biometric system assumes that each individual in the target population possesses the modality in question. However, this hypothesis is not always verified, and a subset of the population risks being excluded by the monomodal system if no other alternative is offered. For example, a voiceless person cannot be enrolled in a voice recognition system, nor can a person with hand paralysis be enrolled in a signature recognition system. In these cases, the persons do not have biometrics, and Failure To Enroll (FTE) errors are generated by the system. Another problem of non-universality is to possess the biometric trait but not in a way that the biometric system can use. For example, a person suffering from an ocular cataract will not provide iris images of enough good quality for automatic recognition, or a person suffering from extreme dryness of fingers will not provide usable fingerprint images. The system will generate a Failure To Capture (FTC) error in this case.

- **Lack of individuality:** while a biometric trait is supposed to vary considerably from one individual to another, there can be great similarities between classes with this trait. For example, identical twins have an almost identical facial appearance. This limitation reduces the discriminatory capacity of the biometric system, which is more likely to accept people who should not be.

- **Intraclass variations:** the biometric data acquired from an individual during authentication can be very different from the data from which the individual's model was generated during the enrollment phase. This variation may be caused by a poor interaction between the user and the sensor (for example, by changing the pose) or when the characteristics of the sensor are modified (for example, by changing the sensor), or by a variation in the ambient environmental conditions (for example, a change in lighting) or even by a variation in the psychological makeup of the individual (for example, inducing a change in expression, tone, dynamics, whether it be for the gait, signature or keyboard typing).

- **Sensitivity to attacks:** biometric traits are much more difficult to counterfeit than traditional means of identification such as passwords and access cards. However, impostures exist particularly in behavioral modalities such as signature and voice. In addition, physiological traits are also susceptible to spoofing attacks. In particular, fingerprints can be reproduced with silicone and the face with a photograph.

To overcome these limitations, multimodal biometric systems have emerged. Their purpose is to increase security by eliminating any possibility of identity spoofing. Indeed, it is unlikely that a person can usurp several types of biometric traits simultaneously, and, in this thesis, face and iris traits are considered.

2.6 Multimodal biometric systems

The principle of multimodality in the general sense is to combine several sources of information from monomodal systems [Jain and Ross, 2004]. According to the combined sources of information, five multimodal scenarios are possible (see Figure 2.3):

- **Multi-sensor systems:** when different sensors are used to acquire the same biometric trait, for example, a digital camera, a webcam, and a thermal camera to capture facial images.
- **Multi-sample systems:** when several samples of the same trait/modality are associated, for example, the images of the left and right iris or the imprints digits of the index and middle fingers.
- **Multi-instance systems:** when several instances of the same stroke, of the same biometric sample and having been acquired by the same sensor are associated, for example, several images of faces taken with pose changes (frontal or profile), expression or lighting.
- **Multi-algorithm systems:** when the processing of an image is done by combining several algorithms either in the extraction module of the characteristics or for comparison.
- **Multi-biometric systems:** or multi-trait systems or systems multimodal in the strict sense of the term, when several modalities/tracts are involved, biometrics are combined, for example, voice and signature or face and iris.

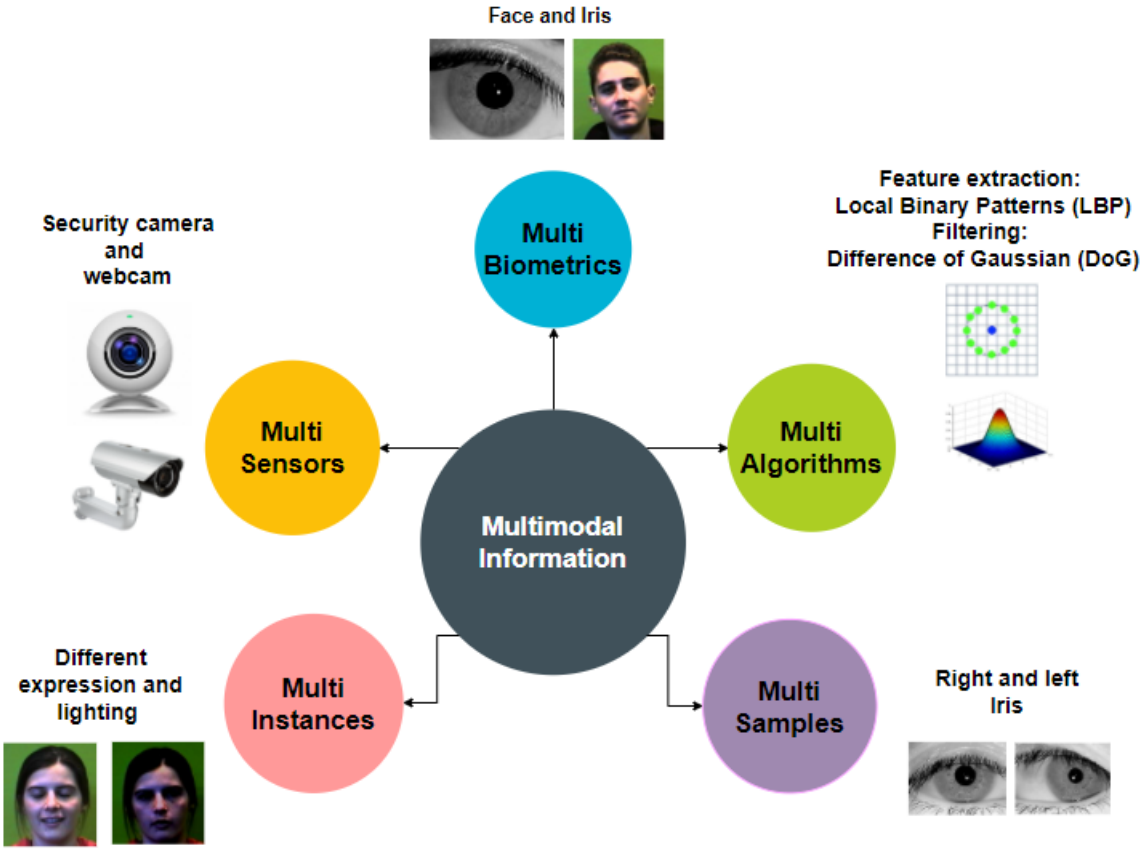


Figure 2.3: Multiple sources of information in a multimodal biometric system [Damer, 2018]

From these scenarios, it can be imagined any combination. A comparison of the interest of multimodal scenarios was carried out in [Tissé, 2003] according to three criteria (see Table 2.2): the gain in performance, the hardware cost of the system, and finally, its consumption in terms of acquisition or execution time for digital processing.

No single scenario can meet all the requirements of a biometric system. Based on the comparison in Table 2.2, these requirements are relatively less satisfied by the multi-sensor and multi-jurisdictional scenarios compared to the others.

The multibiometric scenario is potentially the one that achieves the best performance in terms of FAR and FRR. However, this scenario has drawbacks related to ease and

cost of use. Indeed, using several biometric modalities may increase acquisition and processing time. If it is assumed that the capture of multiple biometrics and the processing of the information can be done in parallel, then this scenario meets the criterion of time consumption. On the other hand, using several sensors necessarily implies an increase in material cost. For this reason, in practice, multimodal systems comprising more than two modalities are rarely used.

Table 2.2: Comparison of multimodal scenarios according to the criteria of performance, material cost and time consumption. +/- designates the level of satisfaction (adapted from [Tissé, 2003]).

	Multi sensors	**Multi instances**	**Multi samples**	**Multi algorithms**	**Multi biometrics**
FRR and FAR performance	+	-	++	+	+++
Material cost	-	++	++	+	–
Time consumption	-	+	-	++	+

2.7 Design of a multimodal biometric system

A multimodal biometric system can be designed according to three architectures [Jain and Ross, 2004] [Fierrez-Aguilar, 2006]: serial (or cascade), hierarchical, and parallel architecture.

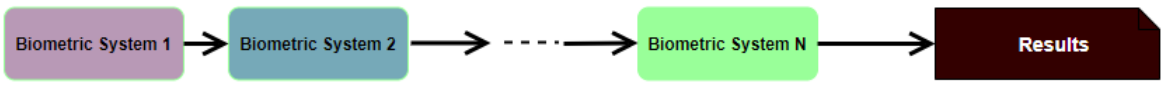


Figure 2.4: Architectures of a serial multimodal system [Damer, 2018]

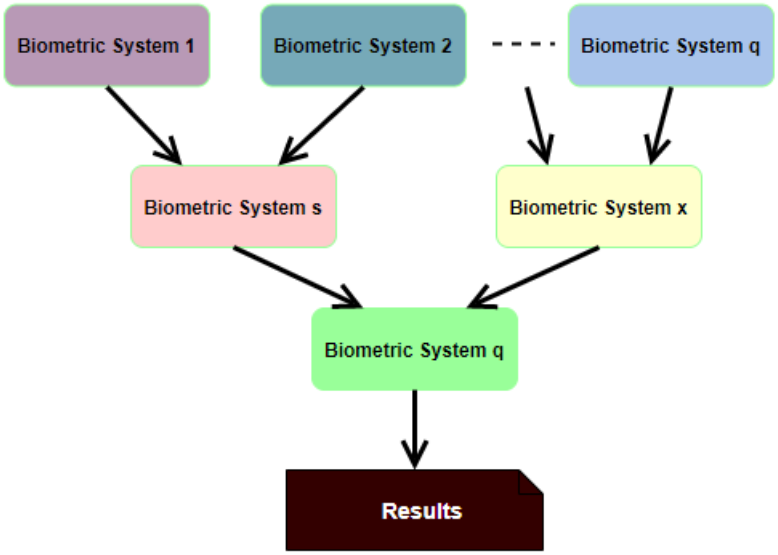


Figure 2.5: Architectures of a hierarchical multimodal system [Damer, 2018]

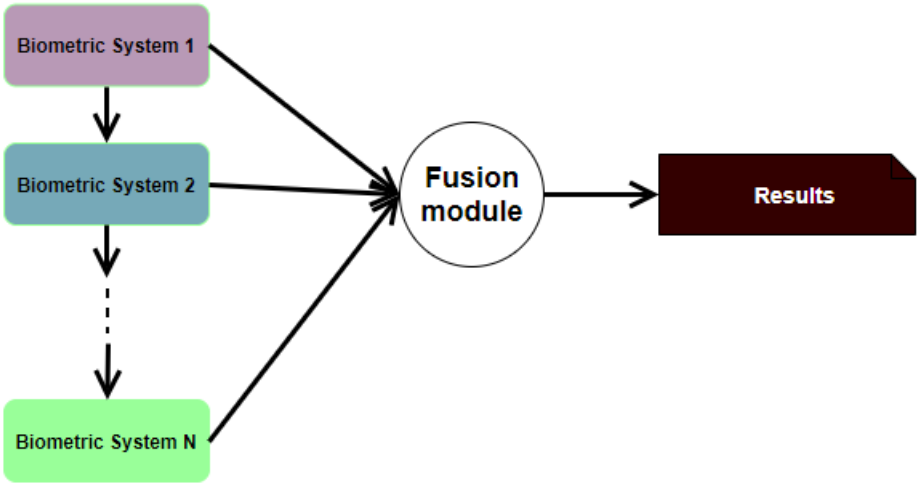


Figure 2.6: Architectures of a parallel multimodal system [Damer, 2018]

The individual systems are invoked in sequence in the serial architecture (Figure 2.4). Some of them may be used only when a possible condition occurs at the output of the previously invoked systems, thus allowing them to decide without involving all of these systems. This architecture can be used as an indexing schema to reduce the number of possible identities before using the next data. It also increases efficiency by using lower cost and less expensive systems first and then the too expensive but more accurate systems.

In the hierarchical architecture (Figure 2.5), the individual systems are combined into a tree structure. This architecture is considered the most flexible and allows to cope with the problems of missing or poor quality data often encountered in biometric systems.

In the parallel architecture (Figure 2.6), the information from the different systems is used simultaneously to perform the recognition task. Using all biometric information is then required to produce a decision, which is likely to bring more improvement than in the case of a serial architecture. As a result of these advantages, most of the methods proposed in the literature belong to this category of architecture which is also the case for the methods proposed in this thesis work.

The choice of a system's architecture depends on the application's needs. For example, in user-friendly and low-risk applications, serial architecture is preferred for its cost advantages in terms of time and hardware compared to parallel architecture requiring the acquisition and processing of a large amount of biometric data.

2.8 The levels of fusion

A biometric system comprises four modules: a data acquisition module, a feature extraction module, a matching module (also called a "classification" module), and a decision module. Multimodal fusion can be performed at the output of any of these modules, giving rise to four levels of fusion which can be grouped into two main families: fusion before the match and fusion after the match [Jain et al., 2005b].

2.8.1 Fusion before the match

This fusion is done before the matching module, at the sensor level, or at the feature level.

2.8.1.1 Fusion at the sensors

It is a matter of merging raw data from the sensor(s). This type of fusion can only be done between different instances of the same biometric and requires compatible data, for example, combining several facial images in different poses to form a 3D model.

2.8.1.2 Fusion at the features

It is a matter of combining characteristic vectors from different sensors or obtained by applying different algorithms to the same biometric data. When the characteristic vectors are homogeneous, e.g., several impression prints fingers, a weighted sum can be applied to obtain a single characteristic vector. If the feature vectors are heterogeneous, for example, if they come from different characterization algorithms or different biometric traits, and if they are compatible, which is not always the case as is the case with the principal component analysis coefficients of a fingerprint, they can be concatenated into a single characteristic vector.

2.8.2 Fusion after the match

2.8.2.1 Fusion at the decision level

This involves combining the decisions of the biometric systems, each of which gives a response (accepted: 1 or rejected: 0, in the case of verification) depending on the input. This level of merging can be achieved by simple rules such as AND, OR, and majority voting, as well as by more complex rules such as weighted voting or classification in the decision space [Jain and Ross, 2004]. Fusion at the decision level has the advantage of being simple. On the other hand, the information it uses is very limited (0 or 1).

2.8.2.2 Fusion at the score level

The goal is to combine the scores from each individual comparison module. Scores have the advantage of being independent of biometric systems and, therefore, much easier to access than features. Also, they overcome the compatibility and large feature space constraints encountered with feature-level merging. Thus, score-level merging provides the best compromise between richness of information and ease of implementation.

The choice of a score combination approach for the fusion involves a prior normalization step before the fusion. This step is essential for the following three reasons:

- the outputs of the matching modules may be non-homogeneous (distances/similarities),
- scores may be in different ranges,
- The statistical distributions of the outputs of each matcher can be different.

2.9 Machine learning

Artificial Intelligence (AI) is a discipline that seeks methods for solving very complex problems in logic or algorithms.
Machine Learning (ML) is an approach to achieving AI using algorithms to determine or predict patterns based on existing data. Machine learning algorithms then automatically deduce rules to distinguish classes.

Image processing is a beneficial technology, and the demand from the industry seems to increase every year. Historically, image processing using machine learning emerged in the 1960s to simulate the human vision system and automate the image analysis process. As the technology developed and improved, solutions for specific tasks appeared.

Generally, three steps are needed for intelligent image processing using machine learning: preprocessing, feature extraction, and finally, the matching step (classification/recognition).

2.9.1 Preprocessing

Image processing is a method of performing certain operations on an image to obtain an improved image or extract useful information from it. Before doing image processing, one usually goes through a preprocessing or data cleaning step before building the intelligent model. This step aims to prepare the images to facilitate their analysis and computer processing. The preprocessing method choice depends on the data's quality and nature.

The acquisition of images for facial and/or iris recognition generally depends on the camera equipment and lighting conditions. In order to compare images under iden-

tical conditions, image processing methods for lighting problems generally attempt to normalize all facial images to canonical lighting. In this thesis, two preprocessing methods, namely contrast and Gamma correction, are used.

2.9.1.1 Contrast Enhancement

The contrast of an image is one of the most important factors influencing its subjective quality. Contrast enhancements improve the perceptibility of objects in the scene by enhancing the brightness difference between objects and their backgrounds. The contrast enhancement algorithm described in [Ramesh and Ramesh, 2016] is used used in this thesis, where images are converted from RGB to HSV color space to achieve the enhancement of the luminance component (V) using the Class Limited Adaptive Histogram Equalization (CLAHE). Discrete Wavelet Transform is applied to the Saturation (S) components, and the decomposed approximation coefficients are modified by a mapping function derived from the scaling triangle transform. The enhanced S component is obtained through Inverse Wavelet transforms. In the end, the image is converted back to the RGB color space. Figure 2.7 shows an example of contrast enhancement.

(a) Original image (b) Preprocessed image

Figure 2.7: Contrast preprocessing

2.9.1.2 Gamma correction

Gamma correction enhances the local dynamic range of the image in dark or shadowed regions while compressing it in bright regions. Gamma correction is a nonlinear operation that replaces the values of I by $I^{1/\gamma}$. Thus, the overall tone of an image can be lightened or darkened depending on the gamma value used while maintaining the image's dynamic range. Figure 2.8 shows the result of Gamma correction with $\gamma = 0,4$.

(a) Original image (b) Preprocessed image

Figure 2.8: Gamma correction

2.9.2 Feature extraction

Feature extraction prior to classification is an essential task. Feature extraction methods take out images' most representative and relevant features to achieve a high classification accuracy. In this research, a static face and iris recognition based on the Extracted Feature Vectors obtained from the Discrete Cosine Transform (DCT), the Discrete Wavelet Transform (DWT), the Hough Transform (HT), and the Malakooti Transform (MT) is performed. In our proposed experiments, the Singular Value Decomposition (SVD) is applied to each sub-image, and its dominant features are extracted and stored inside the feature vector.

2.9.2.1 Discrete Wavelet Transform

The real-world signals are mostly displayed in the time domain. The amplitude of signals versus times is depicted because these signals represent the air temperature, pressure, humidity, or heart bits of patients. The time-domain signals only display the variation of the signal amplitudes at different instances of time. However, some vital information are hidden, and they will be revealed in the frequency domain. The frequency spectrum of signal or representation in the frequency domain shows the signal's frequency components and reveals its behavior at different frequencies. The transformation of signal from the time domain to frequency can multiply the transformation matrix with the time domain signal or original data. The Fourier Transform (FT) is one of the most powerful tools used to transfer the signal from the time domain to the frequency domain to analyze the signals' frequency contents. When FT is applied to any signal, it reveals its frequency contents from low frequency to high frequency. Since the noise or unwanted signal usually has a high frequency, noise can be removed or reduced when the FT is applied to the mixture of signal and noise [Furht, 2008].

If we apply the FT over entire windows of time, we can only reveal the signal's frequency components that are distributed over the frequency axis from low frequency to high frequency. However, we cannot determine which frequency component belongs to which time window if FT is applied over the whole time axis. Short-Time Fourier Transform (STFT) is the key solution that uses FT of small windows of time and slides slowly over the entire time axis. The FT of each short time widows can be displayed to reveal the signal's frequency components in each selected window and obtain the information on both time and frequency. However, the STFT has solved the problem of time-frequency components but still has a resolution problem due to the limitation of the information in a small window of time. One transform that can simultaneously provide the time and frequency information is Wavelet Transform

(WT), which displays the time-frequency representation of the signal and reveals the signal's frequency behavior at each instance of the time[Furht, 2008]. Therefore, WT seems to be the best solution that can be used to solve the problem of time-frequency components and the frequency resolution due to the length of selected windows. If the signal has all its frequency components (stationary signal) at all times, the FT is appropriate for the spectrum analysis. However, suppose the signal has a specific frequency component at certain times(non-stationary). In that case, the WT is more appropriate than STFT, and FT is not suitable and has the weakness to simultaneously reveal time-frequency information simultaneously [A. et al., 2002].

The Continuous Wavelet Transform (CWT) was developed to overcome the resolution problem of the Short-Time Fourier Transform, especially for the nonstationary biological signals, such as Electroencephalograph (EEG), Electrocardiograph (ECD), and Electromyography (EMG). An essential characteristic of the wavelet transform is that the window width will be changed as the transform operation is performed for every signal spectral component. The continuous wavelet transform is defined as the following equation:

$$\Psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int x(t)\psi(\frac{t-\tau}{s})dt \tag{2.4}$$

Where the function $\psi(t)$ is called the mother wavelets, and all other wavelets are derived from it by applying time shift and scaling operations. The mother wavelets are functions that should satisfy specific properties, such as integrating to zero and waving above and below the x-axis. Also, all wavelets derived from mother wavelets must be orthogonal to each other. The sets of orthogonal wavelet functions can span dimension N because all vectors in that space are mutually orthogonal. The vector space spanned by a set of orthogonal wavelet functions is an orthogonal matrix. Any discrete vector in that space can be decomposed into the liner of those wavelet functions. The parameter τ is called the translation and indicates the location of the

selected windows. Parameter S is called the scaling factor and refers to frequency resolution. It is similar to the scaling factor of a city map, which provides fewer details for large S and more details for small S. Similarly, in term of frequency, the high scales refer to the low frequencies and corresponds to global information, whereas the low scales refer to high frequencies and corresponds to the detailed information of a hidden pattern in the signal [Furht, 2008][Campo et al., 2016].

The DWT provides sufficient information for analysis and synthesis of the original signal, significantly reducing computation time. The DWT is considerably easier to implement when compared to the CWT, and it was introduced in 1976 by Croiser, Esteban, and Galand. They were working to find an efficient algorithm to decompose discrete-time signals. There is some similarity between the Discrete Fourier Transform (DFT) and DWT, making DWT that makes DWT so popular. The DFT decomposes the discrete-time signals with linear combinations of the sinusoidal function as the DFT basis with different frequencies. However, the DWT decomposes those signals with a linear combination of the wavelet functions derived from the mother wavelet function with different time shifts and scaling. When DWT is applied to the face image, as shown in Figure 2.9, it will transfer the face image into four different sub-bands, approximation subband (LL), diagonal detail sub-band (HH), Horizontal detail subband (LH), Vertical detail sub-band (HL)[Campo et al., 2016].

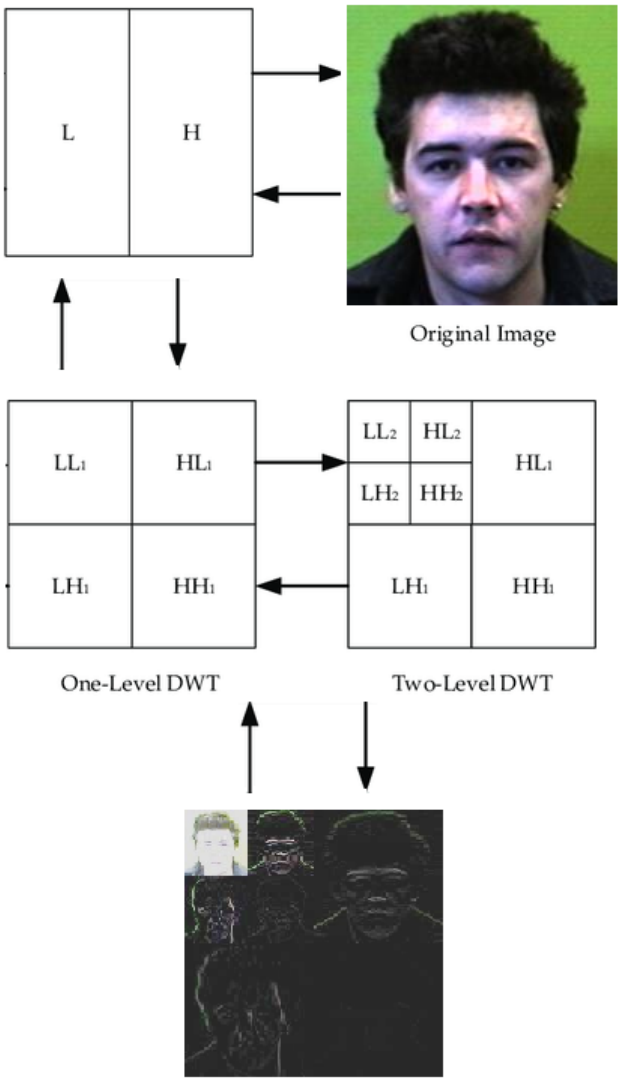


Figure 2.9: Two-Level DWT

2.9.2.2 Discrete Cosine Transform

In Digital Signal Processing (DSP), the concept of data compression or source coding is related to the process of information encoding with fewer bits than the original representation. The data compression aims to save memory space where data are stored on the storage device, flash drive, hard disk, cloud computing, or other storage facilities. The data compression can be classified as either Lossy or Lossless. In Lossy compression, the numbers of bits are reduced by eliminating unnecessary information, such as converting an image from bitmap format into the JPG format. In Lossless compressions, the numbers of bits are reduced by identifying and elimi-

nating those redundant bits, where the locations and corresponding bit pattern will be registered for the decompression process, such as Zip files. The data compression will reduce the information size useful for transmission, storage, and reconstruction [Malakooti and Khederzdeh, 2012].

DCT is one most important data compression algorithms that can be used to convert the string of numbers from the time domain into a sequence of coefficients in the frequency domain, where DCT coefficients are ordered in Low Frequency (LF), Middle Frequency (MF), and High Frequency (HF) regions. The DCT coefficients are in descending order in which the LF coefficients have higher vales than the MF and HF coefficients. The first coefficient of DCT has the highest value or energy of other coefficients. The following coefficients' energy is in decreasing order, i.e., the lowest energy is related to the last coefficient. The DCT coefficient can reconstruct the original information by applying the inverse of DCT (IDCT) to the coefficients. If all coefficients are used in the reconstruction process or IDCT, the original data will be reconstructed with zero error rates. But if a low-frequency coefficient and portions of MF or HF are used in reconstruction, the error rates will be reduced. The error rate of reconstruction is relatively proportional to the number of DCT coefficients used in the reconstruction. One-dimensional DCT is used for the compression, approximation, reconstruction, and even identification of the one-dimensional signals corresponding to voice, pressure, temperature, humidity, etc. Two-dimensional DCT is used for the same purpose but is related to two-dimensional signals such as a face image, an iris image, fingerprints, etc. Since most energy is stored inside the low-frequency DCT coefficients of the face image, they can be used as the reliable coefficients to be stored in the corresponding feature vectors for storage reduction and retrieving individual identification [Ahmed et al., 1974].

The DCT of the one-dimensional array (1D-DCT) for signal $f(i)$ with length N can

be defined by equation 2.5:

$$C(t) = \alpha(t) \sum_{I=0}^{N-1} f(i) cos \left[\frac{\pi(2i+1)}{2N} \right] \tag{2.5}$$

For $t = 0, 1, 2, \ldots, N-1$.

Similarly, the IDCT of the array of DCT coefficients, C(t) with length N can be defined as equation 2.6:

$$f(i) = \sum \alpha(t) C(t) cos \left[\frac{\pi(2i+1)_t}{2N} \right] \tag{2.6}$$

For $t = 0, 1, 2, \ldots, N-1$.

In both equations 2.5 and 2.6 $\alpha(t)$ is defined as:

$$\alpha(t) = \begin{Bmatrix} \sqrt{\frac{1}{N}} & for & t = 0 \\ \sqrt{\frac{2}{N}} & for & t \neq 0 \end{Bmatrix} \tag{2.7}$$

The following analysis has shown the power of DCT coefficients in compression.

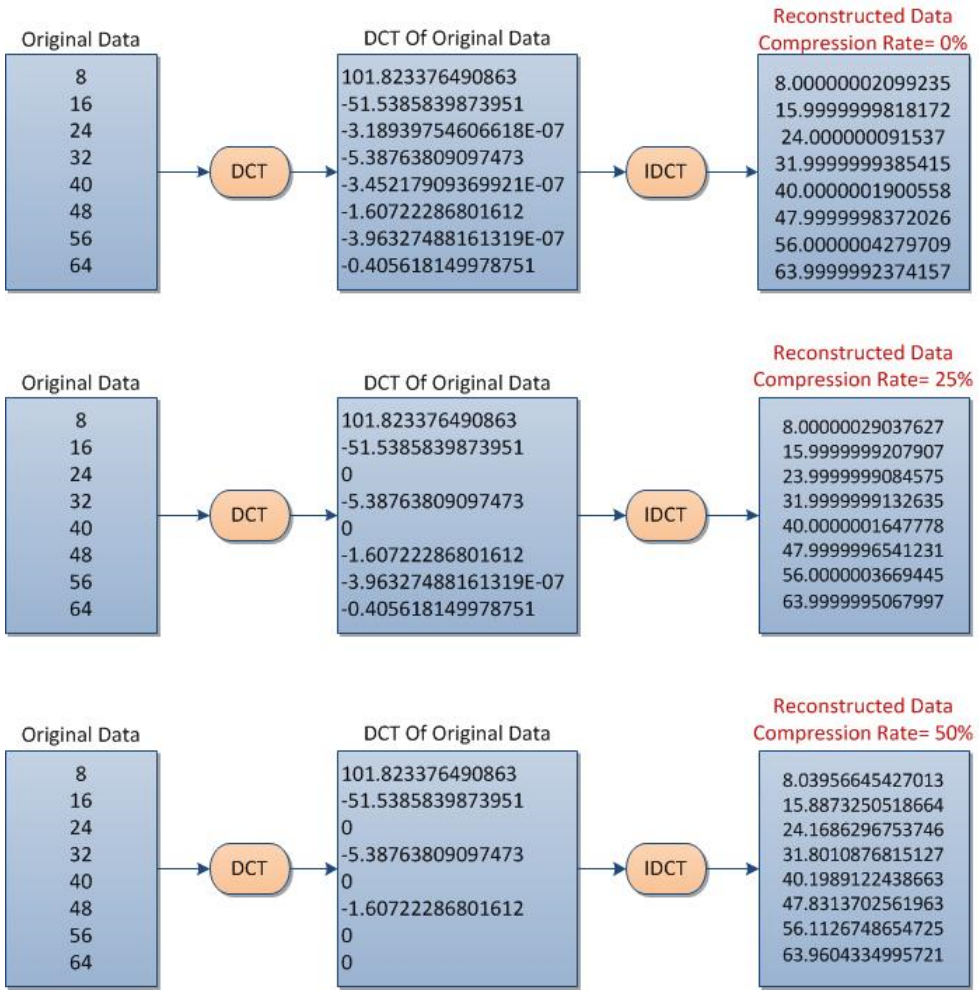


Figure 2.10: Comparison of original and reconstructed data using different compression rates

The tabulated values for original and reconstructed data clearly show that if all DCT coefficients are used in the reconstruction process, then the calculated error rates would equal zero.

The DCT of the two-dimensional array (2D-DCT) for signal f(x,y) with length N by N can be defined by equation 2.8:

$$C(i,j) = \alpha(i)\alpha(j) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) cos \left[\frac{\pi(2x+1)_i}{2N} \right] cos \left[\frac{\pi(2y+1)_j}{2N} \right] \tag{2.8}$$

for $i = 0, 1, 2, \ldots, N-1$, $j = 0, 1, 2, \ldots, N-1$, where $\alpha(i)$ and $\alpha(j)$ are defined in

equation 2.7.

Similarly, the IDCT of the matrix of DCT coefficients, $C(i,j)$ with a length of N by N can be defined by equation 2.9:

$$f(x,y) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \alpha(i)\alpha(j)C(i,j)cos\left[\frac{\pi(2x+1)_i}{2N}\right] cos\left[\frac{\pi(2y+1)_j}{2N}\right] \tag{2.9}$$

for $x = 0, 1, 2, \ldots, N-1$, $y = 0, 1, 2, \ldots, N-1$.

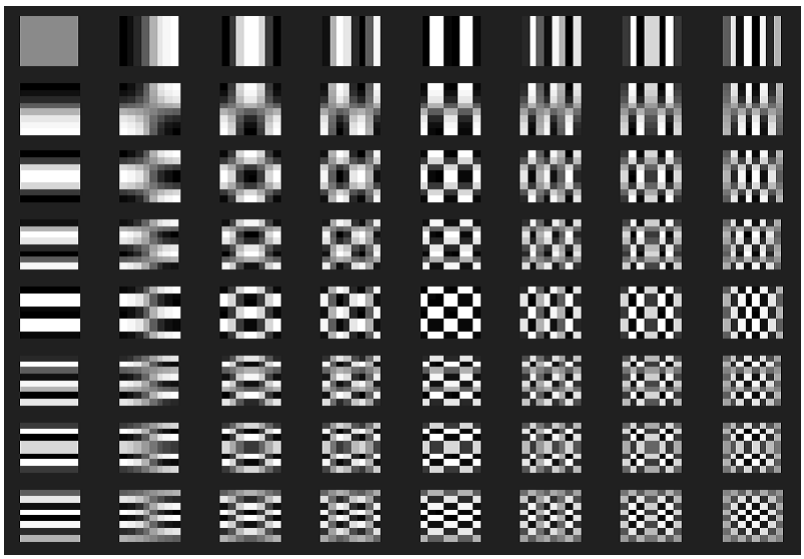


Figure 2.11: Two Dimensional DCT coefficients, N=8[Malakooti and Khederzdeh, 2012]

$$\begin{bmatrix} 162.3 & 40.6 & 20.0 & 72.3 & 30.3 & 12.5 & -19.7 & -11.5 \\ 30.5 & 108.4 & 10.5 & 32.3 & 27.7 & -15.5 & 18.4 & -2.0 \\ -94.1 & -60.1 & 12.3 & -43.4 & -31.3 & 6.1 & -3.3 & 7.1 \\ -38.6 & -83.4 & -5.4 & -22.2 & -13.5 & 15.5 & -1.3 & 3.5 \\ -31.3 & 17.9 & -5.5 & -12.4 & 14.3 & -6.0 & 11.5 & -6.0 \\ -0.9 & -11.8 & 12.8 & 0.2 & 28.1 & 12.6 & 8.4 & 2.9 \\ 4.6 & -2.4 & 12.2 & 6.6 & -18.7 & -12.8 & 7.7 & 12.0 \\ -10.0 & 11.2 & 7.8 & -16.3 & 21.5 & 0.0 & 5.9 & 10.7 \end{bmatrix}$$

Figure 2.12: Two Dimensional DCT coefficients in matrix format, N=8 [Malakooti and Khederzdeh, 2012]

Figure 2.11 and 2.12 have shown the display of the 2D-DCT coefficients for a block of 8*8 image vales. The coefficients of 2D-DCT have been shown by different light intensities where the energy of coefficients at the upper left corner have the highest value (Low-Frequency Coefficients). The energy of coefficients at the lower right corner has the lowest values (High-Frequency Coefficients). However, the upper right corners and lower left corners are referred to as Mid Frequency Coefficients (see Figure 2.13).

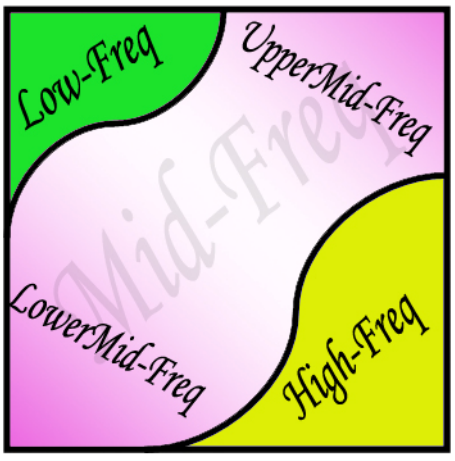


Figure 2.13: Two Dimensional DCT Frequency Regions [Malakooti and Khederzdeh, 2012]

Equation 2.8 can be written in matrix form for simple representation. If we use the

symbol of Img for the target image, f(x,y), and matrix T for the 2-D DCT transform, then the matrix form of 2-D DCT coefficients for equation 2.8 can be shown as equation (37).

$$C = T * Img * T \tag{2.10}$$

Similarly, equation 2.9 can written in matrix form for the Inverse of 2D-DCT transform (equation 2.11).

$$Img = T^t * C * T \tag{2.11}$$

Because human faces have different patterns, automatic face recognition is a complex operation. In our face recognition technique by DCT, the target image features are extracted and form a feature vector. Then, the feature vector corresponding to the target image will be compared with the other face images stored inside the database. If the distance between the feature vector of the target and other feature vectors is greater than some predefined threshold value, the no-match flag will be set.

2.9.2.3 Hough Transform

Hough Transform (HT) [Hough V, 1962] is a feature extraction technique used in image analysis, computer vision, and digital image processing. HT is a technique that can isolate features of a particular shape within an image. This technique is particularly useful for computing a global description of a feature.

2.9.2.4 Malakooti Transform

Malakooti Transform (MT) [Malakooti and Dobuneh, 2012] is a new orthogonal transform, and it has been developed to represent the time series signals with a set of coefficients called the M coefficients. Many time-series signals are highly redundant; speech, image, and other periodic signals fall into this category. The M-transform is useful for the feature extraction since it represents the image with fewer coefficients, called the M coefficients.

This transform contains a complete orthonormal set that can span n-dimensional space and form a basis of independent and orthogonal vectors. All other vectors in this space can be represented as a linear combination of the independent vectors in the vector space.

In practice, it is assumed that the order-0 MT matrix, M0 is equal to one. Thus, the order-1 MT matrix M1 is formed according to :

$$M_1 = \begin{bmatrix} aM_0 & abM_0 \\ -abM_0 & aM_0 \end{bmatrix}$$

Where a and b are constant parameters.

We can use the following recursive equation to generate any size M-Transform depending upon the size of the target image.

$$M_k = \begin{bmatrix} aM_{k-1} & abM_{k-1} \\ -abM_{k-1} & aM_{k-1} \end{bmatrix}$$

2.9.2.5 Singular Value Decomposition

Singular Value Decomposition (SVD) is one of the strongest mathematical tools that can be used to decompose any square or nonsquare matrix, A, into the multiplication of two unitary matrices U and V and one Diagonal matrix Σ. This technique from linear algebra can be used to automatically perform dimensionality reduction. SVD is a matrix decomposition method for reducing a matrix to its constituent parts. The SVD is used widely in calculating other matrix operations, such as matrix inverse, and as a data reduction method in machine learning. SVD can also be used in the least-squares linear regression, image compression, and denoising data.
Given a nxm matrix A, the SVD of A is:

$$A = U.\Sigma.V^T$$

The diagonal values in the Σ matrix are the singular values of the original matrix A. The columns of the U matrix are called the left-singular vectors of A, and the columns of V^T are called the right-singular vectors of A where T is a superscript.
SVD can be thought of as a projection method where data with m-columns (features) is projected into a subspace with m or fewer columns while retaining the essence of the original data.

2.9.3 Matching

Image matching is an important concept in computer vision and objects recognition. In face and/or iris recognition, finding a matching means that the system recognizes the person; otherwise, this person is not registered in the database.
Before doing the matching, we must first find descriptive and invariant features for the images of each person. This features vectors of each person in the database will be used in the future to find a match with the test images.

In this thesis, a comparative study between three matching methods, namely the Euclidean distance, the Manhattan distance, and the Cosine distance, is carried out .

The Euclidean distance (ED) is one of the most widely used methods in the state of the art to compare two feature vectors. Such a measure can be used to find the closest person compared to the test image or to indicate that there is no matching if the distance exceeds a predefined threshold.
For an n-dimensional space, Euclidean distance is:

$$ED = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$$

Where p and q are the feature vectors of n dimension.

The Manhattan distance (MD) between two vectors equals the one-norm of the distance between the vectors. The distance function involved is also called the "taxi cab" metric. The Manhattan distance between two vectors (n-dimensional space) is defined as:

$$MD = \sum_{i=1}^{n}(|p_i - q_i|)$$

Cosine similarity (CS) is a measure of similarity that can be used to compare two vectors. A cosine value of 0 means that the two vectors are at 90 degrees to each other (orthogonal) and have no match. The closer the cosine value to 1, the smaller the angle and the greater the match between vectors.

The cosine similarity is defined as:

$$CS = \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} p_i^2}\sqrt{\sum_{i=1}^{n} q_i^2}}$$

2.10 Deep learning

As described in Figure 2.14, Machine Learning is part of Artificial Intelligence and allows machines to perform calculations to solve complex problems. The unique feature of Machine Learning is that the methods used will enable the system to learn how to perform a task based on a large amount of input data. Thus, the algorithm does not merely apply a set-point defined by its designer but adapts to the data transmitted to it to learn how to respond to the given problem.

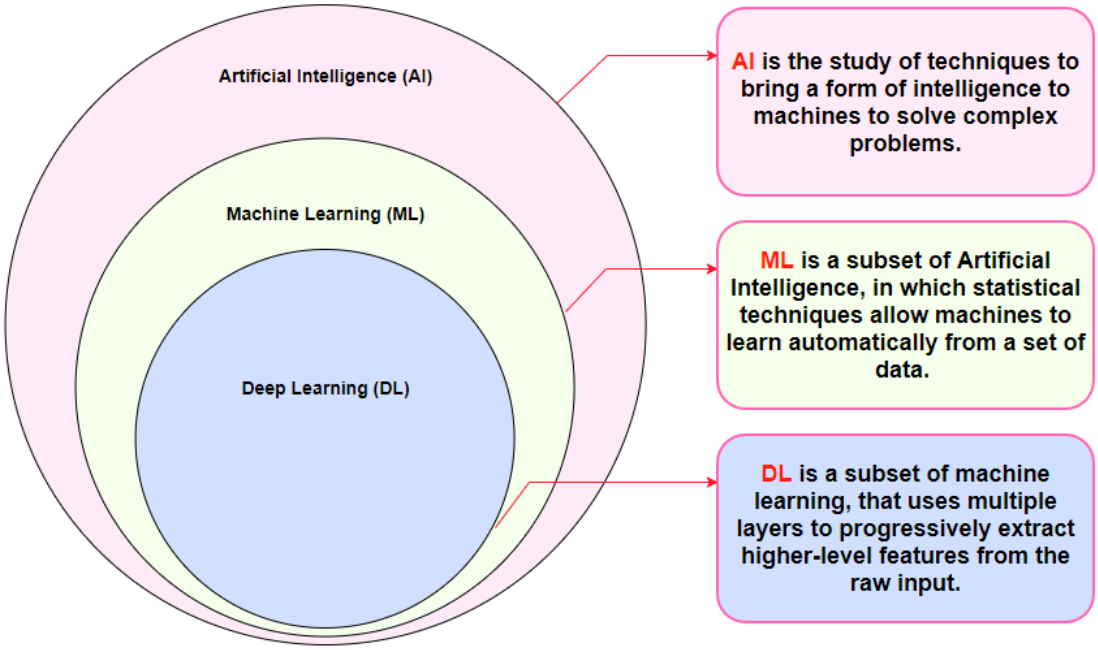


Figure 2.14: Relationship between the artificial intelligence, machine learning and deep learning

Deep Learning (also known as deep structured learning, hierarchical learning, or deep Machine Learning) is a Machine Learning branch. This subset of Machine Learning is based on the principle of Artificial Neural Networks (ANN), used on a much larger scale, based on the fact that increasing the number of layers and neurons in an ANN increases classification performance. Theorized initially in 1986 [Rumelhart et al., 1986] [Dechter, 1986], the concept of Deep Learning has only recently become popular [LeCun et al., 2015], requiring a considerable number of annotated data and a high

computing power.

Deep Learning, described in this section, is a subset of the Learning Machine. In the form of layers, deep learning implements a sequence of algorithmic treatments specific to Machine Learning to answer a complex problem divided into several tasks, each layer using the previous layer's output as input data.

Deep learning is based on artificial neural networks composed of thousands of neurons that perform small and simple operations. For example, the results of the first layer of neurons are used as input for the calculations of a second layer and so on.

The term deep generally refers to the number of hidden layers in the neural network. For example, conventional neural networks have only 2 to 3 hidden layers, while deep networks can have up to 150.

Advances in deep learning have been made possible by the increase in computer power and the development of large databases.

2.11 Artificial neural networks

In the human body, neural networks are the nervous system's building blocks that control and coordinate different human activities. Each neuron or nerve cell comprises a body of cells called Cyton and a fiber called Axon.
Neurons are interconnected by fibrous structures called dendrites using special connections called synapses. Electrical impulses (called action potentials) transmit information from one neuron to another through the network.
Human neural networks inspire artificial neural networks. The basic building block of each artificial neural network is an artificial neuron, i.e., a simple mathematical model (function). Such a model has three simple rules: multiplication, summation, and activation. At the artificial neuron's input, the information is weighted, which

means that each input value is multiplied by the individual weight [Krenker et al., 2011].

The sum function sums up all weighted inputs and biases in the artificial neuron's central part. At the artificial neuron's output, the sum of the previously weighted inputs and the bias passes through the activation function, also called the transfer function (Figure 2.15).

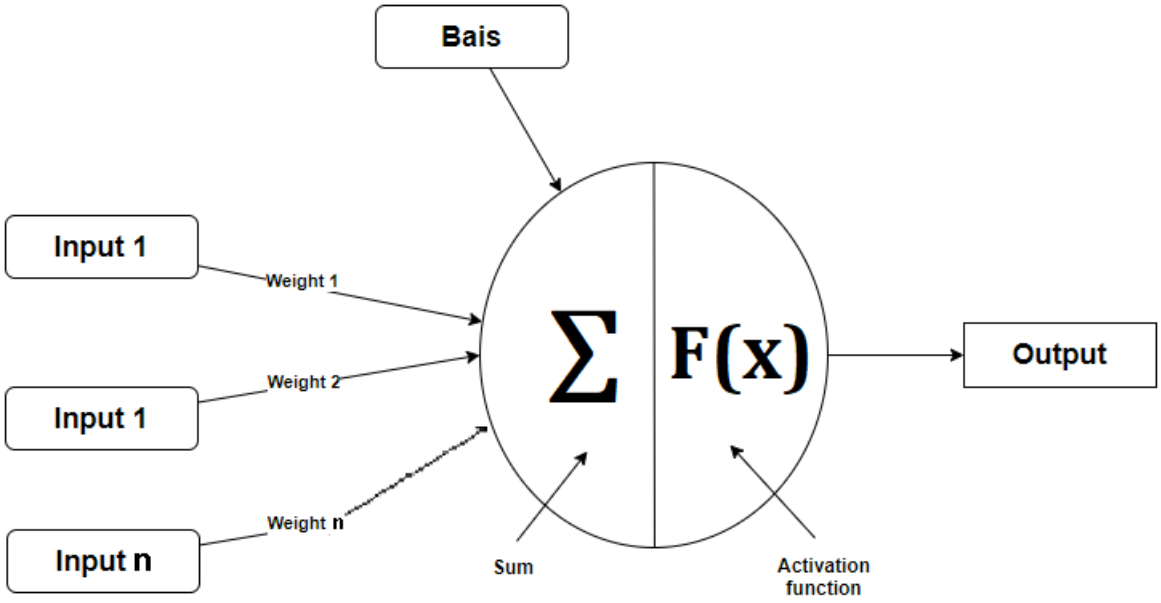


Figure 2.15: Working principle of an artificial neuron

Although the operating principles and the simple set of artificial neurons' rules look simple, these models' potential and computing power become interesting when we interconnect them in artificial neural networks.

An artificial neuron is a function f_j of input $x = (x_1, ..., x_d)$ weighted by a connection weight vector $w_j = (w_{j,1}, ..., w_{j,d})$, supplemented by a neural bias b_j and associated with an activation function f [Krenker et al., 2011], where:

$$y_i = f_j(x) = \phi((w_j, x) + b_j) \tag{2.12}$$

Several activation functions can be considered. Here are some activation functions often found in practice [Altenberger and Lenz, 2018]:

1 **The linear function or identity:** this function is denoted by:

$$\phi(x) = x \tag{2.13}$$

Single-layer neural networks use a step-by-step function when converting a continuously variable input function to a binary output (0 or 1) or a bipolar output (1 or -1).

2 **Binary function:** this function uses a threshold. A binary function with a threshold T is given by :

$$f(x) = \begin{cases} 1 & if x \geq T \\ 0 & otherwise \end{cases} \tag{2.14}$$

3 **The sigmoid (or logistic) function:** is a function of activation where :

$$\phi(x) = \frac{1}{1 + exp(-x)} \tag{2.15}$$

Its range is between 0 and 1. It is an S-shaped curve. It is easy to use and has all the interesting properties of the activation functions: non-linear, continuously differentiable, monotonous, and fixed output range.

4 **The Hyperbolic tangent function (Tanh):** its mathematical formula is :

$$\phi(x) = \frac{exp(x) - exp(-x)}{exp(x) + exp(-x)} \tag{2.16}$$

Tanh is also like logistic sigmoid but better. The range of the tanh function is between (-1 and 1). The tanh function is mainly used to classify between two classes.

5 **The Rectified Linear Unit (ReLu) activation function:** The ReLU is the most commonly used activation function. It is defined as :

$$\phi(x) = max(0, x) \tag{2.17}$$

The ReLU function is non-linear, so we can easily propagate errors backward, and the ReLU function activates several layers of neurons.

There are many classes of neural networks, which also have subclasses. Here we will list the most used ones:

2.11.1 Feed-forward neural network

The forward-propagating neural network was the first and most straightforward artificial neural network design. In this network, information is only routed in one direction, from the input nodes to the output nodes, through the hidden layers, to the output nodes.
There are no cycles or loops in the network. The best-known examples are the single perceptron and its multi-layer version [Fine et al., 1999].

The simple perceptron is a monolayer, acyclic (it has no loop) network whose dynamics (activity) are triggered by the reception of the captured information [Altenberger and Lenz, 2018].
This network is simple because it contains no hidden layers, meaning it consists only of an input layer and an output layer. These structures allow it to be considered a linear classifier: in other words, it can classify data according to two characteristics.

Neural networks, as presented above, comprise a limited number of neurons and therefore have limited capacity to deal with the increasingly complex problems in the

literature. To address these problems, Machine Learning structures must evolve to provide more complex learning models capable of processing information from thousands or even millions of data. One solution is to design neural networks with more hidden layers (and more neurons per layer), called deep neural networks [Schmidhuber, 2015][Montavon et al., 2018]. This increase in network capacity and complexity consequently increases the amount of neuron weight and the amount of information propagated, leading to an increase in the number of computations and thus the need for computing resources.

A DNN consists of an input layer, an output layer, and several hidden layers (see Figure 2.16). Each neuron in one layer directly connects to the neurons in the next layer.

In many applications, the units in these networks apply a sigmoid function as an activation function.

Multi-layer perceptrons are much more useful because they can learn non-linear representations (in most cases, the data presented to us are not linearly separable) [Touzet, 1992]. MLP is widely used to solve problems requiring supervised learning and research in computational neuroscience and parallel distributed processing. Applications include speech recognition, image recognition, and machine translation [Krenker et al., 2011].

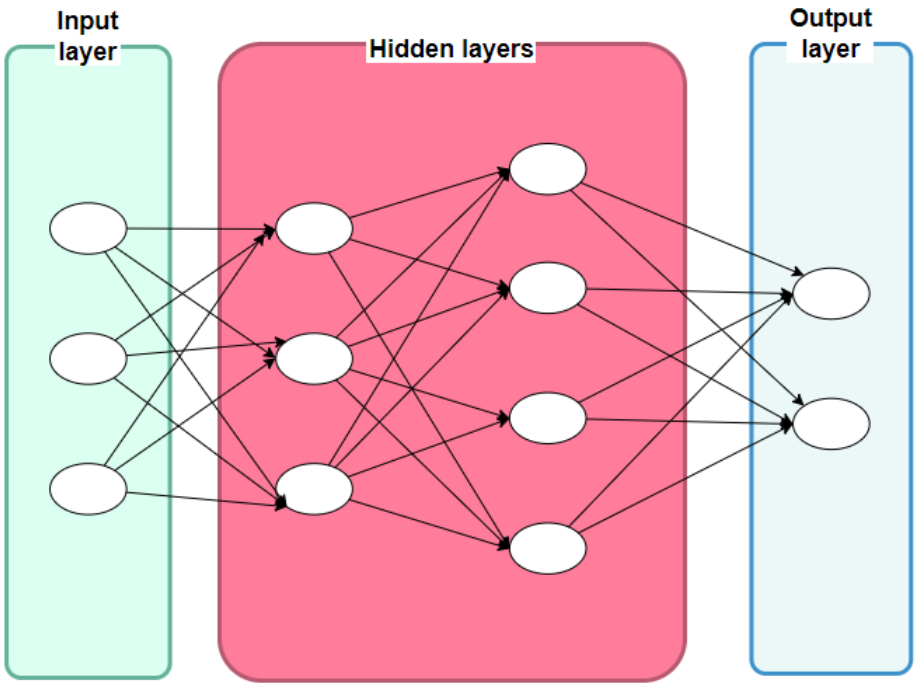


Figure 2.16: Example of a multi-layer perceptron

2.11.2 Recurrent neural network (RNN)

Recurrent neural network (RNN) are also similar feed-forward networks. However, they have recurrent connection loops, propagating the result of a neuron to the previous one or to itself (see Figure 2.17). Thus, the network keeps in "memory" all or part of the previous information and can use it to refine the following results. These networks are mainly used for predictive purposes such as text recognition or translation. Among the most used architectures in this family are Long Short-Term Memory (LSTM) [Hochreiter and Schmidhuber, 1997], Gated Recurrent Units (GRU)[Cho et al., 2014], and Transformers[Vaswani et al., 2017].

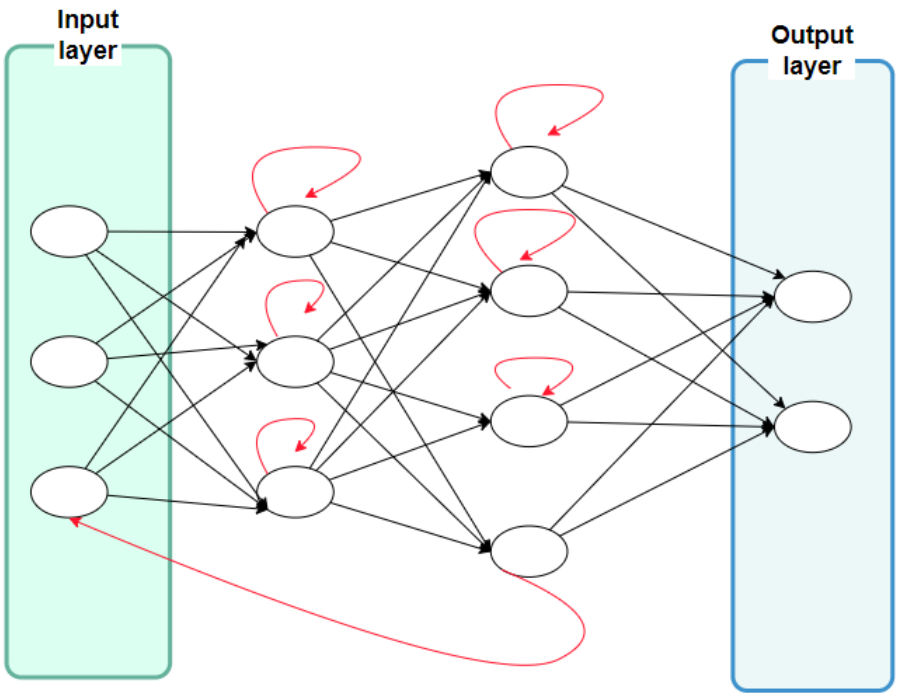


Figure 2.17: Example of a recurrent neural network

2.11.3 Convolutional neural network (CNN or ConvNet)

Convolutional neural networks (CNN) are similar to ordinary deep neural networks, but their architecture is specific to image processing. Indeed, the image information is processed at different points by convolution with several filters for each layer. The recent popularity of Deep Learning is due to these networks allowing, among other things, the latest advances in autonomous driving, complex image analysis, and so on. We will detail this model in the next section.

2.11.4 Convolutional Neural Network

One of the fundamental capacities of the human being is that of analyzing his environment. In most cases, this involves recognizing the elements in our field of vision: finding other people, identifying cars, animals...

Until the emergence of convolutional neural networks in 2012 with Alex Krizhevsky, the task was difficult for a computer. Fortunately, these networks' approach inspired by our eyes (especially since some neurons in our visual area only react to vertical borders and others to horizontal/diagonal ones) has opened many applications, whether in medical imaging, autonomous vehicles, facial recognition, and even text analysis.

Dedicated to image analysis, convolution neural networks (CNN) embark on the entire processing chain. Contrary to the classic Learning Machine and as shown in Figure 2.18, these CNNs can be seen as a black box using a set of learning images of the same size to adjust the many parameters of the network and thus specialize in a certain task. We propose here to present the elements constituting a CNN to facilitate the understanding of the functioning of these models.

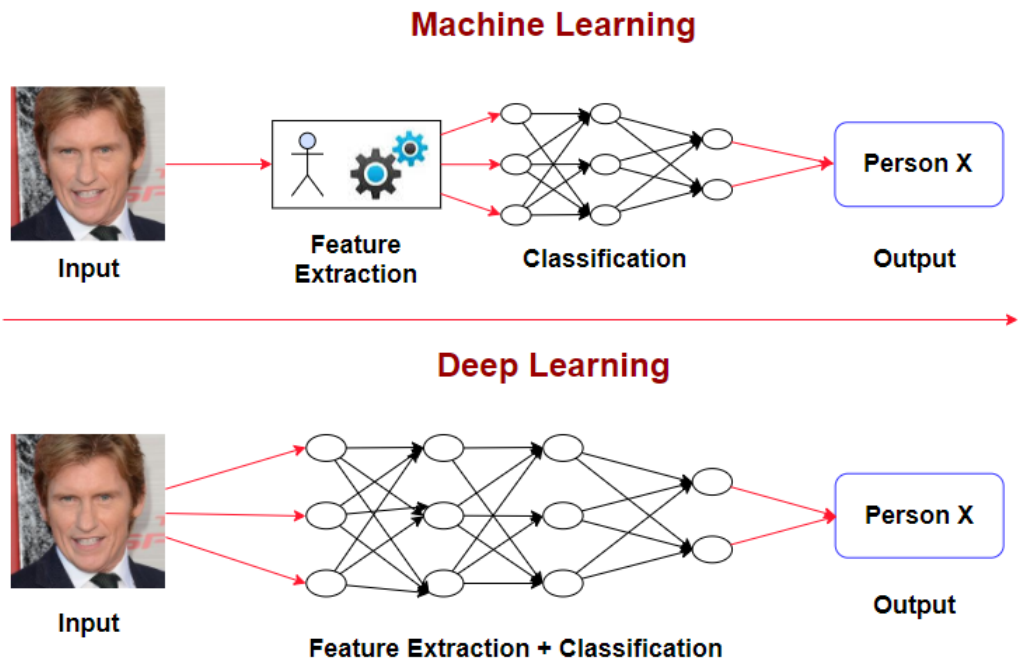


Figure 2.18: Feature extraction using Deep Learning and CNN

Figure 2.19 presents a schematization of the interaction between the various layers of the network with a color input image on three channels. A CNN is divided into several layers, each composed of other layers representing fundamental tasks of the

network. The part performing the extraction of the characteristics of an image is composed of layers called "hidden layers," and the classification part is called the classification layer or dense layer.

A hidden layer consists of one or more convolution layers associated with an activation function and a sub-sampling layer ("pooling"). The non-linear activation function, like for simple ANNs, allows the data values to be corrected by normalization. Many functions exist, the most used in the literature being the ReLU (Rectified Linear Unit) function, the Sigmoid function, or the hyperbolic tangent function. As its name indicates, the sub-sampling layer reduces the amount of data output from the convolution layer. Different types of sub-sampling exist, such as local averaging or local maximum.

The classification layer is divided into two layers. The first one, the vectorization layer, combines the local characteristics detected by the previous layers, freeing itself from their spatial structure. A second layer, the fully connected layer, allows a classification of the characteristics generated by the network, which have then been transformed into a data vector.

The last activation function, often of the "SoftMax" type, allows the normalization of the scores associated with each class.

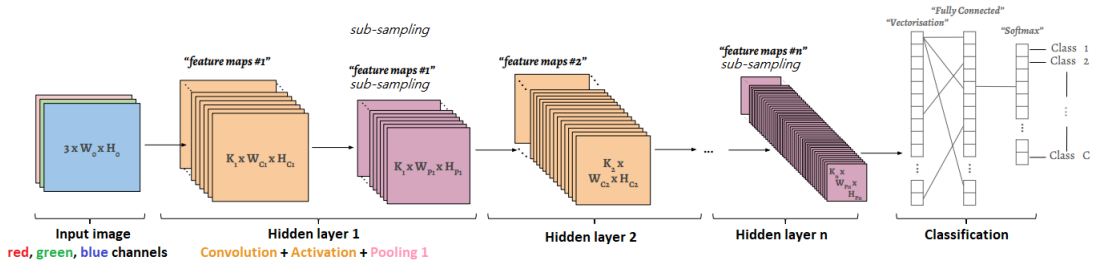


Figure 2.19: The layers of a CNN network

Each hidden layer "i" is made of K_i convolution filters (whose cores are of the same size for a layer), producing K_i new images (called "feature maps") of dimension $W_{Ci}.H_{Ci}$ smaller than the input image dimension $W_{i-1}.H_{i-1}$. An activation function is then applied to these feature maps, which are then subsampled, producing K_i subsampled images of $W_{Pi}.H_{Pi}$ dimensions. Several hidden layers are chained up to the classification layer. The information extracted before classification, initially abstract (outlines), forms high-level features (better representing the image) as the information progresses through the layers.

As shown in Figure X, the CNN contains different parameters and layers. Each of these layers has a particular function.

2.11.4.1 Convolution layer

Let's consider an image (or feature map) in the network. This image is of dimensions WHm, where "W" is its width in pixels, "H" its height in pixels, and "m" the number of channels of the image (m = 1: grayscale image; m = 3: color image).
The objective of the convolution layer is to extract characteristics of the input volume (an image of dimension "WxH" on "m" channels).
To do this, a convolution kernel of dimension "D_KxD_K" is applied at several points of the image following a sliding window moving by a step "s" ("stride"). This convolution consists first of a scalar product between a pixel of the image and the factor of the convolution kernel at the corresponding location and then to sum each of these products. The result of a convolution of layer "i" at a given location of the image, as shown in figure2.20, is thus a single scalar. The convolutions on the image represent a new image of reduced dimension "W_ixH_i".

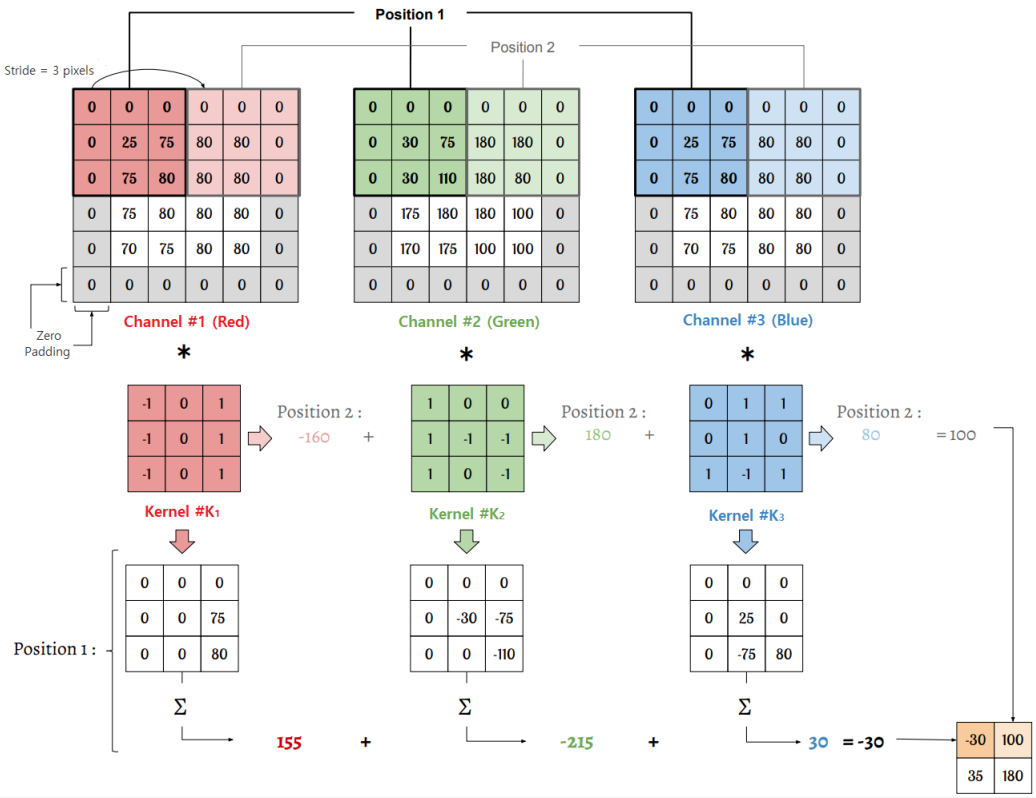


Figure 2.20: Illustration of a convolution for a 3-channel 4x4 image (RGB), a 3-pixel stride, a zero padding and a 3x3 convolution kernel

The stride (step) and the size of the convolution kernel are intended to control the size of the output image. For example, for a stride of 2, the convolution kernel moves two pixels between two convolution positions. The stride is selected so as to produce an image with integer (not fractional) dimensions. In the case of a stride smaller than the convolution kernel's width, there is an overlap with pixels used at the previous position of the kernel. To not lose the information on the edges of the image, the CNNs use "padding". This technique consists of adding a line and a column to each image border, allowing to consider the pixels at the image's border. These new pixels are usually set to zero paddings.

The image transferred at the network's input has pixels whose values are between a lower and an upper limit (0 and 255 for an 8-bit image). However, the resulting image may have values beyond these limits after a convolution. Therefore, a non-linear activation function, which can take different forms, is applied to them to normalize these values. As shown in figure 2.21, the main activation functions to determine a

new value based on the current pixel value. There are some activation features like Linear, Sigmoid, and Tanh, but the most commonly used is the "ReLU" function ("Rectified Linear Unit"). This is determined by the function "max(0, x)", which transforms negative values into zeroes and keeps the other values. This function does not change the feature maps' spatial resolution and does not include any network parameters.

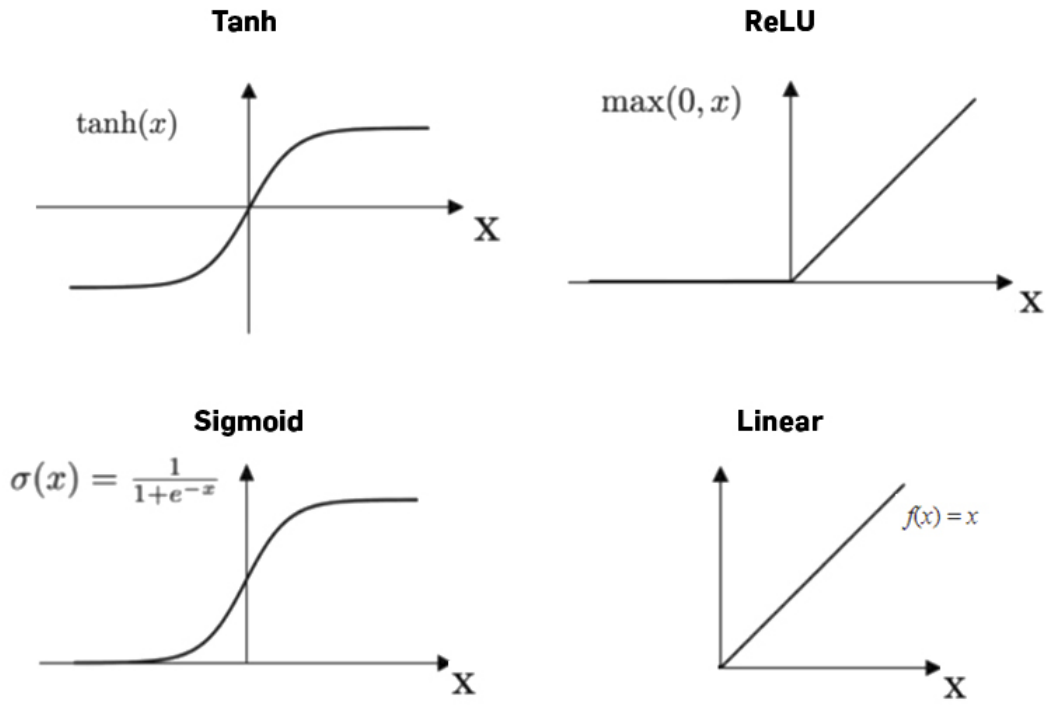


Figure 2.21: Examples of the best-known non-linear activation functions

The convolution layers have hyperparameters, i.e., parameters whose values are set during the network architecture design. There are four of them:

- the number of filters "K",
- the DF size of the filters (each filter is $D_F D_F m$ elements),
- the "s" stride with which the filter is dragged over the image,
- the zero-padding "P".

2.11.4.2 Subsampling layer (Pooling)

The objective of sub-sampling (or "pooling") is to reduce the spatial dimension of an input representation (image or feature map) in terms of height and width (not depth), allowing to simplify the features contained in the grouped sub-regions. The network's computational complexity is thus reduced by reducing the number of parameters to be learned in the following layers, and "pooling" produces translation invariance and some overfitting control. Similarly, a sliding window is moved over the image with a specific pixel size and a particular stride for convolution. There are different functions such as "max pooling" or "average pooling", the most commonly used is "max pooling". Max pooling selects the maximum value present in the window, while average pooling calculates the pixels' average in the window. This function operates on each image channel independently and does not require any parameters.

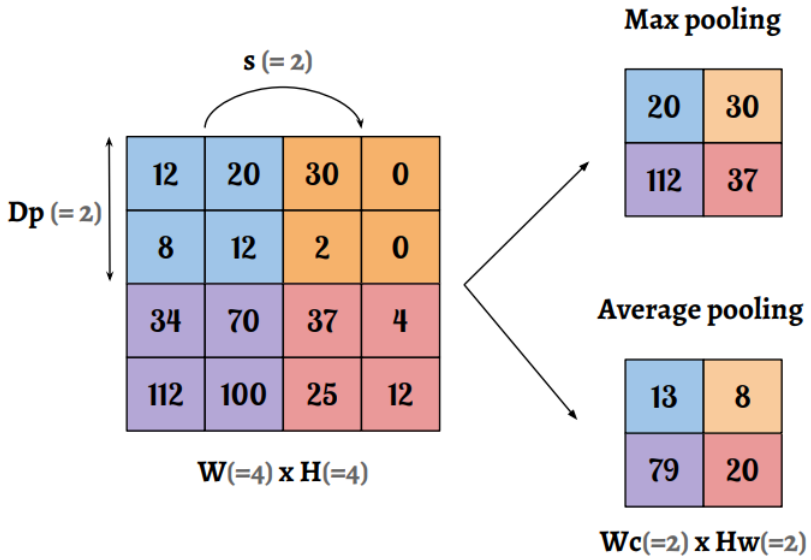


Figure 2.22: Illustration of the sub-sampling step (pooling)

Like the convolution layer, the pooling layer has two hyperparameters:

- the "D_P" size of the cells: the image is cut into square cells of size $D_P D_P$ pixels,
- the "s" stride: the cells are separated from each other by S pixels

2.11.4.3 Vectorization layer (Flattering)

After chaining several convolutions/pooling layers, the network's input image turns out to be a feature map with a very small width and height compared to this input image and a very large depth (number of channels). The classification is done by a fully connected layer, which we will see later, which does not consider any spatial structure. It is, therefore, no longer necessary to keep the information in the form of a feature map. Thus, a "flattening" or vectorization step is required to combine the local features detected by the previous layers. As shown in figure A.8, each channel's last feature map's elements are stacked to create a data vector. This feature map at the output of layer "n" thus has three dimensions, $W_n x H_n x D_n$, and the vectorization phase transforms it into a vector of dimensions $W_n.H_n.D_n$.

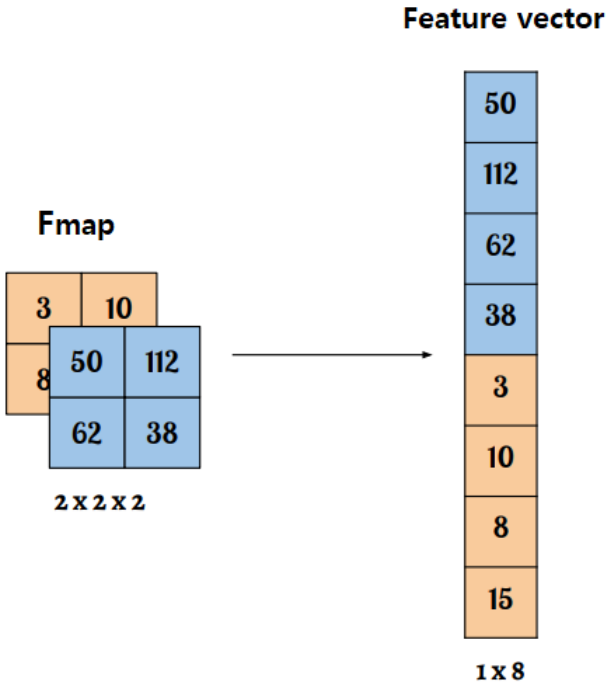


Figure 2.23: Illustration of the vectorization (flattering)

2.11.4.4 Fully-connected layer (classification)

As the name suggests, fully connected layers connect each neuron to the others. This set of fully connected layers forms a traditional neural network. To classify the characteristics generated by the network, which have then been transformed into a data vector, the last of the layers uses an activation function of the "*Softmax*" type, or exponential normalization function. The Softmax activation function normalizes the classes' scores to obtain values between 0 and 1. The sum of the results associated with each class equals one and represents a probability associated with each class (probability distribution). This activation function "fj(z)" is presented by equation 2.18, where "z" is a vector of "K" elements at the output of the network comprising the scores associated with the "K" classes and j corresponds to the jth class of the "z" vector.

$$f_j(z) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}, \quad \forall j \in [1; K] \quad with \sum_{k=1}^{K} f_j(z) = 1 \tag{2.18}$$

2.11.4.5 CNN parameters

A network has different parameters defining its architecture, and their number determines the storage size of the network. These parameters represent each convolution's kernel weights, the classification layer, and the associated biases. During training, it is possible to modify hyperparameters, which in turn determine the behavior of the network without modifying its architecture (such as stopping criteria or the precision of the parameter adjustment). For a better understanding of the interactions between the layers, C. Olah [Olah et al., 2018][Olah et al., 2017] proposes an advanced visualization of the progression (or propagation) of an image through the GoogLeNet network.

During the training, the network parameters are randomly initialized, and the learning images are propagated through the network. With the associated data labels, classification performance is evaluated by calculating a loss function (or cost function).

This function measures the classification error and the difference between the prediction probability and the field truth. The network weights and biases are then adjusted by an iterative optimization algorithm called gradient descent. A new evaluation produces a shift in the error towards a local or global minimum of the gradient of this cost function (Figure 2.24a). The step of this shift at each iteration is called the "learning rate". This step is a hyperparameter that can be adjusted to influence the convergence of learning. As described in Figure 2.24b, a large step covers a larger region of the gradient, but slopes towards a minimum, which can be very short, maybe missed. On the other hand, a small learning rate is more accurate and makes it easier to capture variations in slope but has a high cost in computing time, strongly related to the number of parameters to be adjusted. A compromise is therefore indispensable.

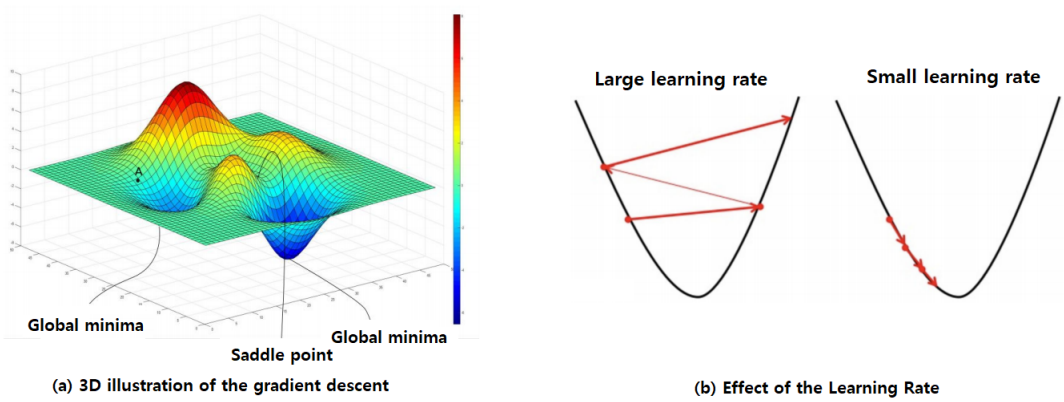


Figure 2.24: Optimization by gradient descent [Gómez Blas et al., 2020]

There are several types of gradient descent, differing mainly in the amount of data they use (called "batch"). The three main ones are :

- batch gradient descent: the error is calculated for each image of the learning set (m images), and the network is only optimized afterward (batch=m),

- stochastic gradient descent: The parameters are adjusted for each image. of learning (batch=1),
- the mini-batch gradient descends: the training data is divided into groups of n images, and optimization occurs after the error of these images has been calculated (batch = n).

The mini-batch size for the mini-batch gradient descending can be adjusted as a network learning parameter, just like the Learning Rate. The number of times all the learning images are presented to the network corresponds to the number of learning steps (or "epochs"). The number of iterations corresponds to the number of batches needed to complete an epoch.

2.12 Risk management

Today, companies do not want to leave anything to chance to develop their business. Therefore, to produce quality or succeed in the market, you must develop different strategies: commercial, political, operational, etc.

2.12.1 Definition of the risk

The risk is a potential danger that must be identified in a very precise manner. Its definition is intrinsically linked to the potentiality that it occurs, to its probability, which makes it feared. The seriousness of this one is evaluated on the importance of the consequences of such an event were to occur, but also of its acceptability.

Generally, there are three types of risk to consider and monitor. First, there is what is called strategic risk. It arises from decision-making and the definition of certain organizational orientations, but it can also arise from external factors. It is

therefore important here to monitor the strategy and the governance. Secondly, there are so-called operational risks, which impact an organization's ability to achieve the objectives it has set for itself, its production capacity, or its commercial strength. Again, this is more related to decision-making and the policies. Finally, there is an environmental risk, which is of external origin. This one is linked to a change of situation, which can be of several kinds (political, legal, economic, social, and even ecological).

2.12.2 Life cycle of risk management

Risk management experts work in many fields and with different structures. For example, they may have to advise companies, local authorities, or government departments on various projects (development, urban planning, politics, health, transport, etc.). Therefore, experts in risk management generally proceed in three steps.

2.12.2.1 Risk identification

To begin with, specialized consultants start by analyzing all the factors that could constitute a risk for the organization or the project. This involves a complete internal audit to assess the situation from the inside and then analyze the environment through a precise market study. As a result, it determines the nature of the risks and classifies them according to the typology mentioned above (strategic, operational, and environmental risks).

2.12.3 Assessing the seriousness of risks

Depending on the probability/severity ratio, it will be possible to classify the different types of risks identified on a scale of 1 to 4. To do this, a criticality matrix is created as a table, and a score is assigned to each risk. Then, the experts will develop a

more in-depth analysis of inherent criticality and residual criticality, i.e., the risk remaining after treatment. It is also necessary to distinguish those that must be treated in priority.

2.12.3.1 Risk management

This is the most crucial step, as it will allow the implementation of preventive measures within the organization. For example, some activities have quality controls and certifications (health, work, production, construction, crafts, services, technologies). This is also the case when employee training is implemented (for handling a machine or a motorized machine, for example). Preventive measures can also include establishing processes to be followed to limit the risk of a problem occurring. For example, care and cleaning protocols have been implemented to avoid nosocomial infections in hospitals. Experts can also advise you to avoid it simply by suspending an action. This could be a ban on using a product or beyond a specific date.
Then, when preventive measures have not been sufficient to avoid a risk, it will be a question of implementing corrective actions to reduce the consequences. For example, using a machine represents a risk of serious injury. Wearing safety equipment does not reduce the risk of a problem occurring, but it does mitigate its effects.
Palliative actions consist in transferring the risk to a third party. This is the role of insurance, for example. Finally, some risks are accepted because they are less critical. This is the case of a hairdresser who accepts ammonia-based products or a painter who prefers to work without gloves.

2.13 Cybersecurity

Information and Communications Technology (ICT) has grown and evolved rapidly due to the fast-growing computer technology, hardware, software, memory module, and data transmission. ICT devices and components are generally interdependent,

and disruption of one may affect many others. Over the past several years, experts and policymakers have expressed increasing concerns about protecting ICT systems from cyberattacks, which have greatly damaged the transmitted information, storage devices, and communication technologies.

The act of protecting ICT systems and their contents from unauthorized users and intruders is known to be cybersecurity. Cybersecurity is a useful tool that can be used to prevent unauthorized surveillance and protect information transmission, privacy, information sharing, and intelligence gathering. Risk management in any information system, along with cyber defense, is the key component of the operational plans. The risks associated with an attack depend on three factors: threats, vulnerabilities, and impacts.

2.13.1 Cyber threats

Most cyberattacks have limited impacts on the private sector. However, some attacks target the nation's critical infrastructure, such as the healthcare and financial sectors, which could have significant effects on national security, the economy, health and the safety of individual citizens. This type of risk must immediately be recognized by removing the source of threats, addressing the vulnerabilities, and reducing its impacts [Von Solms and Van Niekerk, 2013].

Cyber threats are not static and change quickly. Cyber adversaries are a growing threat to every business sector and government authorities involved in protecting personal data, customer data, banking information, military, and healthcare information. The traditional methods of data protection and cyber defense have been largely ineffective against current threats, and complex cybersecurity solution is required to protect the network security from intruders and prevent data transmission

from hackers and unauthorized users. Cyberattacks can be costly for individuals and organizations, and their economic impacts and damages are hard to measure. Hardware and software tend to get outdated at a much faster velocity than required to protect the information from the intrusion software and cyberattack tools, which are growing quickly. Also, most industries are vulnerable because they do not provide enough budget for their annual upgrade or cyber defense. A cybersecurity defense solution based on multimodal biometric authentication has been proposed to address these shortcomings with current cyber protection methods.

2.13.2 Security requirements

Hardware and software tend to get outdated much faster than is required to protect information from rapidly developing intrusion software and cyber attack tools. In addition, most industries are vulnerable because they do not put enough budget for annual upgrades or cyber defense.
A few requirements are necessary for better security:

2.13.2.1 Increase cybersecurity awareness

In addition to having robust security systems in place, having a team with cybersecurity skills and knowledge is your best defense against cybercriminal activity. Data breaches due to human error and negligence can cost millions, whereas organizing a cybersecurity awareness program is much more practical and relatively affordable.

2.13.2.2 Learn about different cyber-attacks

Another cybersecurity requirement is ensuring your team knows and understands the threat landscape or set of potential cyber threats they are most likely to face.

2.13.2.3 Keep your software and devices up-to-date

Cyber threats are constantly evolving and becoming more complex by the day. That is why software developers strive to keep their products up to date to ensure that their users can stay on top of the most common and recent cybercrimes.

2.13.2.4 Improve passwords

Passwords are crucial in protecting systems and databases from hackers and malicious attackers. Having strong passwords is the best way to keep cybercriminals at bay, as it greatly reduces the likelihood of them accessing your company's account. You and your team should know a few tips to meet this cybersecurity requirement. First, using different and unique passwords for each device and account is essential. Generally, passwords should be long and complex, preferably with upper and lower-case letters, numbers, and symbols.

2.13.2.5 Use biometrics

For security systems with physical person access, biometrics is still more advantageous and secure because it ensures that the person is physically present at the location. To go further, combining several biometric modalities can be more effective and secure. Therefore, this thesis proposes a cybersecurity defense solution based on multimodal biometric authentication to address the shortcomings of current cyber protection methods.

In a biometric system, the hardware (a camera, for example) captures the data and transfers it to a network to be read and processed by the authentication or identification platform.
The idea of safe and secure data transfer over the networks or cloud computing storage

facilities is the desire of all users, bankers, businessmen, scientists, and researchers. While the data is transferring over the network, it is exposed to a high risk of threats from hackers and attackers [Nemati et al., 2010]. Data can be changed, corrupted, stolen, or even lost during transmission. Also, the intruders may be able to attack the network and create Denial of Service in which people will be denied from accessing our website, web servers, and database information. The major concern is that data will not be damaged or tampered with by intruders while it is transferring across the networks. Cryptography is the art of data security and protection over wired or wireless networks. Cryptography is based on the symmetric keys algorithm like DES, AES, transformation algorithm based on the orthogonal transforms [Malakooti and Mansourzadeh, 2014b][Nemati et al., 2010], or asymmetric keys algorithm like RSA [Nemati et al., 2010], and each one has its advantage and disadvantages. The asymmetric encryption algorithm is more secure than the symmetric encryption algorithm, but the speed of operation is lower than the symmetric one. In both encryption methods, the generations of the keys are very important, and it must be a randomized sequence of numbers and unpredictable by the intruders. The correction of the randomized numbers is calculated, and along with concepts of the autocorrelation, it will generate the keys that are unpredictable [Yassein et al., 2017].

The cybersecurity problem cannot be solved by cryptography, although the encryption algorithm is very useful in protecting the content of databases, data centers, and cloud computing storage facilities. Among the important challenges of cybersecurity is the presentation of fake samples at the identification site to act as a genuine client by spoofing the recognition system in which a fake fingerprint can complete the authentication process, printed cosmetic lens pattern, or even by face mask [Alam, 2022][Guarnera et al., 2022]. Indeed, if the system does not detect the difference between this fake biometric and the real one, the impostor will be allowed access, hence the advantage of using multimodal biometrics.

Another type of attack can occur if a hacker gets between the acquisition device and the authentication server and replaces the image of the impostor with an image of

an authorized person, hence the advantage of encrypting the image at the time of acquisition and the platform takes care of decrypting it just before authentication. However, although this solution provides more security, it adds a layer of complexity and additional processing time due to the need to encrypt the image on edge and decrypt it in the processing server.

2.14 Cryptography

We live in an era of emerging new technology and a revolution in ICT, where a huge amount of data are transmitted over local area networks, distributed networks, intranets, and the internet. Some of the information will be stored on different networks, and cloud storage will eventually be out of our control. When any information is transmitted from our network, we are not the only ones who can trace the information. Some hackers, intruders, and illegal users also try to access the information to steal or tamper with its contents. One way to protect our information during the transmission is using a virtual private network or secure communication so that eavesdroppers or interceptors cannot intercept our communication. However, it cannot guarantee that our information will be safe on the network facilities or cloud storage. Encryption is the final solution for the information security problem during the transmission and while it's stored on the cloud facilities [Paar and Pelzl, 2009]. In an authentication system, the biometric recognition information must be encrypted in the network to ensure its integrity and security.

Encryption is a method of coding in which the original data are converted into some format so that only those who know the algorithm's structure and some secret keys can convert the coded data into original data. Information cryptography is the process of hiding information from hackers, intruders, and illegal users. It studies mathematical techniques related to information security such as confidentiality, data

integrity, entity authenticity, and data origin authentication.

Modern cryptography intersects the disciplines of mathematics, computer science, computer engineering, probabilities, and statistics as well as engineering. Cryptography applications include ATM cards, computer passwords, e-commerce, bank transactions, online shopping, social media, and voice-over IP. Traditional encryption started with the invention of writing, and it was the science of writing secret codes in ancient art. Humans have used encryption to send secret information, especially during wartime plans or secret commercial or diplomatic messages. Since the invention of the digital computer, modern cryptography has been used to deliver secure messages over the communication media, trusted networks, and particularly the internet. Cryptography is the science of keeping the message secure and converting it into a coded format unknown to those not authorized to know the coding structure and secret keys. In cryptography, the original message is called the plaintext, and the encrypted message is called the ciphertext [Paar and Pelzl, 2009].

2.14.1 Encryption/Decryption

Encryption is the process of transforming information or plaintext into a coded form by using a mathematical algorithm to make it unreadable code or ciphertext. The result of the operation from the encryption is called encrypted information or ciphertext. The reverse operation of the encryption is called decryption, in which the encrypted information is the input, and the result of decryption will create the decrypted information of plaintext [Paar and Pelzl, 2009][Malakooti et al., 2012]. The government organization, militaries, defense departments, banks, big corporations, and enterprises were the primary users of the encryption. However, now, all private sectors, social media, news media, and every business activity have used encryption to protect their information from illegal use.

For example, Sandvine's Global Internet Phenomena Reported that over 50% of internet traffic is now encrypted, 71% of companies utilize encryption for some of their data during transmission, and 53% utilize the encryption for their data stored in the cloud or even their storage facilities [Shbair et al., 2020].

If we use cryptography to encryption some or all vital information which is stored in our storage facilities, cloud, or our personal computer, the security of information has been increased, and in the event of losing our laptop or being stolen by a thief or breach of security the contents of our vital information is already protected.

The process of encryption can only protect the confidentiality of the information, but other techniques are required to protect the integrity and authenticity of the information. For example, the network itself must be protected and secured and use other software and facilities to ensure the authenticity of data passing through our networks, such as verification of a message authentication code or a digital signature.

2.14.2 Scrambling

Digital image scrambling is a mathematical technique and algorithm that can be used to convert a regular digital image into some type of meaningless image that no one can recognize the content of the original image from it. The algorithm must be reversible, which means the digital scramble image can be converted back to the original image with an inversion algorithm. This technique can be used to hide the identity of information from illegal users and increase security. A scrambling algorithm is a simple encryption algorithm that can be used for low-level security encryption. Researchers have developed various image scrambling techniques to hide the content of the original image and convert it into an unrecognizable or unintelligent image [Jiancheng Zou et al., 2004]. The Home Box Office (HBO) and Cinemax have started the full-time signal protection by applying the scrambling algorithm on their entertainment programs, and home dish owners must sign a contract to receive the

descrambling program to be able to see the contents of each movie channel.

2.15 Multimodal biometrics

Biometric authentication and recognition based on the features obtained from the individual are the most reliable technique of authentication and recognition because most of the individual characteristics and traits, i.e., an iris image, will not be changed from fetal life to end of life.

Cybersecurity has encountered many challenges, especially when intruders have learned how to use advanced technologies and sensitive substances to build fake identification tools such as a fake finger, fake face, and printout of iris on the cosmetic lens to spoof the system and breach the security.
Over the years, several researchers have tried to find solutions to its problems through multimodal biometrics using machine learning or deep learning. In what follows, some state-of-the-art works related to this thesis will be presented.

2.15.1 Multimodal biometrics using machine learning techniques

The good recognition of multimodal systems depends on multiple factors, such as the fusion scheme, the fusion technique, the features extraction techniques, and the used classification method.

In recent years, several researchers have focused on the proposal of reliable, multimodal biometric systems. The proposed models have been based on combining at least two characteristics. Among the works are those based on face and voice [Jain

et al., 2005a][Lin Hong and Anil Jain, 1998], face and fingerprint [Brunelli and Falavigna, 1995], face and palm print [Ross and Jain, 2003], fingerprint and iris [Elhoseny et al., 2018], face and iris [Morizet and Gilles, 2008] [Ammour et al., 2020], or fingerprint and hand geometry [Kittler and Messer, 2002] [Feng et al., 2004].

This thesis proposes a biometric system based on the face and iris. This choice is motivated by the fact that the face is the natural means of identifying people, and the iris is currently considered one of the most accurate biometric systems [Ammour et al., 2020].

Eskandri and Toygar proposed to use left and right iris patterns with optimized features of local and global based facial feature extraction methods using Particle Swarm Optimization (PSO) and Backtracking Search Algorithm (BSA) to remove redundant data for the fusion of face-iris multimodal system with tanh score normalization and Weighted Sum Rule fusion method where the weights are also optimized using PSO and BSA [Eskandari and Önsen Toygar, 2015]. Matching score level and feature level fusion techniques were used to test the proposed models. The authors used the Chinese Academy of Science Institute of Automation (CASIA) V1 Iris Distance database, Print Attack face database, Replay Attack face database, and IIIT-Delhi Contact Lens iris database. The obtained identification and verification rates clarify that the proposed fusion schemes significantly improve their study's unimodal and other multimodal methods.

In another work by [Morizet and Gilles, 2008], the authors proposed an adaptive combination approach for merging score levels for facial and iris biometrics by combining wavelets and statistical moments. They showed that they could achieve good recognition using the Log-Gabor Principal Component Analysis (LGPCA) method for facial feature extraction, the 3-level wavelet packets method for iris feature extrac-

tion, and the recognition was performed using Cosine similarity. The face recognition technology (FERET) and CASIA V3 databases were used to test their approaches. The results show that the method is very competitive in FAR and FRR.

In the study [Rattani and Tistarelli, 2009], the authors proposed a robust multimodal and multi-unit feature level fusion of face and iris biometrics. They used Scale Invariant Feature Transform (SIFT) for feature extraction, the spatial sampling method used for the selection process, and the Euclidean distance for the matching. They used the CASIA V3 iris and Equinox faces databases to test their face and iris feature level fusion approaches. The reported results show the performance improvements in multimodal and multi-unit biometrics classification compared to uni-modal classification and score level fusion.

In their article [Huo et al., 2015], the authors have proposed a face–iris multimodal biometric scheme based on feature-level fusion. They used a 2D Gabor filter with different scales and orientations for feature extraction on the face and iris, then transformed them by histogram statistics into an energy orientation. After that, the PCA method was used for dimensionality reduction. Finally, they used Support Vector Machine (SVM) for the matching. Compared with some state-of-the-art fusion methods, experimental results demonstrate that this method provides higher recognition accuracy.

In the work [Eskandari, 2017], the authors proposed a multimodal face–iris biometric system that combines the advantages of score level, feature level, and decision level fusion by considering the optimized information of face and iris biometrics at each level of fusion. The optimized output of one fusion level provides appropriate input for the next fusion level to construct a new and efficient scheme. The optimized scores are computed based on the extracted and fused optimized features of the face and

iris modalities. Finally, the decisions are made according to the optimized Receiver Operator Characteristic's (ROC) obtained from score level fusion. The authors have used the Log-Gabor transform to extract the facial and Iris features and the BSA feature selection algorithm to select the relevant features. Experimental results on four databases (Olivetti Research Laboratory (ORL), CASIA V1, PIE-illum, and CASIA V4-Lamp) demonstrate the proposed combined level fusion scheme significantly over unimodal and multimodal fusion methods.

In the study [Matin et al., 2017], the authors proposed a weighted score level fusion technique to combine face and iris. They used the Daugman method for iris recognition, where an automatic segmentation is performed using circular Hough transform to localize the iris and pupil area. A 1D Log-Gabor filter was used to encode the unique features of the iris into a binary template. A Principal Component Analysis (PCA) based method was used for face recognition. For the matching, they have used the Hamming distance for iris and the Euclidean distance for faces. The min-max normalization technique has been used to normalize the iris' matching score and face recognition. The normalized scores are merged as a single score using the weighted sum rule. The obtained results on the CASIA V4 and ORL Face databases using the developed multimodal technique improve the system's recognition accuracy and robustness.

In [Bouzouina and Hamami, 2017], the authors proposed an Iris and face recognition system based on PCA and Discrete Coefficient Transform (DCT) for facial features extraction. In contrast, iris features were extracted with the 1D Log-Gabor filter method and Zernike moment. They used feature selection with Genetic Algorithms (GA) and scores level fusion with SVM. They obtained good performances using the CASIA-IrisV3-Interval database.

In another work [Ammour et al., 2018], the authors proposed a Face-Iris multimodal biometric system based on hybrid level fusion. They used a multi-resolution two-dimensional Log-Gabor filter combined with Spectral Regression Kernel Discriminant Analysis (SRKDA) for the feature extraction. For the matching, they used the Euclidean distance. The proposed system was evaluated on the CASIA Iris Distance database and compared to existing state-of-the-art systems. The obtained results have shown that the proposed model outperforms the existing methods in the verification mode.

In a recent work [Ammour et al., 2020], the authors presented a Face–Iris multimodal biometric identification system based on a multi-resolution 2D Log-Gabor filter for iris features extraction and the Singular Spectrum Analysis (SSA) combined with the Normal Inverse Gaussian (NIG) statistical features derived from wavelet transform for the facial features extraction. The matching was performed using the Fuzzy K-Nearest Neighbor (FK-NN). They used a chimeric database consisting of ORL and FERET for face and CASIA v3.0 iris image database (CASIA V3) interval for iris. Experimental results show the robustness of the proposed model.

Table 2.3: Summary of the state of the art of machine learning methods

Work	Face and iris feature extraction	Fusion	Advantages	Inconvenients
[Eskandari and Önsen Toygar, 2015]	PSO + BSA	Consideration of all face and both left and right iris	Good feature extraction	Score level fusion classifies each modality separately. One image for face and iris. Need a very hight resolution image to capture iris information
[Rattani and Tistarelli, 2009]	SIFT	Feature level	Easy implementation	One image for face and iris. Need a very hight resolution image to capture iris information. Only support a one-to-one verification model.
[Huo et al., 2015]	2D Gabor filter + PCA	Score level	Good feature extraction	Score level fusion classifies each modality separately.
[Eskandari, 2017]	Log-Gabor transform + BSA	Score level + feature level + decision level	Good feature extraction	Complex method
[Matin et al., 2017]	Daugman method + 1D Log-Gabor filter	Score level	Good feature extraction	Score level fusion classifies each modality separately.
[Bouzouina and Hamami, 2017]	DCT + PCA + GA Zernike moment + 1D Log-Gabor	Feature level	Good feature extraction	Complex approach. No comparison is given with the state of art.
[Ammour et al., 2018]	SRKDA	hybrid level fusion	Good feature extraction	Complex approach.
[Ammour et al., 2020]	2D Log-Gabor filter + SSA + NIG	Score level	Good feature extraction	Score level fusion classifies each modality separately.

Works in literature, summarized in Table 2.3, have shown that combining biometric templates at different fusion levels and using different feature extraction techniques improve the system's accuracy. The cited works use different datasets to evaluate their algorithms, such as the CASIA iris dataset, IIT Delhi Iris dataset, ORL face dataset, FERET faces dataset, Face94 face dataset, etc. These works generally use two different algorithms for feature extraction of the face and Iris, making the system more complex. Also, works using the score-level fusion classify each modality separately and do not benefit from the powerful of the multimodal scheme. As a solution, the goal of this thesis, in its first part, is to propose a single algorithm for extracting features on the face and iris.

2.15.2 Multimodal biometrics using deep learning techniques

Multimodal biometric systems have been the subject of much research, and several approaches have been proposed to build these systems by efficiently combining biometric data from several sensors.

The previous work in the literature shows that when using a machine learning approach to recognize biometrics, the biometrics images require specialized feature extraction algorithms depending on the biometric type. Sometimes the images need several pre-processing stages [Veluchamy and Karlmarx, 2017]. Feature extraction is a component of image processing that often goes hand in hand with classification. Indeed, to establish a classification rule (supervised or unsupervised), we generally base ourselves on a set of numerical criteria describing the object.

In classical Machine Learning, when working on a case such as image recognition, selecting the feature extraction method is crucial and will influence the prediction. If the characterization is not representative of the image, the model will have difficulties recognizing people, which remains a significant challenge in the field.

Deep Learning, with its ability to extract features automatically, has made recogni-

tion more efficient and accurate.

Recently, deep learning has provided great results in multimodal biometrics systems [Arora et al., 2021][Alay and Al-Baity, 2020a]. In addition, the limitations of classical machine learning algorithms, particularly those associated with feature extraction techniques, have been overcome by deep learning algorithms. A deep network with a hidden multi-layer neural network architecture is a promising study for exploiting efficient information for data collection, inspiring multimodal deep learning.

Omara et al. [Omara et al., 2017] have proposed a deep learning multimodal biometric system that uses the face and ears. The deep features extracted for the face and ear images are based on VGG-M Net and Discriminant Correlation Analysis (DCA). They are used for fusion and reduction of the vector size, then SVM is used for the classification. The system has achieved a recognition rate of 100% on the USTB dataset [Soleymani et al., 2018] with the collection I and II for ear and ORL dataset [Gunasekaran et al., 2019] for the faces.

Kurban et al. [Kurban et al., 2017] proposed a multimodal biometrics system that integrates the image of the face and energy gesture. VGG's deep learning model was used as an extractor of characteristics for the face database, and a method of energy imaging was used to extract the characteristics of the gestures. Then, PCA is used to perform a dimensionality reduction of the feature vectors. Finally, similar scores were obtained using the Euclidean distance. These scores were merged using the sum rule.

Shams et al. [Shams et al., 2017] presented a multimodal biometric fingerprint and face retrieval system based on Adaptive Deep Learning Vector Quantiza- tion (ADLVQ). This paper's proposed system extracts the input modalities' characteris-

tics using the Local Gradient Pattern with Variance (LGPV). Then, the K-means algorithm is applied to the vector quantification. Finally, these characteristics are classified using a deep neural network based on the knowledge acquired during the vector quantization.

Al-Waisy et al. [Al-Waisy et al., 2017a] proposed a multimodal biometrics identification that uses a parallel architecture to fuse the results obtained with the face and left and right irises. First, the 3-layer Deep Belief Network (DBN) architecture is used to extract facial features. The first two DBNs are used as characteristics detectors, and the latter is used as a discriminant model associated with softmax units for multi-classification purposes.
Secondary, for iris recognition, a deep learning system is used. It is based on combining a convolutional neural network and a softmax classifier to extract the features discriminating image of the iris.

In [Talreja et al., 2017], the authors presented a secure multibiometric system that combines the face and the iris. In addition, it uses CNNs for feature extraction. Two fusion architectures, one fully connected architecture and a bilinear architecture are implemented to produce a shared robust multibiometric representation.
Then, to generate the multibiometric model, the dimension of the final merged feature vector is reduced by applying a feature selection process.

In [Al-Waisy et al., 2018], the authors presented a fast multimodal biometric system based on the right and left iris of the same person using a score level fusion method. They used a combination of a CNN architecture and a softmax classifier for feature extraction and classification.

Yang et al. [Yang et al., 2018] proposed a new multimodal biometric recognition model based on stacked methods ELMs (Stacked Extreme Learning Machines) and Canonical Correlation Analysis (CCA) methods. The model, which has a symmetrical structure, has a high potential for multimodal biometrics.

The model works as follows: First, it learns the representation of the hidden layer of images using extreme learning machines, layer by layer.

Second, the CCA method is applied to map the representation on a feature space, which is used to reconstruct the representation of multimodal image characteristics. Thirdly, the reconstructed features are used as inputs for a supervised classifier. To test the validity and effectiveness of the method, the authors adopt new hybrid datasets obtained from face and finger vein images.

Kim et al. [Kim et al., 2018] presented a multimodal biometric system that combines finger vein and finger shape using a near-infrared camera sensor based on a deep convolution neural network. Corresponding distances calculated based on vein characteristics and finger shape obtained using ResNet models were merged using various fusion methods such as weight sum, weighted product, and the perceptron.

Soleymani et al. [Soleymani et al., 2018] proposed a common CNN architecture with feature-level fusion for multimodal recognition using multiple modalities: face, iris, and fingerprint.

Rather than merging networks at the softmax layer, the optimal compression characteristics of all modalities are merged at the fully connected layers without loss of accuracy but with a significant reduction in the network's number of parameters.

Tiong et al. [Tiong et al., 2019] proposed a multi-deep learning network for the facial recognition system by combining the periocular and the facial characteristics. They further improved the recognition accuracy by combining the textural and multimodal

features.

Umer et al. [Umer et al., 2020] combined the periocular and iris features for a person's biometric recognition. They deployed different deep learning-based CNN frameworks such as ResNet-50, VGG-16, and Inception-v3 for feature extraction and classification. They demonstrated that combining the features from various traits improves the system's performance.

The features from fingerprints and Electrocardiogram (ECG) were fused by Jomaa et al. [M. Jomaa et al., 2020] to detect the presentation attacks. They deployed three CNN architectures, including fully connected layers, 1-D CNN and 2-D CNN for the feature extraction from ECG, and an EfficientNet for the feature extraction from the fingerprint.

In [Alay and Al-Baity, 2020a], the authors proposed a multimodal biometric human identification system based on three convolutional neural networks (VGG16 architecture) to extract features from iris, face, and finger vein biometric modalities. The classification was performed using a softmax classifier, and the feature vectors were combined using feature and score level fusion approaches.

[Leghari et al., 2021] proposed a multimodal system based on the feature-level fusion of fingerprint and online signature. The authors proposed using an early (before the fully connected layer) and a late (after the fully connected layer) feature fusion scheme to combine the proposed CNN architectures of the fingerprint and the online signature.

From the previous work in literature, it can be noticed that when using a machine

learning approach to recognize biometrics, the biometrics images require specialized feature extraction algorithms depending on the biometric type. Sometimes the images need several pre-processing stages [Veluchamy and Karlmarx, 2017]. While in the deep learning approach, the deep learning network extracts the features from the images automatically.

The performance of deep learning approaches is generally better than the machine learning approaches. However, while effective, these deep learning-based approaches are computationally expensive and time-consuming [Arora et al., 2021][Alay and Al-Baity, 2020a]. Studies deploying deep learning algorithms in multimodal biometric systems [Al-Waisy et al., 2018] begin the experimentation process by applying region detection methods before entering data into a deep learning model. The use of region detection methods requires selecting a suitable technique for a particular trait. Also, the process can be time-consuming [Anil K. et al., 2011]. It is also noticed that the physical biometric traits deliver better performance than behavioral biometric traits [Arora et al., 2021]. Moreover, the iris trait tends to increase the accuracy rate [Al-Waisy et al., 2018][Al-Waisy et al., 2017b].

Based on what has been observed in previous studies, this work develops an identification multimodal biometric system that combines face and iris using a combination of some CNN architectures. Feature level fusion and decision level fusion were applied to determine the most effective approach. An end-to-end CNN algorithm will be used to extract features and recognize images.

2.16 Summary

After introducing the operating modes, characteristics, and evaluation tools of a biometric system, it can be observed that the design of a multimodal biometric system depends on several factors such as the sources of merged information, the architec-

ture, and the level of fusion.

Several works carried out in the literature related to our thesis have also been presented. The study of these methods allowed us to synthesize the advantages and disadvantages and propose contributions to the domain.

The proposed methods based on machine learning and deep learning will be presented in the following chapters.

Chapter 3

Proposed Cyber Security Biometric Platform Solution using Machine and Deep Learning Classifiers

3.1 Introduction

Many biometric systems are based on a single characteristic of the human body. These systems have many limitations related to this single characteristic, such as noise, intra-class variation, and identity fraud. [Ammour et al., 2018][Kabir et al., 2018].

To overcome these problems, multimodal biometrics was developed. This technique combines information from several biometric sources and is also known as information fusion. In this thesis, feature-level fusion is used. It combines different feature vectors generated from different biometric modalities to create a single template or

feature vector [Kabir et al., 2018]. Recognition is based on a comparison between the test sample and the template stored in the database as an indicator of similarity for the modalities.

A biometric system has two phases, training and recognition. For the training phase, the biometric modality is captured and processed using specific algorithms to obtain a biometric reference template for each user, which is stored in a database. For the recognition phase, a biometric sample is captured and processed as in the training phase and compared with the biometric templates stored in the database. The result is either a match found (if the user is found in the database) or not recognized [Jamdar and Boke, 2017].

The insufficient accuracy and reliability of unimodal biometric systems have led many end-users to use multimodal biometric systems to provide the maximum level of accurate authentication. These systems use information from two or more biometric systems. A multimodal biometric system provides a greater level of assurance for an accurate match in verification and identification modes because it uses multiple biometric traits, and each trait can provide additional evidence of the authenticity of an identity claim.
Multimodal biometric systems are indispensable in areas where maximum security and accuracy are required and where a single mistake can result in the death of many civilians or cause great material or human damage. Therefore, they are best suited for sectors such as the military, healthcare, civil identification, voter registration, e-passport, and financial sectors.

In this thesis, the problem of recognizing people using multimodal biometrics is considered. The face and the iris are chosen as biometric identifiers since the face is the simplest way to recognize a person, and the iris has a unique texture.
In such an identification system, two cameras are needed, one for the face and one

for the iris (see figure 5.4). The first camera is used to collect the image of the face, while the second sensor is used to capture the image of the iris. These two images will be used for the recognition of this person. Such a model could be used as a multimodal identification system in different sectors.

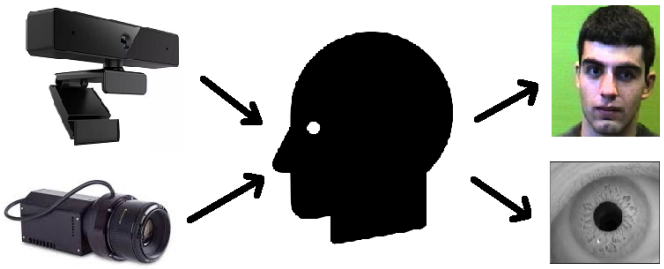


Figure 3.1: Face and iris acquisition [Hond and Spacek, 1997][Kumar and Passi, 2010]

3.2 Face and iris cyber security biometric platform using machine learning

The main contribution of this part of the thesis is to pre-process the images before exposing them to the recognition model. This model exploits the performance of the DWT method for feature extraction and then uses the SVD method to reduce the size of the feature space. The Euclidean distance method was used to find the match in the database.

3.2.1 Proposed multimodal biometric recognition system

This part of the thesis will focus on the face–iris multimodal biometric system based on DWT and SVD for feature extraction and feature level fusion as the fusion strategy of the system. The proposed approach is described and detailed in this section.

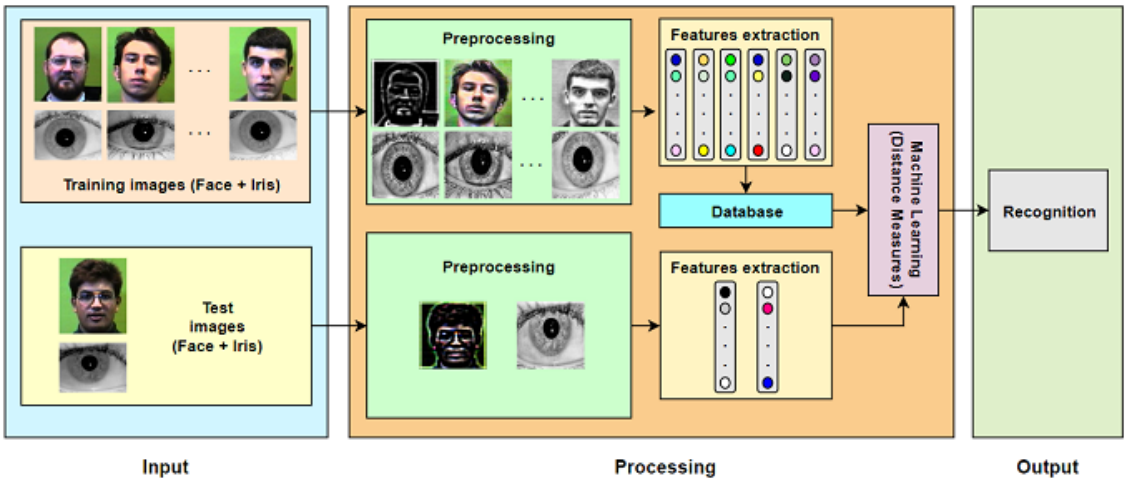


Figure 3.2: Proposed multimodal biometric recognition

Figure 3.2 illustrates the proposed system. A dataset containing images of faces and irises is required to extract features. face and iris images from the same person should be paired. The images first go through a pre-processing step for image normalization, then feature extraction is carried out. A feature vector of each person's facial and iris features is stored in a database. This vector will be used later in the matching step. Matching is a comparison operation using the Euclidean distance to find the desired person if his/her images were stored in the dataset.

To conduct a complete study, unimodal models with multimodal methods are compared. The idea was to confirm the advantages of multimodality over unimodality. For all the experiments (unimodal and multimodal), the different feature extractors with SVD, the different pre-treatment techniques, and the different matrix sizes are tested. Thus, results for the face, iris, and face-iris combination will be provided.

Figure 3.3 summarizes the face and iris recognition process. Suppose the user chooses to do a training on a dataset; he/she first chooses whether or not to do a pre-processing (Gamma Correction or Contrast Enhancement). After that, he/she selects the matrix size, and this value is used to divide the input image into several blocks (2, 4, 8, 16, 32, 64, 128, or 256). Then, experiments using some combinations (DWT+SVD, DCT+SVD, MT+SVD, HT+SVD) are done at the end (see Figure 3.4).

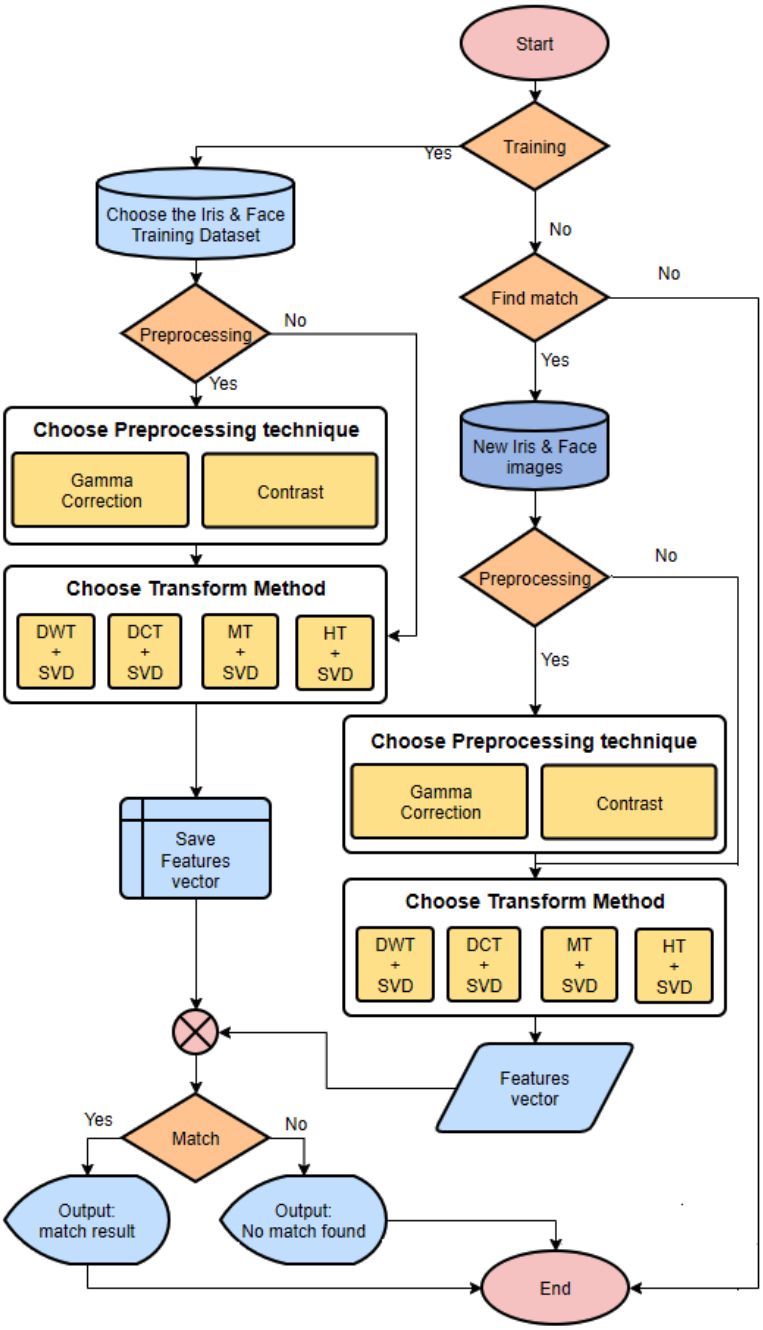


Figure 3.3: Face and Iris recognition flowchart

For the DWT+SVD, DWT extraction is carried out by a two-level wavelet decomposition by the Haar window. The SVD is applied to low-frequency LL sub-bands of the input image and extracted the singular values of Red, Green, and Blue Low-frequency LL sub-bands.

Despite its simplicity, the Haar wavelet was chosen and used because it is sufficient to

obtain the context and texture information of the different frequencies in the image [Wang et al., 2018][Wu et al., 2021].

Once the singular values of each low-frequency LL subband are calculated using the SVD formula, the result will be appended at the end of the corresponding feature vector, and the feature extraction process will be completed. The same dimensionality reduction process is applied to DCT.

For HT and MT extraction, the HT and MT matrices need to be calculated. For the MT, the author's recommended values of a and b (a=1, b=2) are used. The SVD decomposition of MT/HT is calculated by an element-by-element multiplication of the MT/HT matrix with the image block.
This whole process is then applied to the face and iris image.
After this feature extraction step, the feature vectors (of the face and the iris) are stored in a database which will be used later in the matching. Each feature vector (face and iris) is obtained by the concatenation of the S, V, and D components).

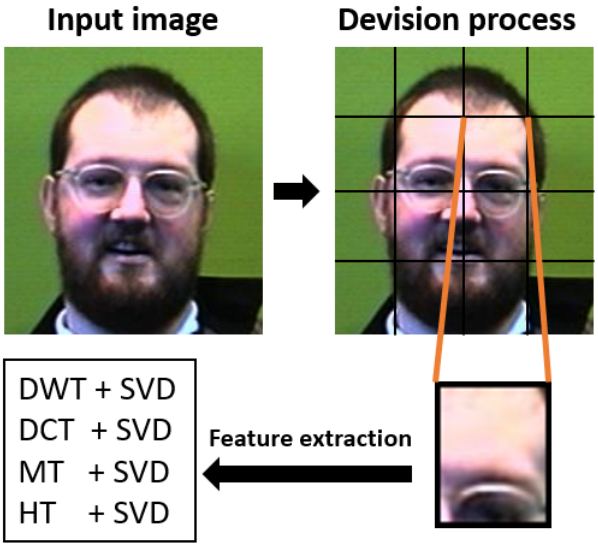


Figure 3.4: Feature extraction process

Face recognition using DWT

Algorithm 1: 2-D Face/Iris Image Detection Algorithm

```
Read the Face/Iris Image from input Device ;
Apply the pre-processing Algorithm ;
Apply three levels DWT on the Face/Iris Image;
Apply the SVD on the three levels DWT ;
Extract the features of Face/Iris Image;
if No Liveness Characteristics found then
    Print "Face/Iris Image does not match with DB Images"
else
    for I= 1 to N do
        //N is the number of DB images;
        Read the Database (DB) image features ;
        Calculate the distance between the two feature vectors;
        if (Decision Threshold Satisfied) then
            Print "Face/Iris Image is matched with DB Image I" ;
        else
            Print "Face/Iris Image does not match with DB Images";
        end
    end
end
```

Algorithm 1 summarizes the Face/Iris recognition process using DWT. In the test phase, the images proposed to the system (facial and iris images) go through the same processing, and the resulting feature vector will be compared to the stored features database to find the best match using a distance measure (Euclidean, Manhattan, or Cosine).

Multiple biometric images obtained from different sensors can be combined at different levels: (i) sensor level: data obtained from different sensors is combined and formed as a new dataset, and further, a feature extraction vector is created. (ii) feature level: at this level, features are extracted from different obtained biometric images and combined to create a new feature vector. (iii) decision level: at this level, different characteristics are extracted, compared individually, and a combined match score is obtained.

Most multimodal biometric systems integrate data at score level due to the good trade-off between the ease of combining the data and better information content. Besides, it is a relatively straightforward way to combine scores generated by different match models. In this thesis, the decision level fusion is used.

3.2.2 Image Encryption/Decryption algorithm

This part of the thesis concerns cybersecurity, in which the database should be secured in the network. To achieve this task, a new algorithm for encryption/decryption of data is proposed. The encryption algorithm (Algorithm 2) allows us to scramble sub-blocks of the image using a new proposed approach (Malakooti-Shebli Scrambling Algorithm) and then transform them using an orthogonal transform (DWT, DCT, HT, or MT). Also, a secret random key using the Malakooti-Raeisi algorithm is generated.

Algorithm 2: Proposed encryption algorithm

Load the original image file from the camera or input device ;

Convert the image file into Bitmap and create three matrices (red, green, blue for RGB Color space) ;

Divide the elements of red, green, and blue matrices into sub-blocks of size 32;

for *I= 1 to N* **do**

    //N is the number of sub-blocks;

    Apply Malakooti-Shebli Scrambling Algorithm on each sub-blocks ;

    Apply one of the Orthogonal Transforms (DWT, DCT, HT, or MT) on scrambled sub-blocks for RGB color space;

    Generate the randomized Secret Keys by using the Malakooti key Gen Algorithm;

    Apply the XOR operation on each transformed sub-blocks;

Combine all scrambled, transformed, XOR-operated sub-blocks for three matrices of red, green, and blue and merge them to obtain an encrypted image file;

Save the encrypted image file into the database;

The decryption algorithm (Algorithm 5) generates a secret random key using the Malakooti-Raeisi algorithm and performs an XOR operation on each image's sub-block. Then, it applies an inverse transform to the result and descrambles the image using the proposed algorithm (Malakooti-Shebli DeScrambling Algorithm).

Algorithm 3: Proposed decryption algorithm

Load the encrypted image file from the database. ;

Convert the encrypted image file into Bitmap and create three matrices (red, green, blue for RGB Color space) ;

Divide the elements of red, green, and blue matrices into sub-blocks of size 32;

for *I= 1 to N* **do**

    //N is the number of sub-blocks;

    Generate the randomized Secret Keys by using the Malakooti key Gen Algorithm ;

    Apply the XOR operation on each encrypted sub-blocks of red, green, end blue matrices;

    Apply the Inverse Operation for one of the Orthogonal Transforms (DWT, DCT, HT, or MT) on encrypted sub-blocks;

    Apply the Malakooti-Shebli DeScrambling Algorithm on each sub-blocks ;

end

Combine all XOR-operated, Descrambled, inverse transformed sub-blocks for the three matrices (red, green, and) blue and merge them to obtain decrypted (original) image file;

Save the decrypted image file into the database;

3.2.3 Malakooti-Shebli scrambling algorithm

Various image scrambling techniques can be used to hide the contents of images efficiently by applying the scrambling algorithm so that illegal users and unauthorized persons cannot recover the scrambled or hidden image content. The rate of data hiding performance depends mainly on the steganography image's visual quality and the scrambling algorithm's hiding capacity. Some researchers have suggested Rubik's cubic rotation techniques for scrambling in which the process of descrambling can

be performed easily [Zhang et al., 2011]. Most researchers use Arnold scrambling algorithm for square images, but most images are not square, and this transform cannot be applied for scrambling of non-square images [Min et al., 2013/11]. Block-based scrambling is another strong algorithm that can be used for scrambling and moving the pixels around so that weekend the strong correlation between adjacent pixels[Xu et al., 2017]. Malakooti and Safari have suggested block-based coding in which images will be divided into sub-blocks of size 32, and a scramble algorithm will be applied to each sub-block. A new robust, block-based scrambling similar to the existing one was developed and tested on several different images and obtained a good rate of image hiding.

This thesis proposes a special block-based algorithm that can be used to scramble image information and an array of vital information such as credit card numbers or employee access IDs to connect to some secure network. The scrambling algorithm is called modified block-based coding or Malakooti-Shebli scrambling because the two algorithms are merged and fitted into the image scrambling algorithm [C. et al., 2007]. The first algorithm used for the design of proposed block-based coding was the Malakooti Scrambling Algorithm [Malakooti et al., 2013] that has been applied to shuffle the digits of credit card for more security before hiding the digits of the credit number inside the target image as follows:

---

for *I=0 to N-1* **do**

    Index=(I+5*7) Mod 16;

    Next I;

The simulation result for N=16 clearly shows that this algorithm provides a unique index number to shuffle the digits of a credit card and scramble the digits.

Table 3.1: Malakooti-Scrambling algorithm for data shuffling

I	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Index	7	12	1	6	11	0	5	10	15	4	9	14	3	8	13	2

The second Algorithm used is the Malakooti-Saffari Scrambling Algorithms (MSSA) [Malakooti et al., 2013]. This method has been used to scramble the pixels of sub-blocks of the image to hide it from unauthorized use as follows :

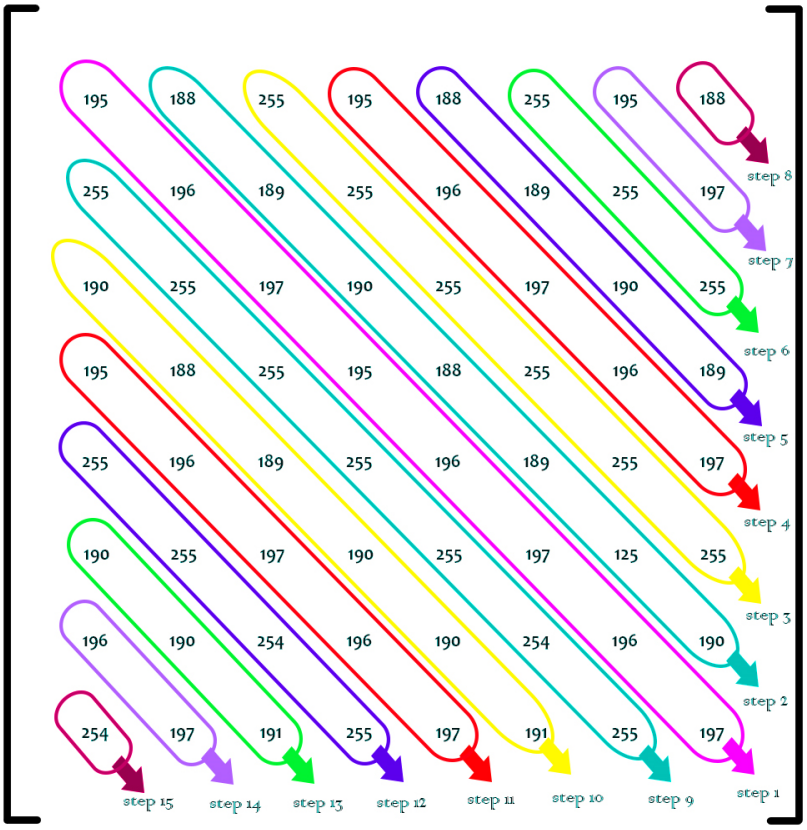


Figure 3.5: Graphical display of the Malakooti-Saffari Image Scrambling algorithm [Malakooti et al., 2013]

They divided the image into sub-blocks of 8*8 and applied MSSA on each sub-block for scrambling. First, each block's main diagonal has been saved into a one-dimensional array, followed by all pairs of upper and lower diagonal elements that have been saved on that array sequentially and respectively. Once main diagonal elements and all pairs of upper and lower diagonal elements are saved on the one-

dimensional array, its contents will be transferred into matrices of the same size as the original red, green, and blues matrices. Finally, these matrices can be combined to create the scramble color image. The result of scrambling and descrambling applied to simulated matrix elements is shown as follows:

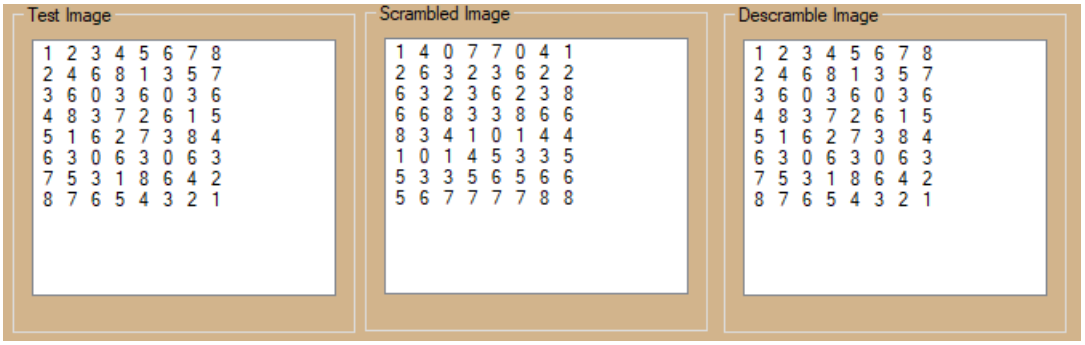


Figure 3.6: Simulated image, Scrambled, and Descrambled results

The Malakooti Scrambling technique is applied to generate 16 randomly unique numbers (7,12,1,6,11,0,5,10,15,4,9,14,3,8,13,2), which can be used to select the corresponding sub-block of red, green, and blues images for scrambling. If the number of sub-blocks is more than 16, the generated random number can rotate by 4 or 8 to generate three groups of randomly selected unique numbers which can be used for sub-block selection [Malakooti and Mansourzadeh, 2014a]. The three rotated random numbers are as follows :

No Rotation: R0= 7,12,1,6,11,0,5,10,15,4,9,14,3,8,13,2

First Index Values: 7,12,1,6,11,0,5,10,15,4,9,14,3,8,13,2

Rotation by 4: R4= 3,8,13,2, 7,12,1,6,11,0,5,10,15,4,9,14

Second Index Values: 19,24,29,18,23,28,17,22,27,16,21,26,31,20,25,30

Rotation by 8: R8= 15,4,9,14,3,8,13,2, 7, 12, 1, 6,11,0,5,10

Third Index Values: 31,20,25,30,19,24,29,18,23,28,17,22,27,16,21,26

Details of the Malakooti-Shebli scrambling algorithm:

Algorithm 4: Malakooti-Shebli scrambling algorithm

Load the image file from the database. ;

Convert the image file into Bitmap and create three matrices (red, green, blue for RGB Color space) ;

Divide the elements of red, green, and blue matrices into sub-blocks of size 32;

Apply Malakooti Scrambling Algorithm to obtain 16 randomly generated unique numbers (7,12,1,6,11,0,5,10,15,4,9,14,3,8,13,2), and use rotation operation by 4, or 8 to get more random number that might be required if numbers of sub-blocks are more than 16 ;

for *I= 1 to N* **do**

    //N is the number of sub-blocks;

    Select the sub-blocks of images corresponding to the indices obtained from randomly generated numbers related to red, green, and blue matrices. Then, apply Malakooti-Saffari Scrambling Algorithm on each sub-blocks for RGB color space;

    Copy the Scrambled sub-blocks inside the scrambled matrices of red, green, and blue in sequential order, i.e., and scrambled sub-block with index 7 will be stored in the first sub-block and index 12 at the second sub-block;

end

Combine all scrambled sub-blocks for three matrices of red, green, and blue and merge them to obtain a scrambled image file ;

Save the scrambled image file into the database;

Details of the Malakooti-Shebli descrambling algorithm:

Algorithm 5: Malakooti-Shebli deScrambling algorithm

Load the scrambled image file from the database. ;

Convert the scrambled image file into Bitmap and create three matrices (red, green, blue for RGB Color space) ;

Divide the elements of red, green, and blue matrices into sub-blocks of size 32;

Apply Malakooti Scrambling Algorithm to obtain 16 randomly generated unique numbers (7,12,1,6,11,0,5,10,15,4,9,14,3,8,13,2), and use rotation operation by 4, or 8 to get more random number that might be required if numbers of sub-blocks are more than 16 ;

for *I= 1 to N* **do**

    //N is the number of sub-blocks;

    Select the sub-blocks of scrambled images sequentially, one by one, from red, green, and blue matrices. Then, apply Malakooti-Saffari DeScrambling Algorithm on each sub-blocks for RGB color space;

    Copy the Descrambled sub-blocks one by one from the Descrambled matrices of red, green, and blue in sequential order. Then, save it at the corresponding sub-block of the descrambling matrix with Index position related to randomly generated numbers, i.e., the first sub-block will be copied at Index location 7, and the second at index 12;

end

Combine all descrambled sub-blocks for three matrices of red, green, and blue and merge them to obtain a descrambled image file ;

Save the descrambled image file into the database;

3.2.4 Secret key generation

Modern cryptographic systems include symmetric key algorithms (such as DES and AES) and public key algorithms (such as RSA). Symmetric key algorithms use a common shared key for encryption and decryption processes. This key needs to be kept as secret data is required for both senders to send the receiver of encrypted information. The asymmetric key algorithm or Public key algorithm uses a public key and a private key. The public key is available to everyone (often using a digital certificate). The sender encrypts the data with the recipient's public key. Only the private key holder can decrypt this data. Because public-key algorithms tend to be slower than symmetric key algorithms, modern systems such as TLS and SSH use a combination of the two: one side receives the other public key and one small piece of data (Or asymmetric key). Or some data is used to generate it, and the rest of the conversation uses the asymmetric (usually faster) algorithm for encryption.

Computer encryption uses integers for keys. Sometimes, the keys are randomly generated using the RNG or quasi-random number generator (PRNG). PRNG is a computer algorithm that generates data that appears under random analysis. PRNGs that use system entropy to seed data usually produce better results because it makes PRNG's initial conditions much harder to guess for attackers. Another way to generate random is to use information outside the system. Random Number Key Generator algorithms such Malakooti-Raeisi Key Gen Algorithm (MR-KeyGen) [89,107] were used to obtain a sequence of secret keys required for encryption and decryption processes. The MR-Key gen Algorithm is presented in the following steps:

1 Enter two large prime numbers P and Q with a constant number for M, i.e. 3.

2 Multiply two large prime numbers together PQ=P*Q.

3 Calculate $A(I) = P * Q \bmod 4096$.

4 Calculate $B(I) = [P * Q/4096]$.

5 Increase $B(I)$; $B(I) = B(I) + I$.

6 Generate new P and Q as following

- $P = [(A(I) + 1) * M]$
- $Q = B(I) + I$

7 Save $A(I)$ as the key Value, $Key(I) = A(I)$

8 Return to step 2 to generate next key value

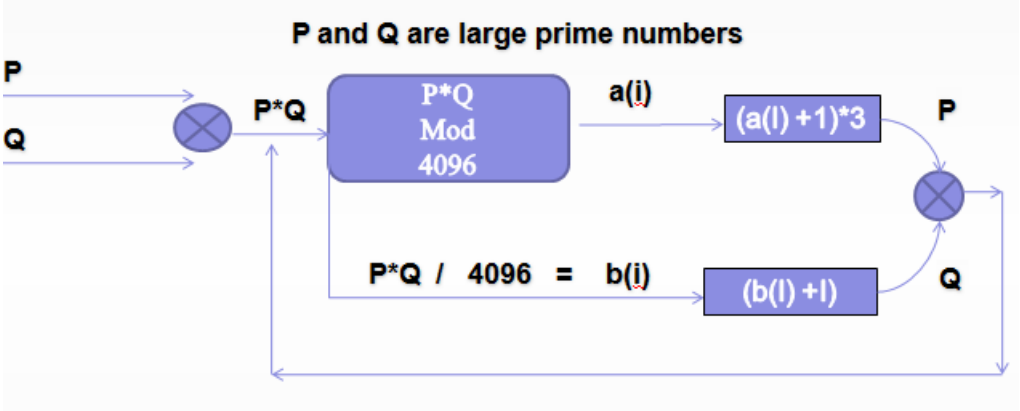


Figure 3.7: Malakooti–Raeisi key Gen algorithm block diagram

The output of the MR-Key Gen Algorithm is shown in Figure 5.6. The $A(i)$ array elements are the pseudorandom numbers used for the third phase of the encryption process called the XOR operation. The key Gen random number will be used for XOR operation, and encryption will be completed in this stage. Similarly, the MR key Gen can be used to generate the same random numbers generated for encryption to be used for the decryption process, which is the first decryption stage.

3.3 Proposed face and iris multimodal biometric recognition solution using deep learning

This section proposes an efficient multimodal biometric system based on Deep Learning. Ten approaches based on Convolutional Neural Networks (CNN) using pre-training models downloaded from the literature are explored.
The VGG16, Resnet50, DensenNet121, MobileNet, and InceptionV3 architectures are used for iris recognition. The VGGFace, FaceNet, InceptionV3, Resnet50, and OpenFace architecture are used for the face. These models were trained on the ImageNet dataset [Russakovsky et al., 2015].

The pipeline of our experiments consists of pre-processing the face images by selecting the part of the image that contains only the face. Then, the second step consists of retraining the ten models on our datasets using fine-tuning techniques. Each model dedicated to the iris will allow us to obtain a feature vector which will be combined with a vector obtained by a model dedicated to the face. Thus, we will have 25 possibilities of combination between the iris and face models. The combinations between the iris and the face will be done at two different levels: at the feature level or the decision level. At the feature level, the two feature vectors (iris and face) will be concatenated and compared to the database to identify or authenticate the person. At the decision level, the verification is done for each modality separately. If the two feature vectors recognize the same person, the decision is validated; otherwise, the person is not recognized.

3.3.1 Used models for iris recognition

In the previous section, all the tools to understand the architecture of a convolutional neural network were introduced. There are several of them in the literature whose

effectiveness varies according to the task because they do not all have the same number of convolutions (nor the same structure).

This section will present the different deep architectures used for the proposed iris recognition system.

3.3.1.1 VGG

The recent arrival of large annotated image databases such as CIFAR-10 and CIFAR-100 [Krizhevsky, 2009], then ImageNet [Liu et al., 2017][Russakovsky et al., 2015], has led to the success of deep convolutional networks in digit classification on the MNIST database [Lecun et al., 1998], in traffic sign recognition [Stallkamp et al., 2011], in Chinese character identification [Liu et al., 2011], and of course in object recognition on ImageNet [Krizhevsky et al., 2012].

The LeNet-5 architecture, developed by LeCun et al. [Lecun et al., 1998], defines the reference structure of a CNN. It consists of two convolutional layers followed by three fully connected layers. The convolutional part of the model performs feature extraction in the image, while the fully connected layers complete the final classification.

The AlexNet [Krizhevsky et al., 2012] network uses a similar approach for color image processing. This model consists of 8 layers, 5 convolutional, and 3 fully connected; it processes red-green-blue (RGB) images of dimensions 224 × 224. The first convolution uses a large 11 × 11 kernel and precedes a subsampling, significantly reducing the image size. Thus, the output activation maps of the first layer are 96×55×55 and 256×27×27 after the second. The convolutional structure of AlexNet is designed to increase the number of activation maps in inverse proportion to the reduction in spatial dimensions. This model allowed Krizhevsky, Sutskever, and Hinton [Krizhevsky

et al., 2012] to win the ILSVRC competition [Russakovsky et al., 2015] in 2012 with an error rate of 15.3%.

The VGG-16 model refines the AlexNet architecture by proposing reducing the convolution kernels' size. Indeed, Chatfield et al. [Chatfield et al., 2014] and Simonyan and Zisserman [Simonyan and Zisserman, 2015] suggest that it is simpler to optimize several successive convolutions of 3 × 3 kernels than a single convolution of dimension 11 × 11. Moreover, the presence of additional nonlinearities is likely to increase the expressivity of the model. The VGG-16 model, therefore, replaces each classical wide convolution with a block of 2 or 3 successive 3 × 3 convolutions, as shown in Figure 3.8. The VGG-16 reference model consists of 16 layers, 13 of which are convolutional and 3 of which are fully connected, following the canonical approach of LeNet and AlexNet. It consists of five 3×3 convolutional blocks, each followed by a sub-sampling of stride 2. The first two blocks have 2 convolution layers, and the next three have 3. The final activation maps are 512×7×7, VGG-16 achieving a dimension reduction by a factor of 32. This vector of 25088 in length is then reduced to 4096 and then to the 1000 classes of ImageNet. To avoid overfitting, the fully connected layers are subjected to dropout. These proposed enhancements to the classical CNN architecture have made it possible to obtain, in 2014, a 7.4% error rate in object recognition during the ILSVRC competition.

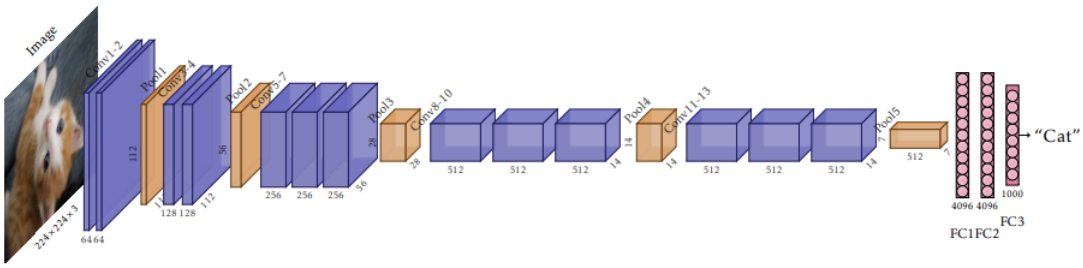


Figure 3.8: VGG-16 architecture [Simonyan and Zisserman, 2015]

3.3.1.2 Inception

Szegedy et al. [Szegedy et al., 2015] propose the GoogLeNet model with 22 layers. In particular, this architecture introduces the Inception module stacking several layers in parallel, not only in-depth. The idea is to realize, for an activation map, the extraction of features at several context levels using either a 1×1 convolution, i.e., a linear combination followed by non-linearity, or a pooling, or 3×3 or 5×5 convolutions. This allows coupling characteristics with invariance to local translations (from subsampling) and characteristics without invariance, allowing a greater variety of cases to be handled. The Inception module is shown in Figure 3.9, while Figure 3.10 details the complete architecture of the GoogLeNet model. Given the depth of the network (22 layers), its authors propose to facilitate the optimization of the lowest layers by adding a classifier at the level of the intermediate representations after the Inception modules (4a) and (4d). This deeply supervised approach had already shown its effectiveness in combating vanishing gradient problems [Lee et al., 2015]. This model obtains an error rate of only 6.4% in object recognition at the ILSVRC. The GoogleNet architecture was subsequently improved [Szegedy et al., 2016] by replacing the 5×5 convolutions of the Inception module with two 3×3 convolutions, as proposed in the VGG model [Simonyan and Zisserman, 2015], and by integrating the Batch Normalization [Ioffe and Szegedy, 2015].

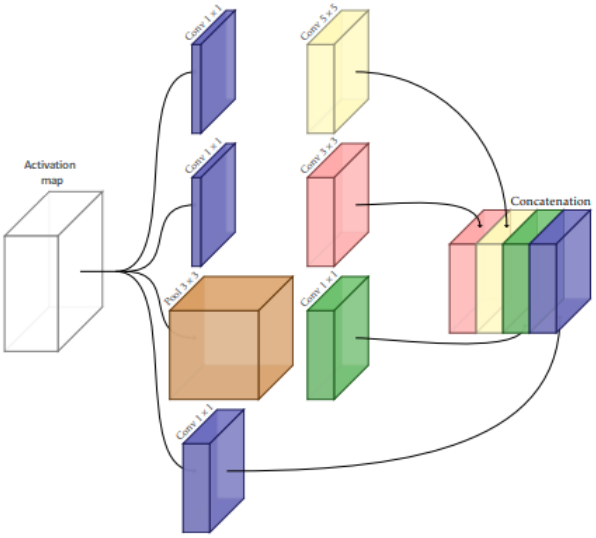


Figure 3.9: Inception module

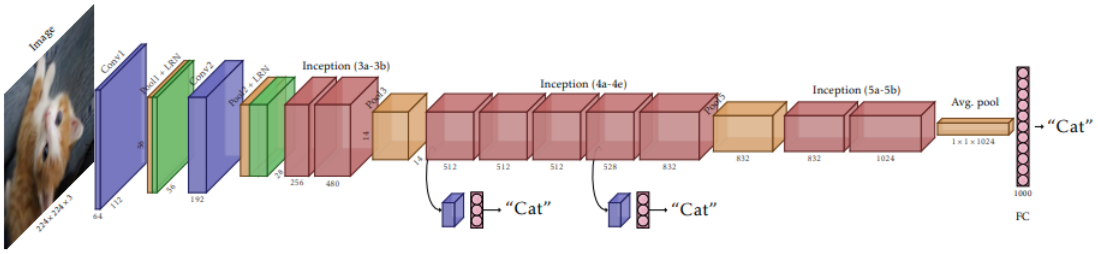


Figure 3.10: Inception architecture

3.3.1.3 Resnet

In 2015, He et al. [He et al., 2016] obtained an error rate of only 3.5% in object recognition during ILSVRC. Their approach consists of a deep network comprising more than 100 convolutional layers. Optimization is made possible on the one hand by Batch Normalization, but mainly by residual learning. The idea is to break the purely sequential structure of forward-propagating networks by adding connections to short-circuit the next layer. These connections, called residuals, correspond to a simple identity operation and allow the activation and the gradient to run through the entire network without suffering evanescence or explosion due to the chain shunt rule. Rather than trying to approach $f : x \to f(x)$, the residual block will approach

$\hat{f} : x \to f(x) - x$, which is simpler because it has a priori a lower amplitude.

The residual convolution block is shown in Figure 3.11, and an example of a 34-layer ResNet model is detailed in Figure 3.12. The introduction of residual learning partly changes the paradigm used until now for the design of CNNs. The basic building block of the network is thus shifted to the residual block. Resnet has many layers but comparatively few parameters because only the last layer is fully connected. The fully connected layers generally concentrate most of the CNN weights. They are also the most sensitive to overfitting, requiring integrating regularizations such as Dropout [Srivastava et al., 2014]. ResNet contains almost only 3×3 convolutions, except for the first 7×7 convolutions, which allows a drastic reduction in the image's spatial dimensions. Interestingly, the final dimension reduction before the fully connected layer is made using adaptive subsampling. Thus, regardless of the input image's size, the subsampling will average the activations to produce the vector of features of the size expected by the fully connected layer. However, it should be noted that the high number of activations and intermediate gradients to be computed makes ResNet expensive in memory and impractical on large images. The Inception architecture will also be enhanced by residual connections [Szegedy et al., 2017].

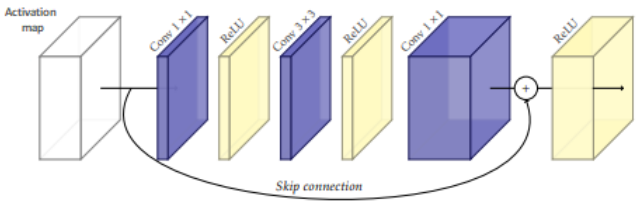


Figure 3.11: Residual convolution block

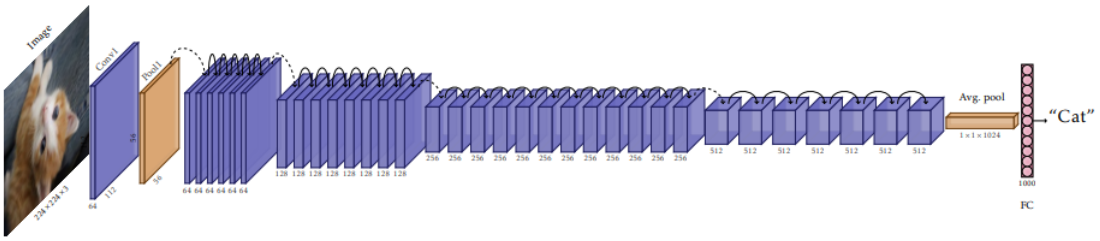


Figure 3.12: 34-layer ResNet model

3.3.1.4 DenseNet

The use of intermediate activation maps and their propagation to the upper layers allows a better classification by considering several levels of abstraction. Also, several works suggest that these approaches allow in practice to combine several models into one, with activations being able to follow several network topology paths [Veit et al., 2016][Huang et al., 2016]. However, the residual connections only provide access to the activations of the previous layer. Therefore, Huang et al. [Huang et al., 2017] proposed a so-called DenseNet architecture with dense connections, building a model in which all activation maps from the lower layers are transmitted to all the upper layers. To avoid the explosion of the number of parameters and activations, the model is divided into several dense blocks, as shown in Figure 3.14. Each block is detailed, as shown in Figure 3.13. The dense connections' presence allows the gradient to propagate immediately from the upper to the lower layers, thus applying an implicit form of deep supervision [Lee et al., 2015]. In addition, a convolutive transition layer is applied between two blocks to reduce the number of planes and is followed by max-pooling to reduce spatial dimensions. This architecture obtains comparatively better results than the ResNet based on validating the ILSVRC 2012. But, like ResNet, if the number of parameters of DenseNet architectures is low, these modifications are costly in the memory space required to store activations and intermediate gradients.

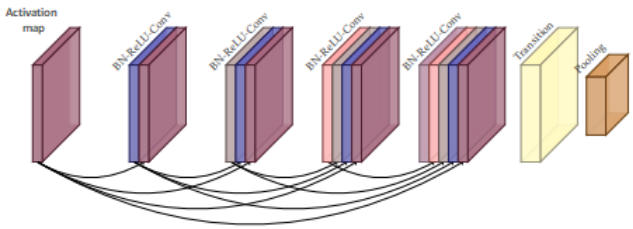


Figure 3.13: Dense convolution block

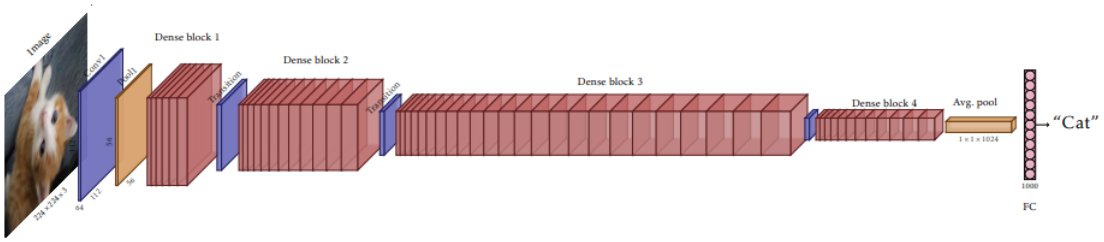


Figure 3.14: DenseNet model

3.3.1.5 MobileNet

Chollet [Chollet, 2017] introduces depth-wise separable convolutions. These operate a filter by a plane of the activation tensor and are recombined pixel by pixel by a kernel point-to-point 1 × 1 convolution. Thus, this is a specific case of the usual convolution in which each tensor plane is filtered by one and only one convolution kernel, as shown in Figure 3.15. These convolutions are introduced to replace the Inception module in the eponymous architecture and have improved its performance on ImageNet and JFT datasets internal to Google. A notable advantage of these convolutions is that they require fewer parameters than classical convolutions. Indeed, a k1 × k2 convolution operating on Nin activation cards producing Nout activation cards requires k1 × k2 × Nin × Nout parameters. The convolution separable in depth requires Nin ×Nin ×k1 ×k2 parameters for the first phase and then Nin ×Nout for the second, i.e., a total of Nin × (Nin - k1 - k2 + Nout), which is advantageous when Nout Nin, which is the most common case. These separable convolutions are thus effective at Using the new software is also very popular for real-time embedded applications [Huang et al., 2016]. MobileNet is a 28-layer architecture built on depth-wise separable convolutions, except for the first layer, which is a full convolutional layer. All layers are followed by batch normalization and ReLU non-linearity. The final layer is fully connected and feeds the softmax for classification.

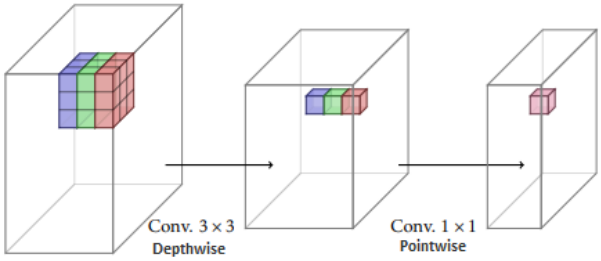


Figure 3.15: Depth-separable convolutions

3.3.2 Used models for face recognition

For face recognition, two of the previously mentioned architectures (Resnet and Inception) are used, and three other architectures are adapted explicitly for this task (VGGFace, FaceNet, and OpenFace).

3.3.2.1 VGGFace

Members of the Visual Geometry Group (VGG) at Oxford University have proposed a dataset for recognizing people using their faces. They proposed a face recognition model, called VGGFace, based on the VGG architecture with a different output layer (the last layer contains 2,622 units instead of 1000). [Parkhi et al., 2015a]
The proposed model starts by classifying faces as persons using the Softmax activation function in the output layer. Then, it recovers the vector representation of the face (also called face embedding) by removing this output layer. The last step is to improve the model by fine-tuning the triplet-loss function to reduce the Euclidean distance between the vectors generated for the same identity and increase it with the vectors generated for different identities.
Triplet-loss training aims at learning score vectors that perform well in identity verification by comparing face descriptors in Euclidean space. It is used to learn a projection that is at the same time distinctive and compact, achieving dimensionality reduction at the same time. The formula for calculating the empirical triplet-loss

of the projection W' is as follows:

$$E(W') = \sum_{(a,p,n)\in T} max\{0, \alpha \left\|x_a - x_n\right\|_2^2 + \left\|x_a - x_p\right\|_2^2\}, \quad x_i = W' \frac{\phi(l_i)}{\left\|\phi(l_i)\right\|_2^2} \tag{3.1}$$

With:

- $\alpha \geq 0$ is a scalar that represents the learning margin,
- T is a collection of training triplets,
- x_t is a a score vector associated by the deep architectures ϕ to each training image $l_t,\ t = 1, ..., T$

A triplet (a, p, n) contains an anchor face image a as well as a positive $p \neq a$ and negative n examples of the anchor's identity. [Parkhi et al., 2015a]

3.3.2.2 FaceNet

The FaceNet model was developed by researchers at Google [Schroff et al., 2015]. As shown in Figure 3.16, This model consists of a batch input layer and a deep CNN followed by L2 normalization, which results in face embedding. The authors used the Inception ResNet v1 backbone and triplet-loss during training. This model allows the direct extraction of vectors of 128 elements (rather than extracting them from an intermediate layer of a model), representing the embedding of the input images and then being used as the basis for the classification training systems. FaceNet is built on the idea of inception, using a 1×1 convolution and pooling layers in parallel to remove possible redundant parameters and make the recognition model lighter.

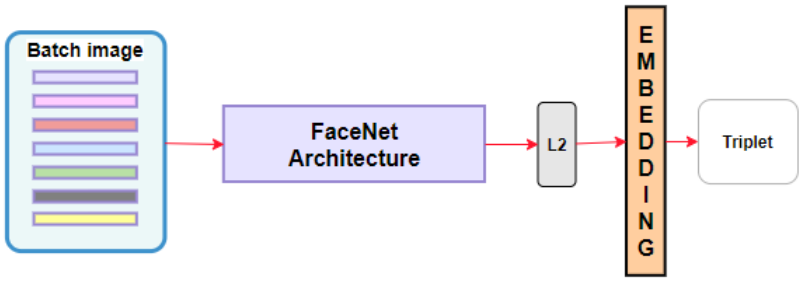


Figure 3.16: Block diagram of FaceNet architecture

3.3.2.3 OpenFace

OpenFace [Amos et al., 2016] provides low-dimensional face representations for the faces in an image which makes it well-suited for mobile scenarios compared to other techniques.

This model is based on FaceNet's architecture used as feature extractor and dlib's [King, 2009] pre-trained model as face detector (see Figure 3.17).

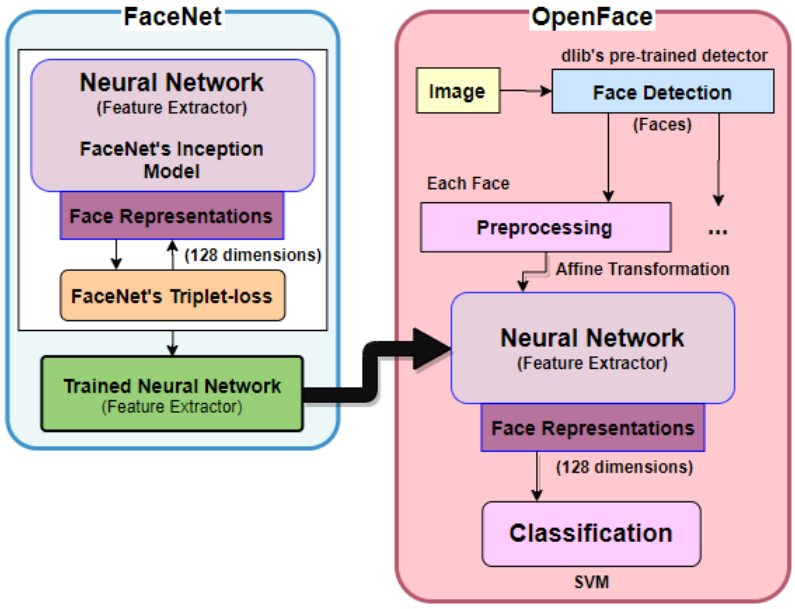


Figure 3.17: Block diagram of OpenFace architecture

In the face detection step, a list of bounding boxes around the faces in an image

is returned. This new face image is normalized so that the eyes, nose, and mouth appear at similar locations in each image.

Figure 3.18 illustrates how the affine transformation normalizes faces. The 68 landmarks are detected with dlib's face landmark detector [King, 2009]. The affine transformation also resizes and crops the image, so the input image to the neural network is 96 × 96 pixels.

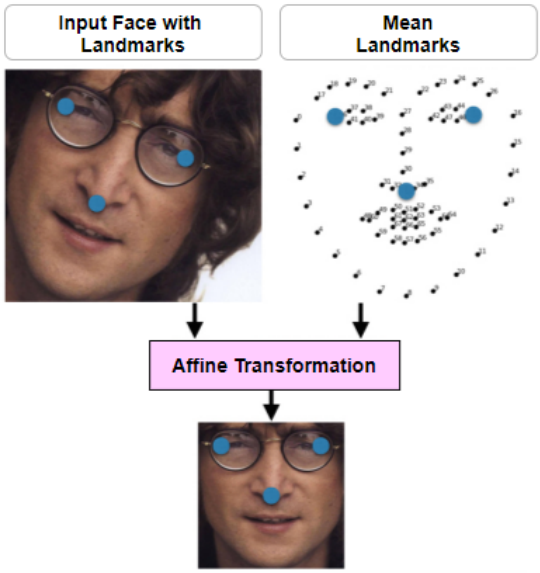


Figure 3.18: OpenFace's affine transformation

3.3.3 Proposed multimodal system

The second part of this thesis's objective is to secure access to a restricted area using a system based on deep learning models. It has been chosen to focus on an image-based biometric security system that meets the various standards, constraints, and security experts' recommendations. A multi-biometric system, using, on the one hand, the face and, on the other hand, the iris, is proposed.

To achieve this task, several steps are necessary:

- Setting up a multimodal database for learning and testing the proposed approaches.

- The training of a robust model based on a deep architecture for iris recognition.
- The training of a robust model based on a deep architecture for face recognition.
- The fusion of the two models for multimodal recognition.
- The comparison of the results with state-of-the-art.

Figure 3.19 shows the flowchart of the proposed method.

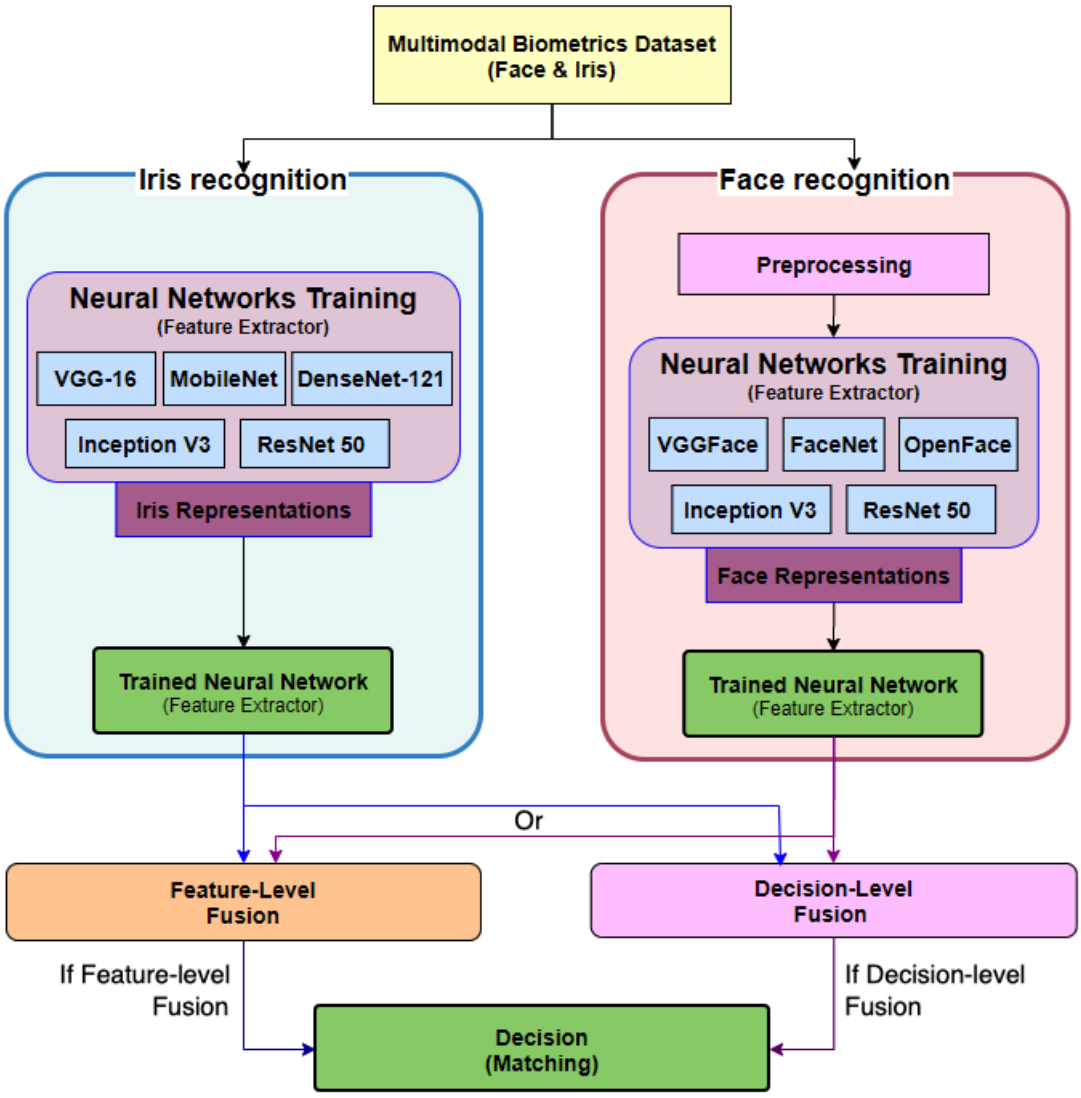


Figure 3.19: Flowchart of the proposed models

3.3.3.1 Face recognition

Face recognition involves preparing the database by pre-processing the images and training the different architectures.

Preprocessing Before considering any recognition or authentication of faces, it is necessary to detect them. Increasing the performance of classification algorithms involves maximizing the amount of face-specific data related to the amount of information in the processed image, thus minimizing the influence of data external to the face (background).

The problem of face detection has been studied in the literature since the first contributions of computer-assisted vision. Research in the field of automatic face detection is indeed motivated by the various emerging applications closely linked to these needs, such as face recognition or authentication [Zhao et al., 2003], face tracking for surveillance [Kalal et al., 2010], facial expression recognition [Kumar et al., 2009], gender and age recognition [Fu et al., 2010] or automatic retouching of facial photos [Wang et al., 2009], etc.

For face recognition or authentication, some researchers have proposed techniques to detect the faces based on deep learning [Shao et al., 2020] or complex 3D alignment [Taigman et al., 2014]. Others have used simpler methods, such as FaceNet, which uses a minimal alignment with tight cropping around the face area. OpenFace uses the dlib's technique that allows moving the corners of the eyes and the nose to medium locations and resizing and cropping the image.

For the proposed approach, the technique used by the OpenFace network is employed [King, 2009].

Deep learning models training After the pre-processing stage, five different networks: VGGFace, FaceNet, OpenFace, Inception V3, and ResNet50, are trained. These models are trained on the VGGFace database using a machine with an i7

4820k processor, 64 GB of Ram, and double GTX1070ti 8GB GPU.

The Total fine-tuning approach for transfer learning is adopted. All layers of each network are re-trained on the pre-processed images. As the weights are initialized with the pre-trained network's values and refined, this new learning is faster.

The five models obtained are used as feature extractors (see Figure 3.20). These models will extract feature vectors from each training image and store them in a database. This database will be used later to decide whether a test image belongs to authorized individuals or not.

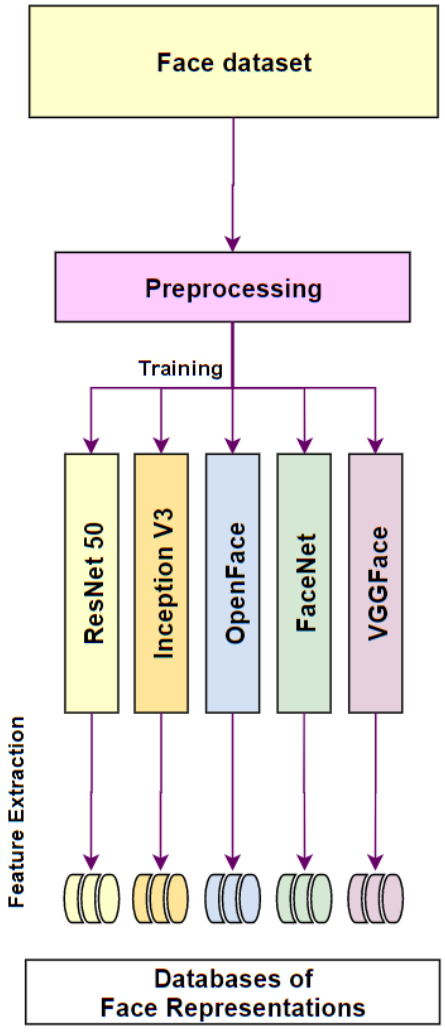


Figure 3.20: Face features extraction process

In the literature, for face recognition or authentication, CNNs are used either as a feature detector and classifier or as the only feature extractor (in the same way as DWT). The classification phase (matching) is performed by an additional method

(SVM, Euclidean distance, Bayesian model ...). This mode of use aims to create a model that optimizes the training data distribution, separating the classes as well as possible and minimizing the interclass variance. Thus, using a very simple classifier, the features of new face images are generated by the network, and a fast classification can be performed.

3.3.3.2 Iris recognition

Iris recognition goes through the same steps as face recognition; the differences concern the architectures used and the fact that there is no pre-processing step (see Figure 3.21).

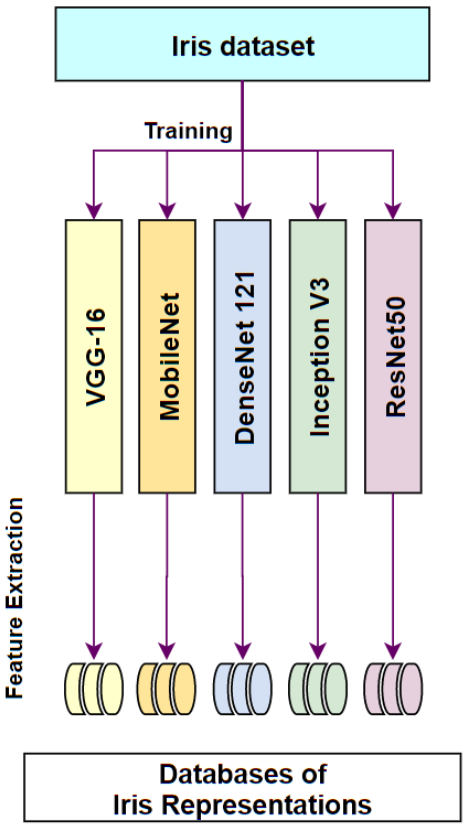


Figure 3.21: Iris features extraction process

3.3.3.3 Face and iris fusion techniques

Two fusion techniques, decision-level and feature-level fusion are tested to choose the best combination technique, which ensures a good recognition rate and good security.

Feature-level fusion This method uses the representations (feature vectors) extracted by the deep architectures. These vectors are concatenated to make a single vector used to find the match of the test image (see Figure 3.22).

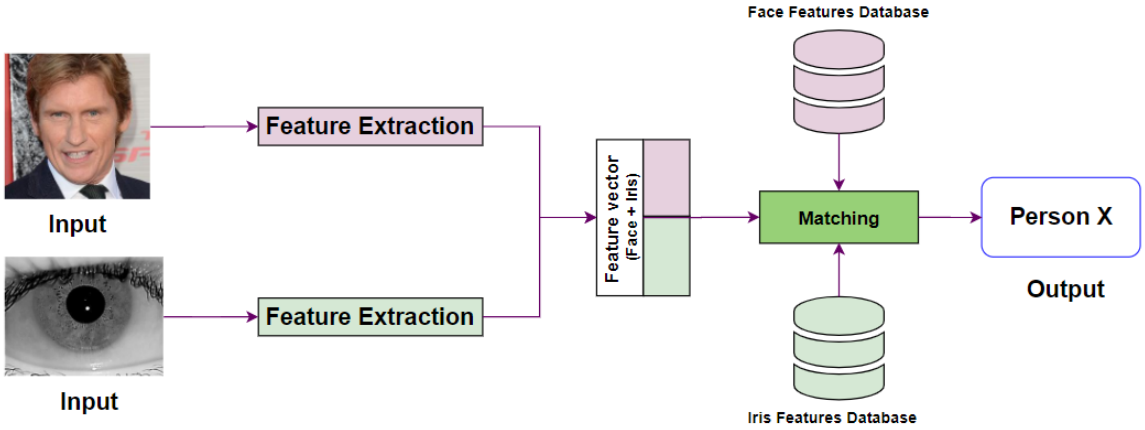


Figure 3.22: Feature-level fusion process

Decision-level fusion Using this technique, each recognition model (face or iris) makes a decision independently of the other. At the end of this step, a fusion of these decisions is performed to make the final decision. If both models have recognized the different features and belong to the same person, the access will be authorized; otherwise, the access will be rejected (see Figure 3.23).

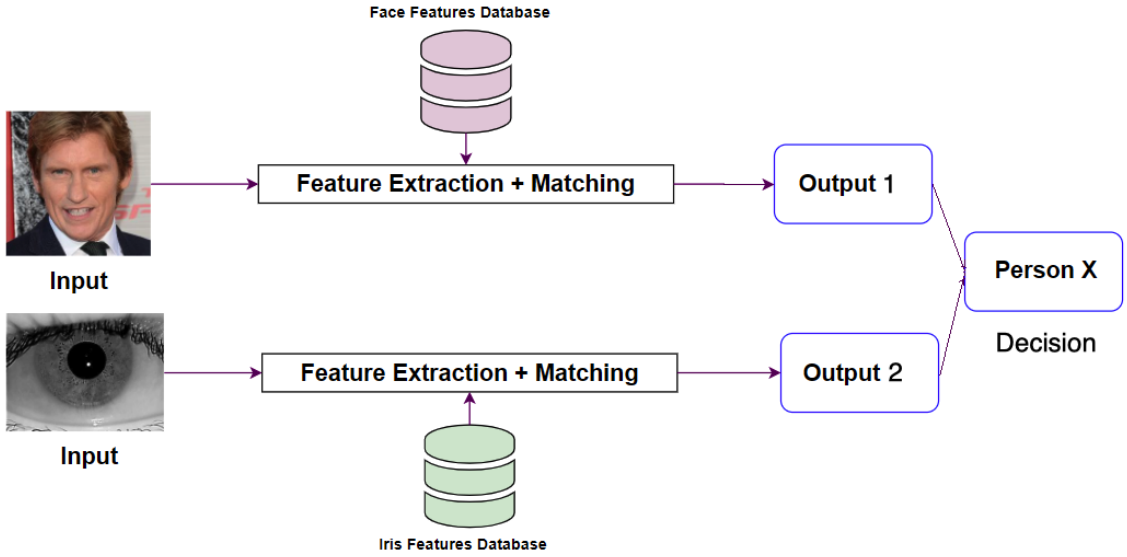


Figure 3.23: Decision-level fusion process

3.3.3.4 Multimodal decision (matching)

Feature extraction, therefore, provides a vector of elements representing these features. The next step in the chain is classification (matching). Its goal is to compute the degree of similarity between two vectors (target feature and measured feature) or between one vector (measured feature) and a set of vectors (forming a class). This comparison can be done in different ways, more or less efficient depending on the data's complexity (dimensions of the vectors, interclass variance, separation of the classes, etc.).

Similarity Measure In general, the most trivial way to compare two vectors of the same characteristic (and thus of the same dimension) is to check their degree of similarity. In statistics, this similarity is expressed as the distance separating these two vectors in their space. In the same way as measuring the distance between two points (norm), calculating the distance between two vectors measures the distance between each element "i" of these vectors.

Let there be two vectors "x" (target features) and "y" (measured features) of "n" elements, belonging to a normed vector space E, such that $\vec{x}(x_1, ..., x_i, ..., x_n)$ and $\vec{y}(y_1, ..., y_i, ..., y_n)$ are both elements of R^n. The distance between these vectors, denoted by $d(\vec{x}, \vec{y})$, is a measure between each of their two-by-two "i" components $(1 < i < n)$ and can be defined in several ways; the most used in state of the art is the Euclidean distance (2-distance or L2) that will be used in the approaches.

The measured distance gives information about the relationship between the recorded feature vector and the measured feature vector. Therefore, the correspondence between the two vectors is defined according to a threshold on this distance value.

Proposed matching method The decision threshold directly impacts the model's performance, making it a key factor in distance-based matching.
To improve the system's performance (correct recognition rate) and minimize the false acceptance rate (FAR) and false rejection rate (FRR), the threshold must be set correctly. Most of the methods proposed in state-of-the-art focus on improving the feature extraction approach and neglect the setting and improvement of the decision threshold.

A new method for automatic matching without setting up a threshold is proposed here (see Figure 3.24).

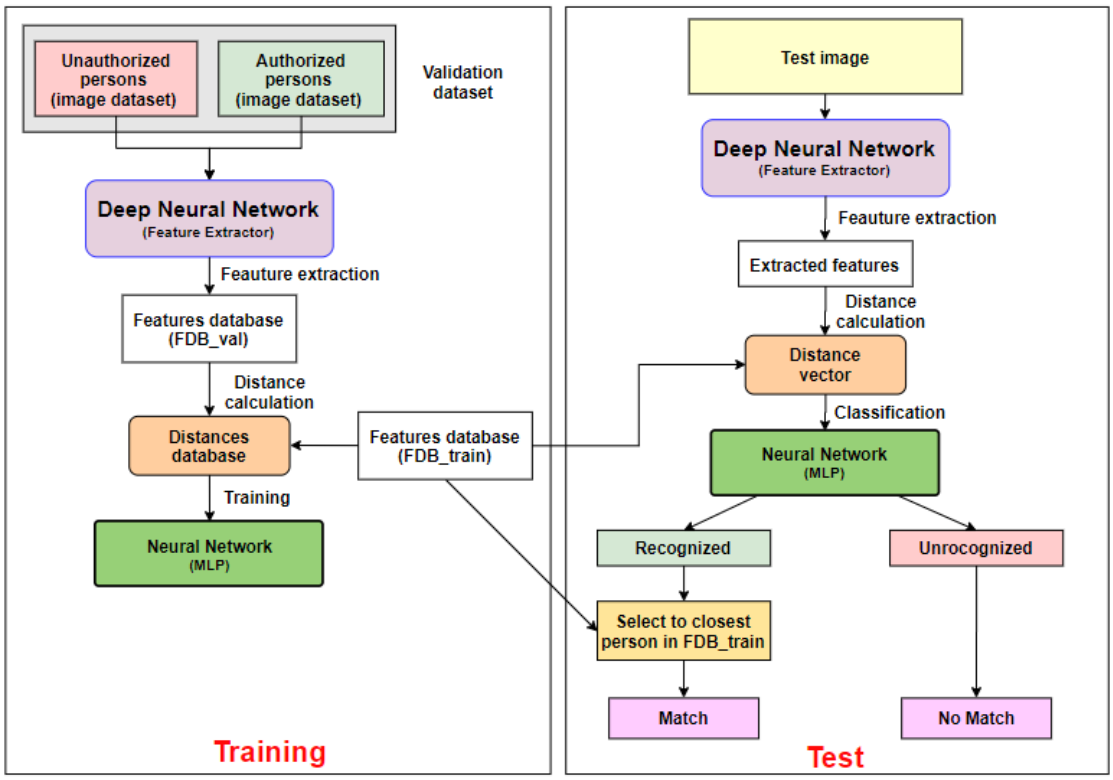


Figure 3.24: Automatic matching process

The proposed method could be used in a unimodal or multimodal system. Its objective is to ensure the best compromise between FAR and FRR.

This technique consists of a neural network model that allows people's automatic classification into two classes: recognized or unrecognized. If the person is authorized, it will be authenticated with the person's identity closest to it in the database of training features (FDB_train). Otherwise, the person is not recognized and will not be authenticated.

To train the proposed model, feature extraction on the validation dataset and on another set containing unauthorized persons is performed using the already trained convolution network. This database is named Feature Database Validation (FDB_val).

After the feature extraction step, the distances between each feature vector of the FDB_val feature database and each vector of the FDB_train database are calculated. The results will be stored in a new database (distance database) which will be used to train a multilayer neural network (MLP). Each row of this database will contain the minimum distance (D_{min}), the average distance (D_{mean}), and the maximum distance (D_{max}) between each feature vector of the FDB_val database and the whole FDB_train database. The last column represents the class: 0 if the person is not authorized, and one if the person is authorized.

Let F be the feature extraction function using a deep learning model. It will have as input an image I and returns a vector that represents it.
Let SM denote a similarity measurement operation using Euclidean distance. Then, for a given image I, the resulting distance between I and the training features database FDB_train can be expressed as a distance-vector D_I:

$$D_I = SM(F(I), FDB\_train) \tag{3.2}$$

Each image can be described by its minimum, average and maximum distances from the FDB_train:

$$image_i = (D_{Imin}, D_{Imean}, D_{Imax}) \tag{3.3}$$

Where:

$$D_{Imin} = min(D_I) \tag{3.4}$$

$$D_{Imean} = mean(D_I) \tag{3.5}$$

$$D_{Imax} = max(D_I) \tag{3.6}$$

The idea is to perform an automatic classification based on the distance. Indeed, an unauthorized person will necessarily be far from the authorized persons. Thus, the system will be able to learn from the data to distinguish between these two classes.

During the test phase, the new image follows the same process: feature extraction, distance calculation with the learning feature base (FDB_train), and classification with the neural network (MLP). Once the decision is made, if this person is considered authorized, the match will be found by taking the closest person on the FDB_train database.

To analyze the distance space and understand the impact of the choice of threshold (θ) on the model's performance, the FAR and FRR are plotted as a function of an alpha (α) parameter, which will be an operating factor of the decision space.
The threshold based on D_{max} is studied. The alpha value is varied from 0 to 100 with a step of 0.05, which will give us 2000 values for each error (FAR and FRR).

$Threshold\ (\theta) = \alpha(max(D\_max))$, max of all D_max values

Such an analysis allows us to study the impact of the choice of θ on the model's performance and especially on the FAR and FRR values. Indeed, the threshold choice depends on the security model's characteristics.

3.4 Summary

The first part of this chapter has given an overview of the approaches used to design the proposed model, namely DWT, DCT, HT, MT, and SVD. This thesis treats the problem of multimodal recognition using a unique face and iris recognition method. This thesis's contribution consists of the proposal of a face and iris image processing pipeline for multimodal recognition. The images are preprocessed using the contrast enhancement; then, each image is divided into 128 blocks, each of which will be characterized using DWT.

The second part of this thesis introduces biometric recognition using deep learning. Here, a deep architecture provides feature extraction: VGG-16, MobileNet, DenseNet-121, Inception V3, or ResNet 50 for the Iris and VGGFace, FaceNet, OpenFace, Inception V3, or ResNet 50 for the Iris. In the end, an automatic classification based on the distance is performed. This step allows to give or not access to an individual. These different proposed ML/DL approaches will be evaluated in Chapter5.

Chapter 4

Design Testing, Validation and Verification of the Developed Cyber Security Biometric Platform Solution using Machine and Deep Learning Classifiers

4.1 Introduction

This thesis aims to answer the problem of person recognition in a secure environment.

The preliminary study on biometrics, presented in chapter 2, led us to consider a multimodal system based on facial and iris features and add, consequently, to the security requirements, a part of encryption/decryption of data.

Whatever the modality used in a biometric system, it is necessary to go through an

enrollment phase (or training of the biometric model). This consists of acquiring a sample, extracting the characteristics, and then training a classifier to recognize them.

For a biometric recognition system, the extracted characteristics are stored in a database to be compared to a new sample when an unknown person passes by.

In this chapter of the thesis, the design, used tools to develop the platform and models, and the obtained results are presented. The discussion will focus on comparing the proposed method based on machine learning and other machine learning techniques. It will also focus on the proposed deep learning approach compared to other state-of-the-art deep learning techniques.
The advantages and disadvantages of deep learning compared with machine learning, the implications of the results, and the objectives achieved in this thesis will also be discussed.
At the beginning of this thesis, it was intended to develop a platform using machine learning for multimodal recognition, hence the choice of the Matlab programming language, which provides a large panel of toolboxes for machine learning. During the thesis, it was found that Deep Learning offers more possibilities and power in image processing. It should be noted that Matlab is not recommended for developing Deep Learning models; the most commonly used language is Python hence the adoption of this language in the second part of this thesis.

4.2 Cyber security biometric platform solution using machine learning classifiers

To develop the proposed cyber security biometric platform solution using machine learning classifiers, MATLAB 9.0 (MATrix LABoratory) was used.

The graphical interfaces are designed using the GUIDE (Graphical User Interface Development Environment) of Matlab. GUIs (Graphic User Interfaces) are programmed to respond to events. Unlike traditional programming, events generated by the user interface (caused by the user's interaction with it) drive the program's behavior.
GUIs allow users to interact with a computer program through various graphical objects (buttons, menus, checkboxes, etc.). These objects are usually operated with a mouse or keyboard.
Although graphical interfaces may seem secondary to developing an application's core, they must nevertheless be designed and developed with care and rigor. Their efficiency and ergonomics are essential in accepting and using these tools by end-users. A good design and a mastered development also allow for ensuring better maintainability.

The following graphical interfaces explain the use of the proposed platform, which allows training and testing on single or multimodal images of the face and iris.

Figures 4.1 and 4.2 represent the training and recognition applications for the face and iris, respectively.
In each application, to do the training, the user can choose the training database, choose if he wants to do a pre-processing and which technique to use, and choose the size of the matrix and the transformation method. During recognition, the user will use the database saved at the end of training (it contains feature vectors of each training set image). This database will be used to find a match with the test image.

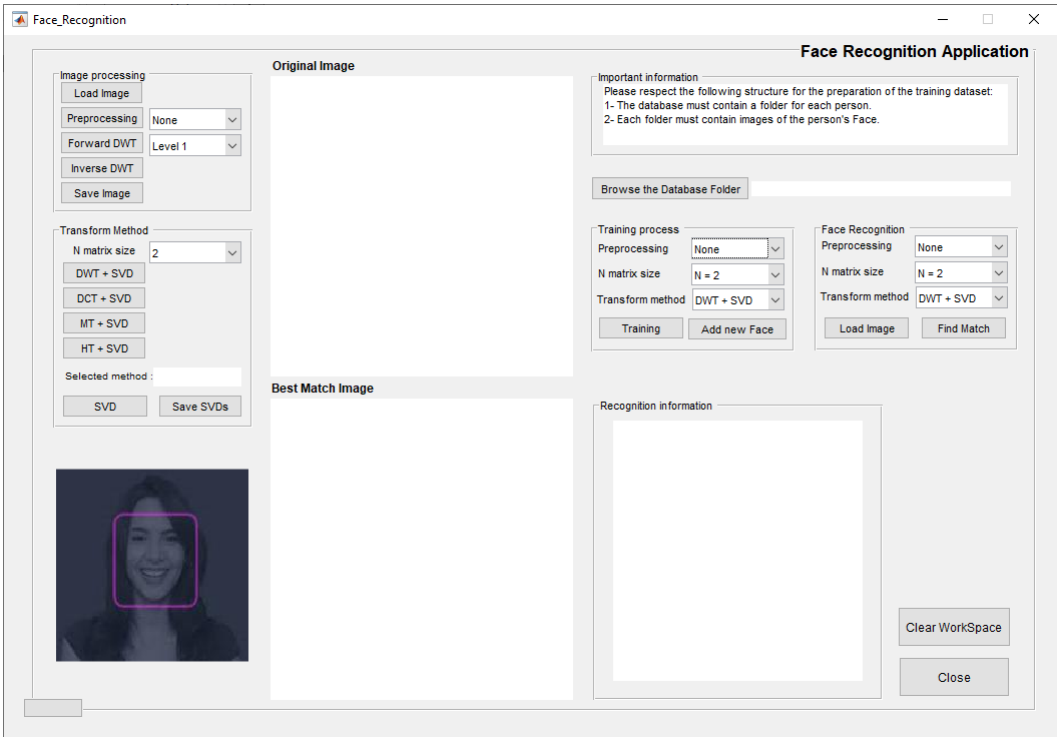


Figure 4.1: Face recognition application

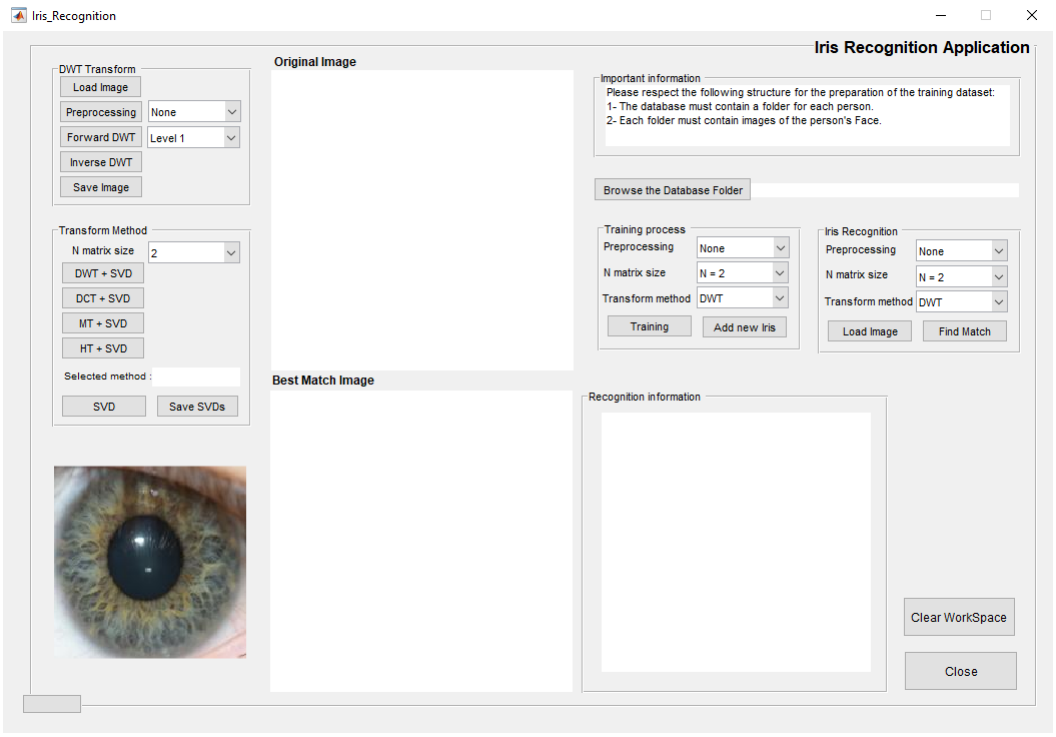


Figure 4.2: Iris recognition application

Figure 4.3 shows the multimodal face/iris training and recognition application. In this interface, the user will be able to choose a training dataset that respects the

following structure:

- The database must contain a folder for each person.
- Each folder must contain two subfolders named "Iris" and "Face."
- These two subfolders contain images of the person's faces and irises.

The user can also select the other training options similar to the face and iris applications (pre-processing, matrix size, and transformation method).

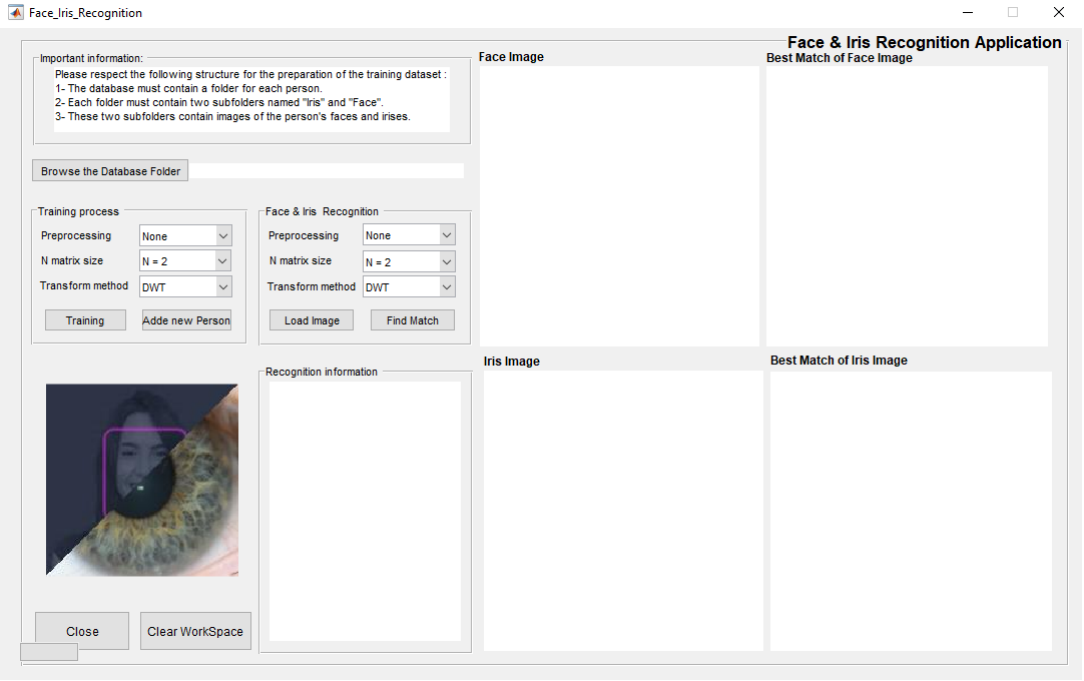


Figure 4.3: Face and Iris multimodal recognition application

During the recognition process, the user selects one image of the face and another of the iris. Suppose both images belong to the same person, and that person is in the database. In that case, the access authorization will be issued to the person, and a message will be displayed indicating that the recognized face and iris belong to the same person (see Figure 4.4).

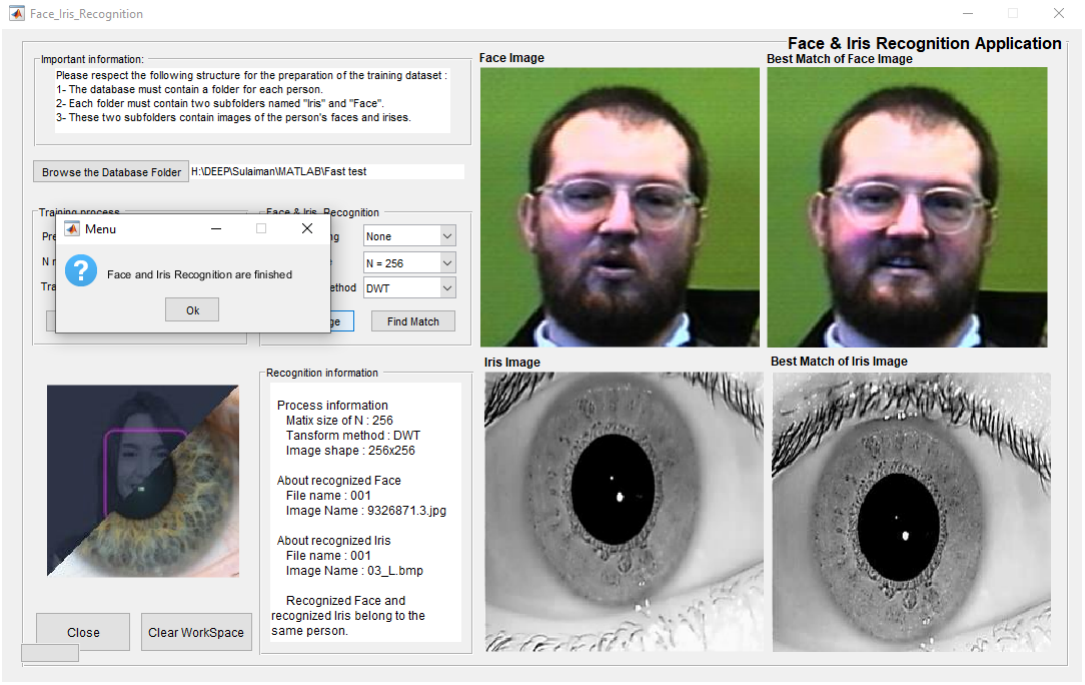


Figure 4.4: Face and Iris recognition - authorized person

Suppose the two images do not belong to the same person, and both persons are in the database. Access authorization will be denied in that case, and a message will be displayed indicating that the recognized face and iris do not belong to the same person (see Figure 4.5).

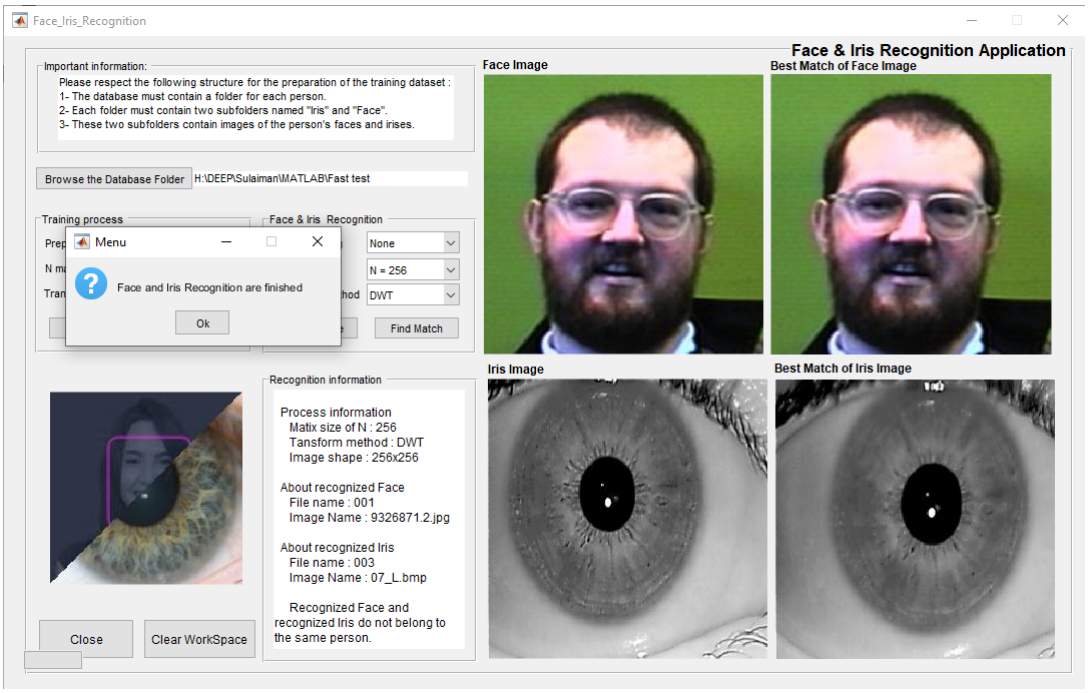


Figure 4.5: Face and Iris recognition - unauthorized person

Otherwise, if one or both biometric traits are not in the database, the user will not be authorized, and a message will indicate that there is no match.

4.3 Cyber security biometric platform solution using deep learning classifiers

Many useful libraries and projects have been created to help solve image processing problems with machine learning or improve processing pipelines in computer vision projects using Deep Learning.

It is possible to build an image processing application from scratch in theory. However, in reality, it is better to stand on the shoulders of giants and use what others have built and extend or adjust it as necessary.

This is where libraries and frameworks come in, and in image processing, where creating practical implementations is often difficult, this is even more true.

4.3.1 Used programming tools

In this part of the thesis, the development of the models was done using some of the most used tools in deep learning, namely Python (version 3.6.5) as a programming language and Keras (2.2.4) and Tensorflow (version 2.1.0) as deep learning frameworks.

TensorFlow is an open-source platform for machine learning and a symbolic math library used for machine learning applications.

Keras is an open-source library written in Python (under the MIT license) based primarily on the work of Google developer François Chollet as part of the ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System) project. This library aims to allow the rapid constitution of neural networks. In this context, Keras does not function as its framework but as an application programming interface (API) for accessing and programming different machine learning frameworks.

4.3.2 Design and training of deep learning models

In general, few people train a CNN entirely (training from a blank network, i.e., a network without weights) because of the difficulty of obtaining the large number of annotated images needed for this type of network. It is more common to find trained CNN models using Transfer Learning. Transfer Learning is a training technique, mostly used in Deep Learning, allowing to train a network to a certain task from a model already trained on a similar task (e.g., original network, m classes: animal classification; new network n classes: cat breed classification; with n being different from m). The knowledge from this pre-trained network is transferred to the new problem to assist in training using new images, and some steps of the network are then fine-tuned. This form of training has the advantage of requiring fewer data (thousands instead of millions) and allows much faster learning, from hundreds or thousands of computation hours to a few hours or even minutes.

Since training a CNN can take several weeks, even with several GPUs' computing power (Graphics Processing Units), it is common for large research structures to

make some of their networks available on the net. These networks are mostly trained on the ImageNet challenge images with 1000 object recognition classes [Russakovsky et al., 2015].

Transfer Learning can exploit a pre-trained CNN in different ways depending on the size of the new input data set and the similarity of the new images to those used in the original learning. The three main forms of Transfer Learning are as follows :

- **Total Fine-tuning:** In this case, the last fully connected layer (classification layer) is replaced by a classifier adapted to the new problem. All layers of the network are then re-trained on the new images. This strategy is used when the new image collection is large. As the weights are initialized with the pre-trained network's values and refined, this new learning is faster.

- **Partial fine-tuning:** As before, the last fully connected layer is replaced by a new classifier, but only some of the feature extraction layers are re-trained, the weights of the first layers being kept at their values. Since the first layers of a network can extract abstract or generic characteristics (outlines, colors, etc.) from the old training image set, and since the last layer progressively produces characteristics more specific to the original data classes, the weights of the latter are adjusted. This strategy is used when the new set of images is smaller and different.

- **Feature extraction:** The last case consists of using the pre-trained network's extracting layers to represent the new images of the new problem. Thus, the last fully connected layer is reset or removed to be replaced by a new classifier, and the parameters of the other layers are frozen. The new classifier will then be trained from the feature vectors extracted by the other pre-trained network layers. This strategy is used when the new collection of images is small and has similarities to the original images.

This thesis used the total fine-tuning of the existing models in the literature. The different architectures were trained using the VGGFace dataset [Parkhi et al., 2015b] for the face and the IIT Delhi dataset [Kumar and Passi, 2010] for the iris.

The first step to performing a fine-tuning using Keras is to initialize the deep architectures with the ImageNet weights. Since total fine-tuning has been adopted for the transfer of learning, the next step is to make all layers trainable. Once the architecture is defined, the final decision layers should be added (fully connected). For this, a Global Average pooling layer is applied to the model's output, followed by a Dense layer of 512 units and an output layer of 224 neurons, representing the number of classes. Furthermore, the Softmax function is used because this problem is multi-class (not a binary classification). The next step is crucial and concerns the preparation of the data. Indeed, the training and validation images are loaded and processed. The processing consists in augmenting the data and putting them in a data generator with batch size 64.

Data augmentation is a technique for artificially creating transformed versions of images in the training dataset that belong to the same class as the original image. Several transformations such as shifts, flips, zooms are applied. Data augmentation will allow the models to be more robust and perform better. The last step is the compilation of the models and the training process. The number of epochs, the optimizer, and callbacks must be specified in this step.

Once the training is finished, the best model will be saved in the file of type h5. This file contains the architecture with the model weights. Once the face and iris models are trained, they will be used according to the chosen combination mode (feature level fusion or decision level fusion). In the first case, the models are used as feature extractors. At the end of this step, two feature vectors will be generated (one for the face and the second for the iris), combined to make the matching. In the second case (decision level fusion), each model produces a classification, and the results are combined.

An automatic matching technique is also proposed in this thesis. It consists of training another binary classifier that will give access or not to an individual based on the distances between its features and those stored in the reference database.

4.4 Results and discussion of the developed cyber security biometric platform solution using machine learning

4.4.1 Datasets

In this thesis, for the machine learning part, two public datasets widely used in state of the art were used, the first is Faces94 which contains facial images [Hond and Spacek, 1997] (used for the face recognition experiments), and the second is IIT Delhi Iris dataset [Kumar and Passi, 2010] which contains iris images (used for the iris recognition experiments).

The IIT Delhi Iris dataset was collected from 224 students and staff at IIT Delhi, New Delhi, India. Iris images were captured in the indoor environment using JIRIS, JPC1000, digital CMOS camera and saved in bitmap (*.bmp) format with a resolution of 320x240 pixels.

The Face94 dataset was collected from 153 subjects (133 male and 20 female) and provided by Computer Vision Science Research Projects.

For the experiment of multimodal recognition (face and iris), the two previously cited datasets were grouped to result in a single dataset containing face and iris images for each person. The irises do not belong to the people on the faces, but this operation tests the different algorithms on multimodal classification (face and iris). The final used dataset contains 150 folders representing individuals. Each folder contains two sub-folders named "Face" and "Iris", which contain the images of faces and irises, re-

spectively. Figure 4.6 shows some examples of images from the Face and Iris datasets.

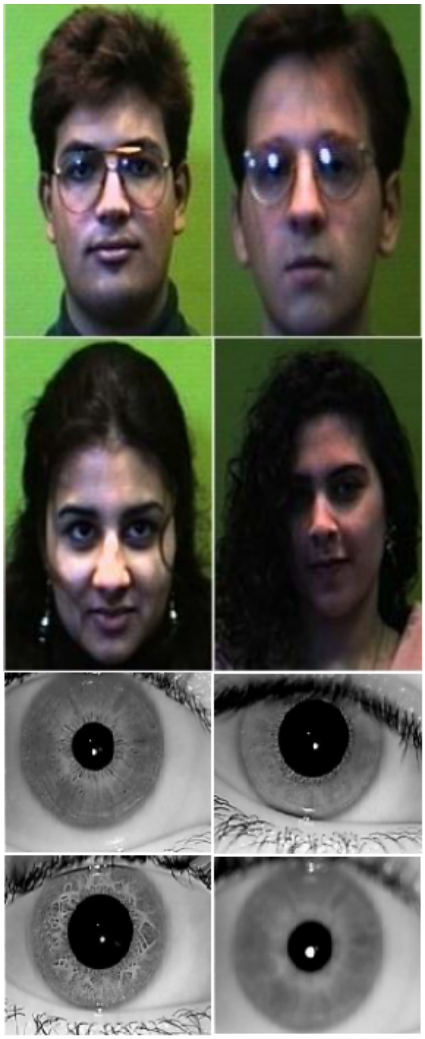


Figure 4.6: Face and Iris images

4.4.2 Experiments

For comparison purposes, experiments in two stages are conducted. The first is comparing different feature extraction techniques (DWT+SVD, DCT+SVD, HT+SVD, MT+SVD) with different parameters, including matrix size and preprocessing technique, and distance measure.

Experiments using 1500 images from the Face and Iris dataset were done. First, the different combinations of the proposed algorithms (DWT, DCT, HT, and MT com-

bined with SVD), the different matrix sizes (32, 64, 128, and 256), the preprocessing methods (contrast and Gamma correction), and the distance measures (Euclidean, Manhattan, and Cosine) are compared.

The goal is to visualize and compare the accuracy of each combination to perform a comparison with the state-of-the-art methods in multimodal biometry.

4.4.3 Results and discussion

Tables 3.2, 3.3, and 3.4 show the accuracy of different combinations with different distances.

It may be observed that the multimodal algorithm (DWT+SVD) gives the best result in the test, especially when the matrix size is equal to 128 and a contrast preprocessing is applied with the Euclidean distance as a matching metric (see Table 3.6).

The DWT gives the best results; this can be explained by the fact that DWT is better than DCT in terms of time and frequency resolution. Its coefficients are calculated by performing the successive Low pass and High pass filter on the Discrete-Time samples.

DWT is also better than HT and MT since it selects only the LL2 band, which contains the most useful features of the input image.

Euclidean distance gives the best results compared to the Cosine and Manhattan distance. Manhattan distance is usually preferred over the Euclidean distance when there is high dimensionality in the data [Aggarwal et al., 2001].

In this study, the use of SVD with DWT reduces the features vector's size, making the Euclidean distance more suitable for this case.

Table 4.1: Results of the different combinations using the Euclidean distance

Method	Matrix size	Pre-processing	Face	Iris	Face Iris
DWT	256	None	90.54	94.54	96.00
		Contrast	92.00	96.36	96.72
		Gamma corr.	91.63	95.27	96.36
	128	None	90.90	95.27	97.09
+		Contrast	94.54	96.72	**98.90**
		Gamma corr.	91.27	96.00	97.81
	64	None	90.18	93.45	95.27
		Contrast	91.27	93.81	96.36
SVD		Gamma corr.	90.90	93.45	95.63
	32	None	89.81	92.90	95.27
		Contrast	90.18	93.45	96.00
		Gamma corr.	90.18	93.81	96.00
DCT	256	None	88.00	92.36	94.90
		Contrast	89.09	93.45	95.27
		Gamma corr.	88.72	92.90	94.90
	128	None	87.63	91.63	94.54
+		Contrast	88.72	93.45	96.00
		Gamma corr.	86.90	91.27	95.63
	64	None	88.72	92.00	95.27
		Contrast	88.72	92.36	95.63
SVD		Gamma corr.	88.00	92.00	95.27
	32	None	86.90	90.90	93.81
		Contrast	87.63	91.27	94.54
		Gamma corr.	86.90	90.45	94.18
HT	256	None	85.81	87.63	90.18
		Contrast	85.81	88.00	90.54
		Gamma corr.	86.18	89.45	91.27
	128	None	86.18	88.90	91.63
+		Contrast	88.00	90.90	93.90
		Gamma corr.	87.63	91.27	93.45
	64	None	86.90	90.90	93.45
		Contrast	88.00	91.27	94.18
SVD		Gamma corr.	87.63	90.54	93.81
	32	None	86.18	89.45	92.00
		Contrast	86.90	89.81	92.90
		Gamma corr.	86.54	89.81	92.36
MT	256	None	82.18	85.18	88.90
		Contrast	83.63	85.54	90.54
		Gamma corr.	84.00	86.90	91.27
	128	None	83.63	86.90	91.63
+		Contrast	84.00	87.72	92.00
		Gamma corr.	86.18	89.09	92.36
	64	None	85.81	88.00	90.90
		Contrast	86.54	88.00	91.63
SVD		Gamma corr.	86.18	88.72	92.00
	32	None	82.18	85.09	87.27
		Contrast	82.54	85.81	87.63
		Gamma corr.	82.90	86.18	88.00

Table 4.2: Results of the different combinations using the Manhattan distance

Method	Matrix size	Pre-processing	Face	Iris	Face Iris
DWT	256	None	87.63	88.00	88.72
		Contrast	88.00	88.00	89.09
		Gamma corr.	84.72	86.18	86.54
	128	None	85.54	86.18	85.54
+		Contrast	87.27	86.90	87.27
		Gamma corr.	90.90	91.27	91.27
	64	None	91.27	91.27	93.81
		Contrast	92.36	92.90	**94.18**
SVD		Gamma corr.	89.81	89.09	90.54
	32	None	90.54	90.90	91.63
		Contrast	90.18	91.27	92.00
		Gamma corr.	89.45	90.90	91.27
DCT	256	None	86.90	87.27	87.63
		Contrast	87.63	88.00	90.18
		Gamma corr.	87.63	87.63	88.72
	128	None	85.81	86.18	87.27
+		Contrast	86.90	87.27	89.45
		Gamma corr.	85.81	87.27	88.00
	64	None	88.72	88.72	89.09
		Contrast	89.09	89.45	90.18
SVD		Gamma corr.	88.00	88.72	89.81
	32	None	87.27	87.27	88.72
		Contrast	88.72	89.09	90.18
		Gamma corr.	86.90	87.63	89.09
HT	256	None	83.27	83.63	84.00
		Contrast	85.81	86.54	87.63
		Gamma corr.	84.00	84.72	86.18
	128	None	86.54	87.63	89.45
+		Contrast	87.63	88.72	89.81
		Gamma corr.	87.63	88.00	89.09
	64	None	84.18	85.54	86.18
		Contrast	85.81	86.90	88.00
SVD		Gamma corr.	85.09	86.18	86.90
	32	None	83.63	83.63	84.00
		Contrast	85.09	86.54	87.27
		Gamma corr.	84.00	84.36	86.18
MT	256	None	81.81	82.18	83.27
		Contrast	82.18	83.63	85.09
		Gamma corr.	82.18	82.90	84.36
	128	None	82.54	82.90	83.63
+		Contrast	83.63	85.09	86.18
		Gamma corr.	83.27	84.00	85.81
	64	None	84.36	84.72	85.81
		Contrast	85.09	85.81	86.90
SVD		Gamma corr.	85.81	85.81	86.18
	32	None	80.00	81.09	82.18
		Contrast	82.54	83.63	84.36
		Gamma corr.	81.45	82.90	83.63

Table 4.3: Results of the different combinations using the Cosine similarity

Method	Matrix size	Pre-processing	Face	Iris	Face Iris
DWT	256	None	90.18	91.63	92.00
		Contrast	91.27	93.45	94.18
		Gamma corr.	90.54	92.36	93.81
	128	None	90.90	92.36	94.54
+		Contrast	92.00	94.90	**96.36**
		Gamma corr.	91.63	93.45	95.27
	64	None	91.27	93.45	94.18
		Contrast	92.00	94.18	95.63
SVD		Gamma corr.	92.36	93.81	94.54
	32	None	91.63	92.36	92.90
		Contrast	92.36	93.45	93.81
		Gamma corr.	92.00	92.90	93.45
DCT	256	None	87.27	88.00	89.45
		Contrast	88.72	90.54	91.63
		Gamma corr.	87.63	88.72	90.18
	128	None	86.54	87.27	89.09
+		Contrast	87.63	88.72	89.81
		Gamma corr.	88.00	89.45	90.18
	64	None	88.00	88.72	89.45
		Contrast	89.09	90.18	90.54
SVD		Gamma corr.	90.18	90.90	91.63
	32	None	88.72	89.09	90.18
		Contrast	89.09	90.54	91.27
		Gamma corr.	89.09	89.81	90.54
HT	256	None	84.72	85.09	86.18
		Contrast	86.90	87.27	88.00
		Gamma corr.	85.54	86.90	87.27
	128	None	88.00	88.72	89.45
+		Contrast	88.72	89.81	91.27
		Gamma corr.	88.72	89.45	90.90
	64	None	86.18	87.27	88.00
		Contrast	87.63	88.72	89.81
SVD		Gamma corr.	87.27	88.00	89.09
	32	None	84.72	85.09	85.54
		Contrast	86.18	87.27	88.72
		Gamma corr.	85.54	86.90	87.63
MT	256	None	83.27	84.00	84.72
		Contrast	84.36	85.09	86.90
		Gamma corr.	82.90	83.27	85.81
	128	None	84.00	85.54	87.27
+		Contrast	85.09	86.18	88.72
		Gamma corr.	84.72	85.54	88.00
	64	None	84.36	85.09	87.63
		Contrast	85.54	87.27	89.81
SVD		Gamma corr.	84.72	86.18	88.72
	32	None	81.45	82.54	83.63
		Contrast	82.90	84.00	85.09
		Gamma corr.	83.27	84.36	86.18

Table 4.4: Best obtained results from different combinations

Distance	Method	Matrix size	Pre-proc.	Face & Iris
Euclidean	DWT+SVD	128	Contrast	98.90
Manhattan	DWT+SVD	64	Contrast	94.18
Cosine	DWT+SVD	128	Contrast	96.36

4.4.4 Discussion of the findings of machine learning with machine learning

The second stage of the experiments concerns a comparative study with some state-of-the-art techniques that address the problem of multimodal classification.
To achieve this objective, the best combination from the previous experiments (DWT+SVD) are used following these steps:

- Apply three-level DWT on the extracted face and iris images and save them into two separate files.
- Apply SVD on the third level DWT face and iris images and obtain the singular values of face and iris images separately.
- Merge the singular values of combined DWT-SVD algorithms into a feature vector.
- Save the feature vector on the database for future retrievals.

Table 4.4.4 shows the results of the proposed method compared with four other techniques. The proposed methods and the above-described face and iris recognition techniques are implemented to judge the outcome of this thesis. The experiment was done in the same environment and on the same training and test samples (same dataset).

Table 4.5: Machine learning comparative results

Method	Feature extraction	Matching	Acc.
G. Huo et al. [Huo et al., 2015] (2015)	2D Gabor filter with different scales and orientations, then transform them by histogram statistics into an energy-orientation. PCA method is used for dimensionality reduction.	Support Vector Machine (SVM)	97.81
Y. Bouzouina et al. [Bouzouina and Hamami, 2017] (2017)	PCA and discrete coefficient transform (DCT) for facial features. 1D Log-Gabor filter method and Zernike moment for iris features. Genetic algorithm (GA) is used for dimensionality reduction.	Support Vector Machine (SVM)	96.72
B. Ammour et al. [Ammour et al., 2018] (2018)	Two-dimensional Log-Gabor filter combined with spectral regression kernel discriminant analysis.	Euclidean Distance	97.45
B. Ammour et al. [Ammour et al., 2020] (2020)	The dataset is pre-processed using the histogram equalization then the features are extracted from face images using singular spectrum analysis (SSA) and normal inverse Gaussian (NIG) combined with statistical features of wavelet. Featureextraction from iris images was performed using multi-resolution 2D Log-Gabor filter and spectral regression kernel discriminant analysis (SRKDA)	Fuzzy K-Nearest Neighbor (FK-NN)	98.18
Our method	The dataset is pre-processed using the contrast method then the feature extraction is made by DWT and SVD with the matrix size equal to 128.	Euclidean Distance	98.90

The obtained results confirm the superiority of the proposed approach in comparison with other techniques. Results also confirm that techniques using Euclidean distance give better results than techniques using a Support vector machine (SVM). Fuzzy k-nearest neighbor (FK-NN) also gives good results, which means that, for the matching, techniques based on distance give better results than techniques based on classical machine learning.

For feature extraction, it can be noticed that DWT combined with SVD ensures a good characterization of the images and thus a good recognition of people. DWT provides simultaneous spatial and frequency domain information of the image, and the SVD selects the best features obtained from DWT.

This approach has proven effective in authenticating individuals using their facial and iris images. This model could be used in a multimodal biometric system using similar images.

4.5 Results and discussion of the developed cyber security biometric platform solution using deep learning

In this part of the thesis, a detailed study on five deep architectures for iris recognition and five for face recognition is performed. A multimodal dataset formed by merging two other datasets is used to train and test the models.

4.5.1 Datasets

Two public datasets widely used in the state-of-the-art are utilized. The first one is the IIT Delhi Iris dataset [Kumar and Passi, 2010], which contains 2240 iris images collected from 224 subjects using JIRIS, JPC1000, digital CMOS camera, and saved in bitmap (*.bmp) format with a resolution of 320x240 pixels.

The second is the VGG-Face dataset proposed by the VGG group [Parkhi et al., 2015b]. The dataset consists of 2,622 identities for a total of 982,803 images. Each identity has an associated text file containing URLs for images and corresponding face detections.

The two previously cited datasets are grouped for the multimodal recognition experiment (face and iris), like in the first experiments. Since the iris dataset contains only 224 subjects, the same number of data from the VGG-Face dataset is randomly selected.

The irises do not belong to the people on the faces, but this combination is necessary to test the different algorithms on multimodal classification (face and iris). The final used dataset contains 224 folders representing individuals. Each folder contains two sub-folders named "Face" and "Iris," which contain the images of faces and irises, respectively. Figure 4.7 shows some examples of images from the Face and Iris datasets.

Figure 4.7: Face and Iris images

4.5.2 Experiments

This study presents many steps for the training and testing phase of the face and iris authentication classifiers.

Some of these steps (i.e., pre-processing of face images, model training, choice of fusion technique, and fusion threshold) are studied under different configurations. To facilitate the understanding of these different settings, they are summarized in the form of two points: training of the face and iris classifiers and evaluation of the results; and study of two fusion techniques of the two modalities with the proposed matching technique.

First, the biometric modalities (face and iris) are evaluated independently, and the test protocols and results are described in this section. Then, both modalities are merged at the feature or decision levels and evaluated with automatic matching.

All experiments were performed on a local machine (desktop computer with an i7 4820k processor, 64GB of ram, and two GTX1070 8GB graphics cards).

The face training was performed on the VGG-Face database [Parkhi et al., 2015b] and the iris training on the IIT Delhi Iris dataset [Kumar and Passi, 2010]. Each database is divided into three parts (60% for training, 20% for validation, and 20% for testing).

Since the iris database in our possession contains only 224 classes, and for comparison purposes, Only 224 files from the VGG-Face database are used to have a multimodal database containing the iris and face images for each person.

For the test, only 448 images per modality are kept to evaluate the fusion. In addition, 60 face and 60 iris images considered unauthorized persons are added. These data are randomly selected from the rest of the VGG-Face database and some fake authorized persons (non-authorized person wearing a mask of an authorized person) for the face images and from the MMU1 Iris dataset [Multimedia-University, Last Accessed (February 2021] for the iris images (see Table 4.6).

Table 4.6: Used face and iris datsets

	# images Training	**# images Validation**	**# images Test**
Face	10386	3410	448+60
Iris	1344	448	448+60

First of all, the datasets were pre-processed; they contain images of different sizes. They are readjusted so that all images will match each model's input. Table 4.7 summarizes the size of input images of each used network.

Table 4.7: Input images size for each model

	Face				
Model	**VGGFace**	**FaceNet**	**OpenFace**	**Inception V3**	**ResNet 50**
Input image size	224x224	160x160	96x96	224x224	224x224
	Iris				
Model	**VGG-16**	**MobileNet**	**DenseNet-121**	**Inception V3**	**ResNet 50**
Input image size	224x224	224x224	224x224	224x224	224x224

During the training, a data augmentation that will allow the models to be more robust and perform better is performed. Data augmentation is a technique for artificially creating transformed versions of images in the training dataset that belong to the same class as the original image. Several transformations such as shifts, flips, zooms, and more, are performed.

Feature extraction is a crucial and indispensable step for classification and allows one to consider only relevant elements by optimally describing the image via image-specific features.

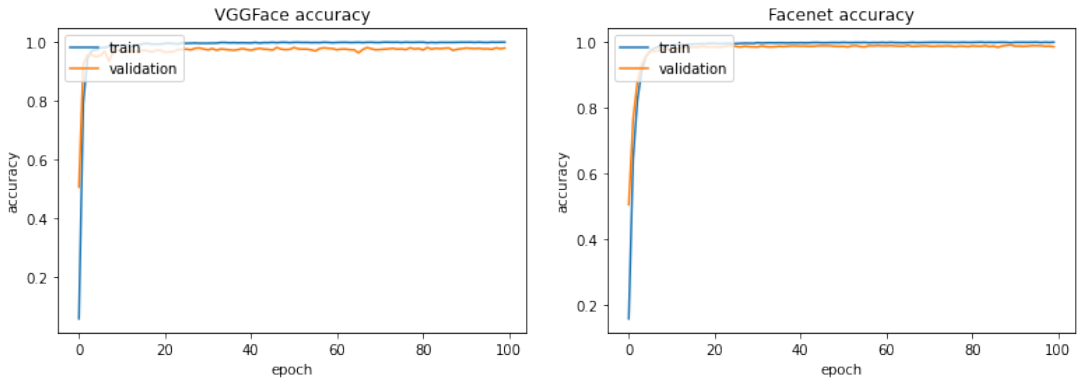
All the models used for the face (VGGFace, FaceNet, OpenFace, Inception V3, and ResNet 50) and iris (VGG-16, MobileNet, DenseNet-121, Inception V3, and ResNet 50) feature extraction were trained during 100 epochs using batches of size 32 and the Stochastic Gradient Descent (SGD) method to reduce the loss function.

4.5.3 Results and discussion

4.5.3.1 Unimodal recognition results

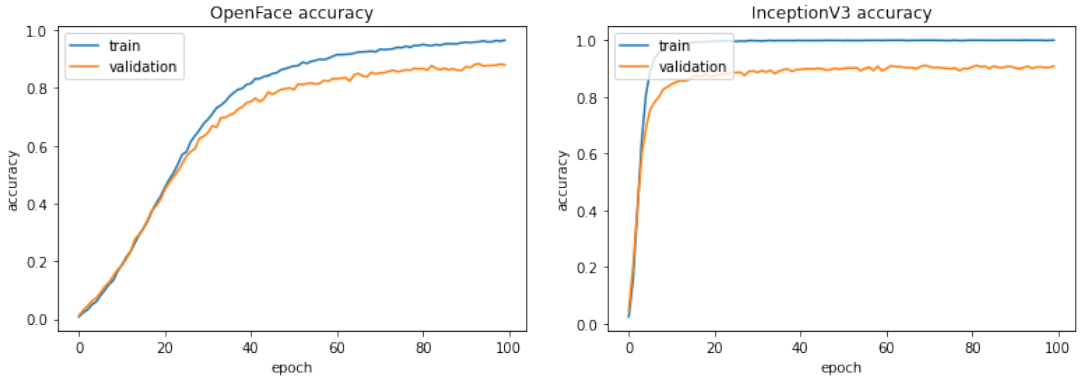
The classifiers were trained following the protocols outlined above. In Figures 4.8 and 4.9, the curves show that all models have learned well from the dataset and suffer neither from over- nor under-fitting. This can be explained by the fact that SGD,

as explained in Hardt et al. [2016], contains the error, stays stable, and prevents overfitting even without any regularization term.



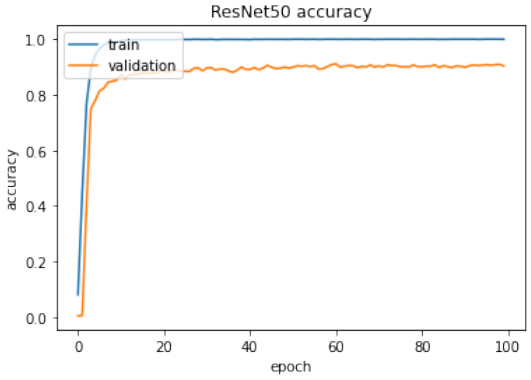
(a) VGGFace

(b) FaceNet



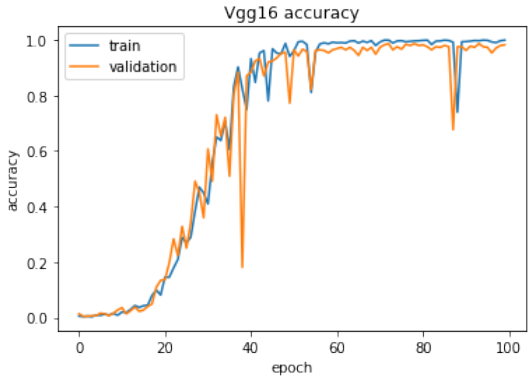
(c) OpenFace

(d) InceptionV3

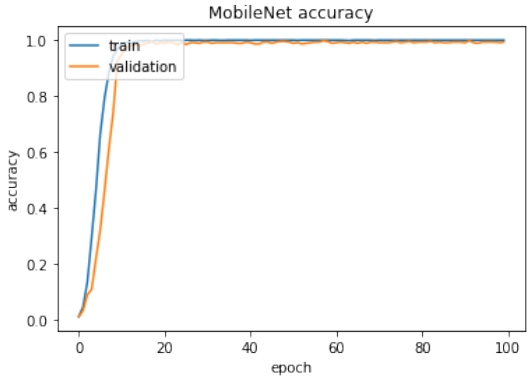


(e) ResNet50

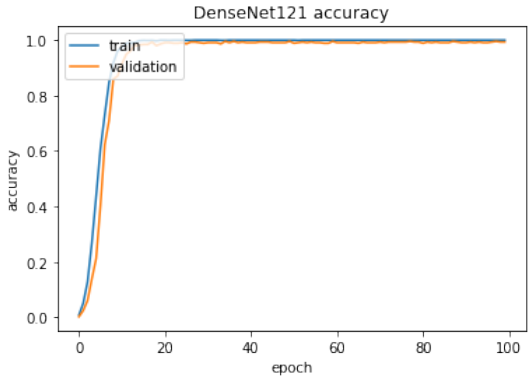
Figure 4.8: Train and validation accuracy of different face models



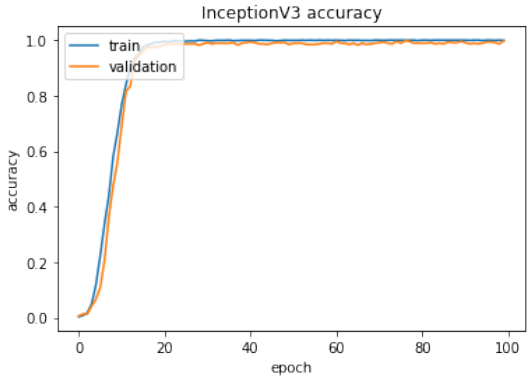
(a) VGG-16



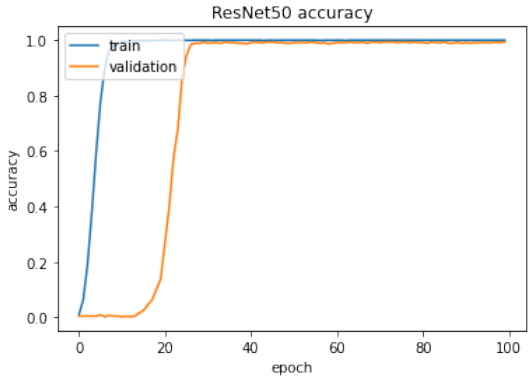
(b) MobileNet



(c) DenseNet-121



(d) InceptionV3



(e) ResNet50

Figure 4.9: Train and validation accuracy of different iris models

As explained in the previous section, the trained models are used as feature extractors, and the next step is to find the matching if it exists. The existence or not of a match depends on the distance used and the decision threshold. During the previous experiments with machine learning, it has been concluded that the Euclidean distance was the best suited for this case; this is why it has been adopted.

In this thesis, an automatic method for matching is proposed. A multilayer neural network classifier is trained on a database containing the distances between each feature of the validation database and the training database's features.

The used neural network is a multilayer perceptron (MLP); it contains an input layer of three neurons representing Dmin, Dmean, and Dmax distances. It also contains two hidden layers, each containing five neurons and an output layer containing one neuron. This network is trained on 100 epochs using gradient descent and a training step of 0.1. The results obtained for face and iris recognition are detailed in Tables 4.8 and 4.9.

Table 4.8: Face recognition results

Metric	VGGFace	FaceNet	OpenFace	ResNet50	InceptionV3
Accuracy (%)	22,77	89,51	62.94	52,23	57.81
Precision (%)	34,67	91,66	74,81	64,15	59.70
Recall (%)	22,67	89,11	62.66	52,00	57.55
FRR (%)	76,79	5,80	33.48	45,54	39.95
FAR (%)	76,67	6,67	36,67	46,67	43,33

Table 4.9: Iris recognition results

Metric	VGG-16	MobileNet	DenseNet-121	ResNet50	InceptionV3
Accuracy (%)	90,63	93,30	93.53	93.53	93.08
Precision (%)	93,36	94.25	94.25	94.25	94.25
Recall (%)	90,22	92.89	93.11	93.11	92.66
FRR (%)	3,35	0.67	0.45	0.45	0.89
FAR (%)	3,33	0	0	3,33	3,33

These tables contain the recognition results for each architecture in terms of accuracy, precision, and recall. Accuracy represents the number of well-recognized individuals out of the total number. The FAR and FRR errors have also been considered, representing the system's tolerance to recognition errors.

By analyzing the face results tables, it can be seen that the FaceNet model largely outperforms the other models and provides a very good compromise between the performances (accuracy, precision, and recall) and the FAR and FRR error. FaceNet provided a very low FAR and FRR, which confirm that the system has good capabilities to detect authorized and unauthorized persons. In this case, the system can therefore be considered restrictive regarding security, which is perfectly suited to this application's context. In comparison, the VGGFace, OpenFace, InceptionV3, and ResNet50 models do not give good results.

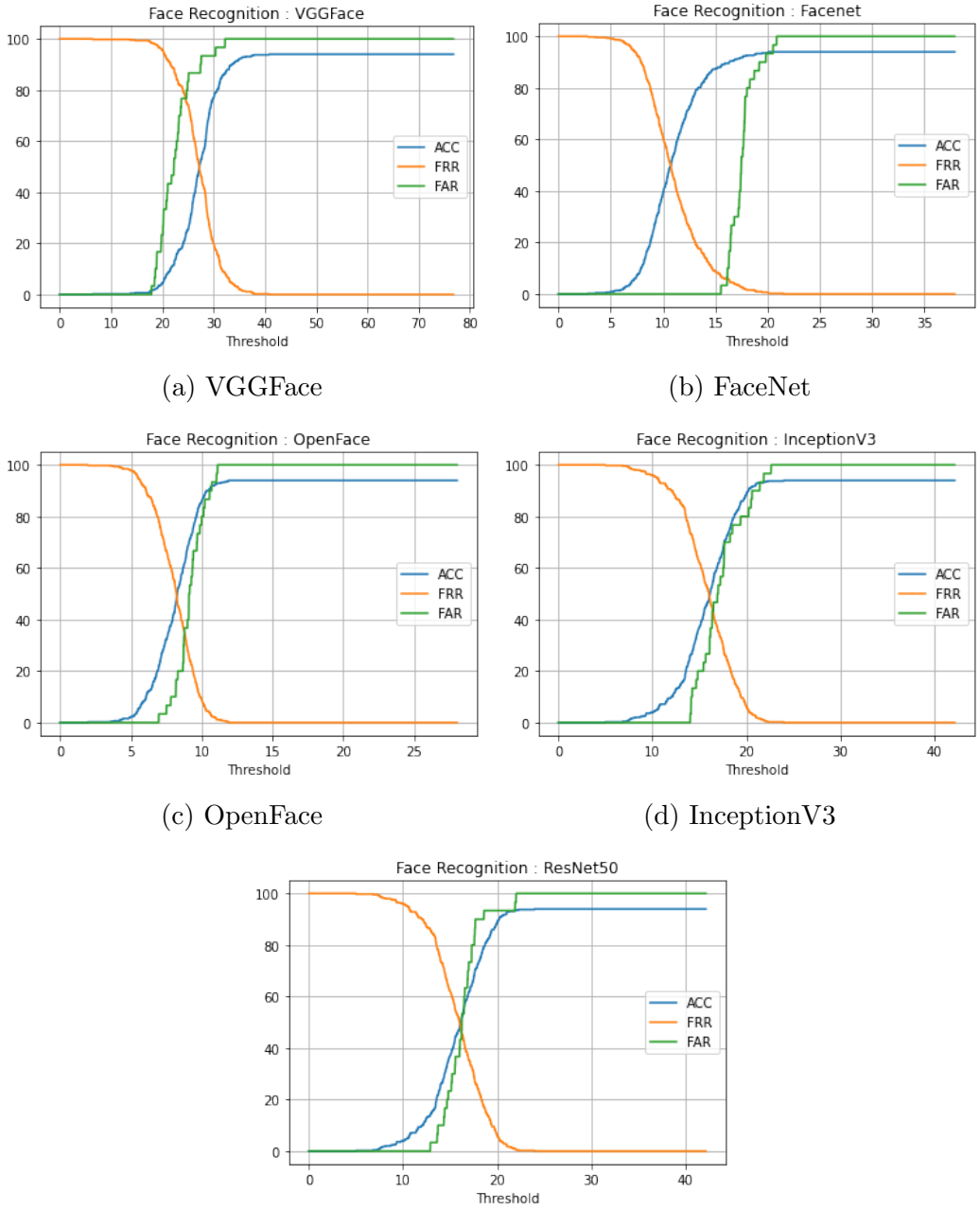
Analyzing the iris results shows that the best results are obtained using DenseNet-121, MobileNet, ResNet50, and InceptionV3. However, DenseNet-121 is slightly better since it gives a FAR rate equal to zero. VGG-16 is the worst performing model with 90.63, 3.35, 3.33 for accuracy, FRR, and FAR, respectively.

A general observation is that the DenseNet-121 and DeepFace models learned well from the data to perform their respective tasks. The rates are obtained by automatically matching without having to manually choose them.

To explain the influence of thresholding on the performance of models, a cross-study is performed. Each architecture is tested with the different values of the threshold θ.

The α scalar allows us to traverse all possible values from 0 to the maximum distance between each element of the FDB_val base and the elements of the FDB_train base. Thus, a global view of the impact of the threshold's choice on the accuracy, FAR, and FRR will be seen. The scalar alpha is chosen differently for each model. This application has chosen a compromise between the FAR and FRR error, so the threshold represents the best compromise between the Accuracy, FAR, and FRR.

Figures 4.10 and 4.11 show the curves for FAR, FRR, and accuracy as a function of threshold. In the literature, the suggested threshold is the EER (Equal Error Rate) point, which is the intersection between the FAR and FRR curves (see Figure 2.2).



(a) VGGFace

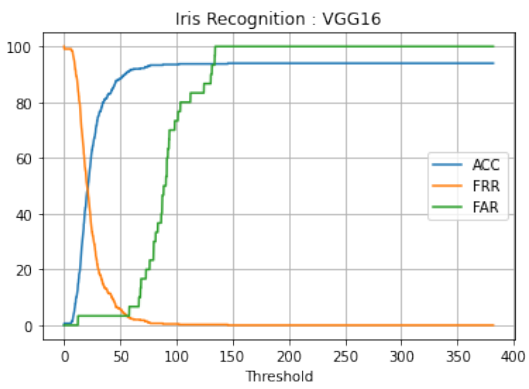
(b) FaceNet

(c) OpenFace

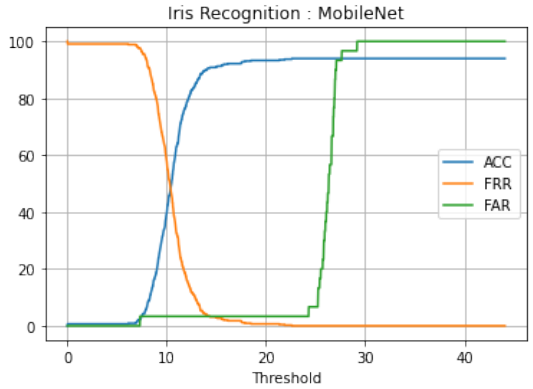
(d) InceptionV3

(e) ResNet50

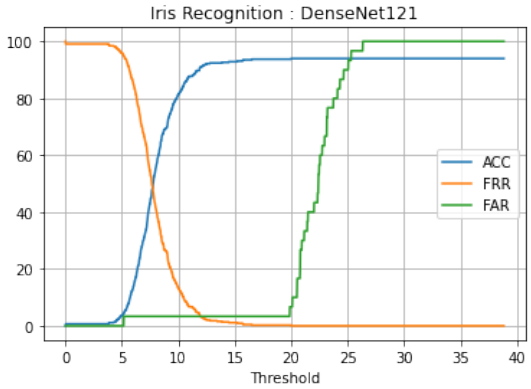
Figure 4.10: Test accuracy, FRR, and FAR of different face models



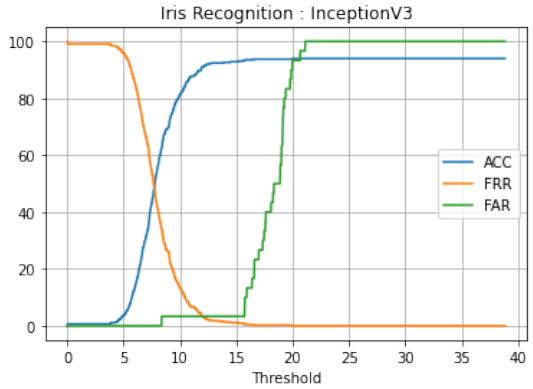
(a) VGG-16



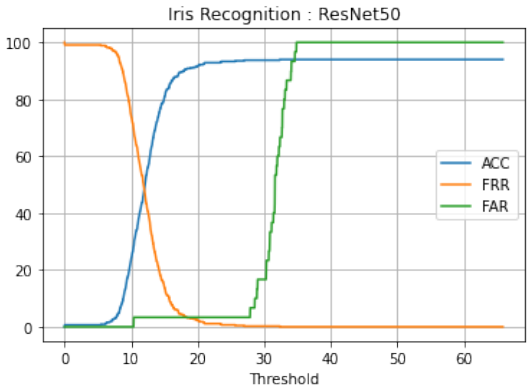
(b) MobileNet



(c) DenseNet-121



(d) InceptionV3



(e) ResNet50

Figure 4.11: Test accuracy, FRR, and FAR of different iris models

The different curves show that the EER point is not always the right compromise between the accuracy and the FAR and FRR errors. The rates obtained by the automatic method and those obtained with the EER are summarized in Tables 4.10 and 4.11.

Table 4.10: Comparaison between matching using MLP and ERR for face recognition

	Metric	VGGFace	FaceNet	OpenFace	ResNet50	InceptionV3
MLP	Accuracy (%)	22,77	89,51	62.94	52,23	87.81
	Precision (%)	34,67	91,66	74.81	64,15	59.70
	Recall (%)	22,67	89,11	62.66	52,00	57.55
	FRR (%)	76,79	5,80	33.48	45,54	39.95
	FAR (%)	76,67	6,67	36,67	46,67	43,33
EER	Accuracy (%)	22,77	89,51	61.61	52.23	54.24
	Precision (%)	34,67	91,66	74,15	64,15	66,37
	Recall (%)	22,67	89,11	61,33	52,00	54,00
	FRR (%)	76,79	5,80	35,27	45,54	43,53
	FAR (%)	76,67	6,67	36,67	46,67	43,33
	Threshold EER	24,54	16,13	8,78	16,23	16,42

Table 4.11: Comparaison between matching using MLP and ERR for iris recognition

	Metric	VGG-16	MobileNet	DenseNet-121	ResNet50	InceptionV3
MLP	Accuracy (%)	90,63	93,30	93.53	93.53	93.08
	Precision (%)	93,36	94.25	94.25	94.25	94.25
	Recall (%)	90,22	92.89	93.11	93.11	92.66
	FRR (%)	3,35	0.67	0.45	0.45	0.89
	FAR (%)	3,33	0	0	3,33	3,33
EER	Accuracy (%)	90,63	90,63	90,85	90,63	90,85
	Precision (%)	93,36	93,73	93,19	93,21	93,19
	Recall (%)	90,22	90,22	90,44	90,22	90,44
	FRR (%)	3,35	3,35	3,35	3,35	3,35
	FAR (%)	3,33	3,33	3,33	3,33	3,33
	Threshold EER	55,59	14,15	11,91	18,15	11,91

According to Tables 4.10 and 4.11, the proposed method gives good results compared to those that used the EER threshold. This can be explained by the fact that the proposed MLP learned how to make a difference between them automatically instead of choosing a threshold variable from the validation dataset and the unrecognized dataset.

This approach could be used with any model in a task that favors a compromise between FAR and FRR error.

4.5.3.2 Multimodal recognition results

In this work, two fusion approaches are tested. The first one concerns decision-level fusion, and the second one feature-level fusion.

In the score-level fusion, each recognition model for each biometric feature (face and iris) performs recognition independently. The fusion consists of grouping the decisions so that the person will be recognized if the face and iris images are recognized and belong to the same person.

Feature-level fusion consists of independently merging the two feature vectors generated by each recognition model (face and iris). The fusion result will be considered a single feature and used to find the match in the database.

The following tables (Table 4.12 and 4.13) summarize the obtained results for each of the possible combinations between the iris and face recognition models.

Table 4.12: Decision-level fusion results

Metric	DenseNet - VGGFace	DenseNet - FaceNet	DenseNet - OpenFace	DenseNet - ResNet	DenseNet - Inception
Acc. (%)	36,83	**57,59**	45,31	39,29	57,37
Prec. (%)	24,94	**49,69**	40,74	32,09	55,72
Recall (%)	36,67	**57,33**	45,11	39,11	57,11
FRR (%)	0,67	**1,56**	10,04	12,95	12,50
FAR (%)	0,00	**0,00**	0,00	0,00	0,00
Metric	Inception -VGGFace	Inception -FaceNet	Inception -OpenFace	Inception - ResNet	Inception -Inception
Acc. (%)	13,62	17,41	10,71	11,61	31,25
Prec. (%)	4,41	7,27	3,30	3,40	18,03
Recall (%)	13,56	17,33	10,67	11,56	31,11
FRR (%)	0,45	0,45	0,22	0,45	0,67
FAR (%)	0,00	0,00	0,00	0,00	0,00
Metric	MobileNet -VGGFace	MobileNEt -FaceNet	MobileNet -OpenFace	MobileNet -ResNet	MobileNet -Inception
Acc. (%)	12,95	17,41	13,84	11,83	30,13
Prec. (%)	4,77	7,27	5,13	3,82	17,05
Recall (%)	12,89	17,33	13,78	11,78	30,00
FRR (%)	0,22	0,45	0,45	0,45	0,67
FAR (%)	0,00	0,00	0,00	0,00	0,00
Metric	ResNet - VGGFace	ResNet - FaceNet	ResNet - OpenFace	ResNet - ResNet	ResNet - Inception
Acc. (%)	5,80	17,41	13,62	12,50	31,03
Prec. (%)	1,94	7,27	4,72	4,04	17,61
Recall (%)	5,78	17,33	13,56	12,44	30,89
FRR (%)	0,45	0,45	0,22	0,45	0,67
FAR (%)	0,00	0,00	0,00	0,00	0,00
Metric	VGG16 - VGGFace	VGG16 - FaceNet	VGG16 - OpenFace	VGG16 - ResNet	VGG16 - Inception
Acc. (%)	19,20	17,41	15,63	12,50	31,47
Prec. (%)	9,05	7,27	6,12	4,04	18,18
Recall (%)	19,11	17,33	15,56	12,44	31,33
FRR (%)	0,67	0,45	0,22	0,45	0,67
FAR (%)	0,00	0,00	0,00	0,00	0,00

Table 4.13: Feature-level fusion results

Metric	DenseNet - VGGFace	DenseNet - FaceNet	DenseNet - OpenFace	DenseNet - ResNet	DenseNet - Inception
Acc. (%)	96,65	**99,78**	99,55	88,39	96,65
Prec. (%)	99,11	**99,56**	99,56	94,22	99,11
Recall (%)	96,22	**99,33**	99,11	88,00	96,22
FRR (%)	3,35	**0,22**	0,45	11,61	3,35
FAR (%)	3,33	**0,00**	0,00	10,00	3,33
Metric	Inception -VGGFace	Inception -FaceNet	Inception -OpenFace	Inception - ResNet	Inception -Inception
Acc. (%)	95,09	99,33	99,33	80,13	93,30
Prec. (%)	97,78	99,56	99,56	90,22	99,11
Recall (%)	94,67	98.89	98,89	79,78	92,89
FRR (%)	4,91	0,67	0,67	19,87	6,70
FAR (%)	3,33	0,00	0,00	20,00	6,67
Metric	MobileNet -VGGFace	MobileNEt -FaceNet	MobileNet -OpenFace	MobileNet -ResNet	MobileNet -Inception
Acc. (%)	96,65	99,33	99,33	56,70	96,65
Prec. (%)	99,11	99,56	99,56	72,89	99,11
Recall (%)	96,22	98.89	98,89	56,44	96,22
FRR (%)	3,35	0,67	0,67	43,30	3,35
FAR (%)	3,33	0,00	0,00	40,00	3,33
Metric	ResNet - VGGFace	ResNet - FaceNet	ResNet - OpenFace	ResNet - ResNet	ResNet - Inception
Acc. (%)	99,55	99,33	98,44	95,31	96,65
Prec. (%)	99,56	99,56	99,11	98,22	99,11
Recall (%)	99,11	98,89	98,00	94,89	96,22
FRR (%)	0,45	0,67	1,56	4,69	3,35
FAR (%)	0,00	0,00	0,00	3,33	3,33
Metric	VGG16 - VGGFace	VGG16 - FaceNet	VGG16 - OpenFace	VGG16 - ResNet	VGG16 - Inception
Acc. (%)	96,88	96,65	96,65	96,65	96,65
Prec. (%)	99,56	99,56	99,56	99,56	99,56
Recall (%)	96,44	96,22	96,22	96,22	96,22
FRR (%)	3,13	3,35	3,35	3,35	3,35
FAR (%)	3,33	3,33	3,33	3,33	3,33

It can be noticed on the table of the decision-level fusion results that the performances are not satisfying. This can be explained by the fact that decision fusion depends on the individual performance of each classifier. Since the DenseNet-121 and FaceNet models gave the best individual results, it would be logical that their combination remains the best performing. Here the goal was to find the best compromise between

FAR and FRR error, but unfortunately, the results obtained were at the expense of accuracy.

Regarding the feature-level fusion results, the combination of the vectors extracted by the DenseNet-121 model and the FaceNet model for the iris and face gave good results. Indeed, the model reached an accuracy rate of 99.78, with an FRR of 0.22 and with 0 FAR.

4.6 Discussion of the findings of deep learning with deep learning

In the second part of this thesis, the individual recognition problem using deep learning is tackled.

To expand this study and validate the results, a comparative study with some state-of-the-art works addressing multimodal biometric recognition is performed.

Table 4.14 summarizes the performance obtained by each technique. All methods are implemented in the same environment and using the same training and testing data. The results indicate that the approach outperforms the other two techniques.

Table 4.14: A comparative study with techniques from the state-of-the-art

Metric	[Leghari et al., 2021]	[Alay and Al-Baity, 2020b]	Our work
Accuracy (%)	97,09	95,091	**99.78**
Precision (%)	97,11	94,67	**99.56**
Recall (%)	96,67	94,67	**99.33**
FRR (%)	0,45	6,25	**0.22**
FAR (%)	**0**	**0**	**0**

The obtained results by the DenseNet121-FaceNet model by adopting feature-level fusion and using automatic matching are very satisfactory and exceed those obtained in state-of-the-art. The proposed approach also ensured 99.78% accuracy, 99.56% precision, and 99.33% recall while keeping FAR at 0.22% and FRR at 0%. On the same test dataset, other approaches were less performed. Such performance will allow security systems to guarantee access to all authorized individuals and reject all unauthorized ones.

4.7 Discussion of the findings of automatic matching

This part of the thesis aims to experimentally find the best combination between state-of-the-art architectures for multimodal face and iris recognition. Significant improvements in existing multimodal deep learning models have been achieved in this thesis to make the resulting model more accurate in terms of accuracy, precision, and recall while keeping a good compromise between FAR and FRR errors.

In this context, a new approach for automatic matching is proposed, based on individuals' automatic classification based on the distance between the feature vectors representing them.

Figure 4.12 illustrates the proposed approach. Combining the DenseNet-121 and FaceNet models for iris and face feature extraction proved very efficient. Indeed, the user avoids worrying about the threshold choice by using the proposed automatic matching technique that classifies each distance vector.

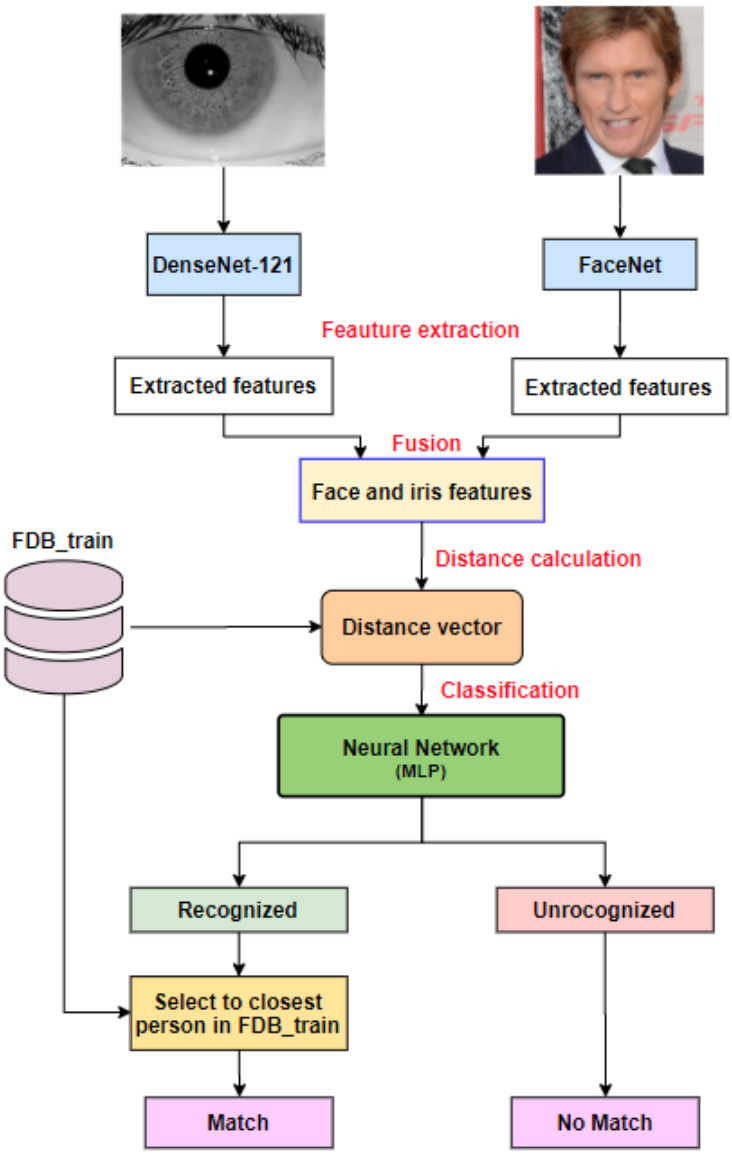


Figure 4.12: Proposed feature-level fusion approach

The proposed automatic matching approach does not use the threshold but tries to ensure a better compromise between performance and FAR and FRR errors. It is worth noting that when plotting the different curves of accuracy, FAR, and FAR, it can be noticed that the choice of the threshold directly impacts the performance, and its choice depends on the use case of the system. Figure 4.13, which represents the performances curve of the DenseNet121-InceptionV3 feature-level fusion, can be taken as an example.

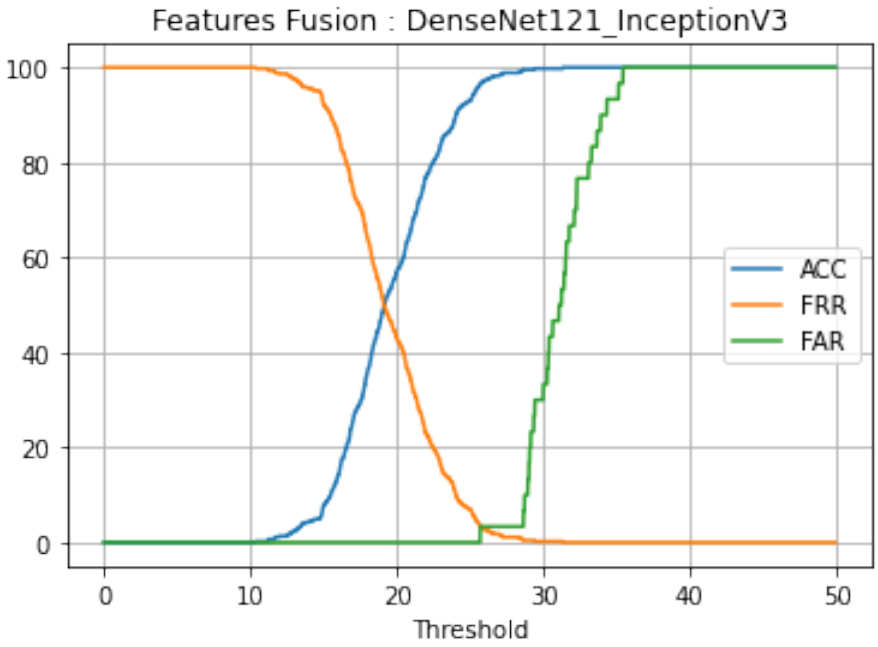


Figure 4.13: DenseNet121-InceptionV3 fusion performances

For applications that require a trade-off between FAR and FRR error, most techniques in the literature use the EER point as the threshold choice. However, it has been shown in this thesis that this choice is not always optimal by comparing the results obtained with this point with the results obtained with the automatic matching technique. This clearly states the most important finding of that study.

For applications that require a minimal tolerance for unauthorized persons (access to a very restricted area), a threshold lower than the EER point is preferred.

For applications that favor rapid recognition of persons at the FAR value's expense (e.g., unlocking a phone), a threshold higher than the EER point is preferred.

4.8 Discussion of the results implications

The complete biometric security system consists of different processing blocks. The images of people (face and iris) presented for recognition are captured and processed by intelligent cameras. Thus, only the results of these treatments are transmitted over the network, the personal data (such as the subject's images) being strictly lim-

ited to the camera. Whatever its architecture or objective, a camera is defined as intelligent when it includes a computing part to directly process the acquired image.

The images acquired by the sensor are therefore processed within the camera itself to extract the relevant information. The results are encrypted and transmitted to the global system via integrated communication interfaces. The hardware target in charge of onboard processing can be of various types. It can be integrated circuits dedicated to a non-modifiable task or more flexible targets, modifying the nature of the applied processing (for example, an NVIDIA Jetson Nano card).
The latter has significant computing power, is particularly efficient for implementing regular tasks, and has strong potential for data parallelism. Integrating the feature extraction approach on these cards, encrypting the result using the approach presented in chapter 3, and transmitting the result into the network allows the server to verify the database and return the result to the user interface.

The quality of the camera and the hardware used for data processing directly impact the speed of the response. Indeed, using a low-performance card to extract features from the image captured by the camera could lead to a delay in response to the request.

Since facial features are subject to variation over time, regular updating of the models is required. However, since no subject's data is stored in the system, the model can update this data periodically. Indeed, if the authentication is successful, the system can use the acquired and validated images, which will only be kept for this process's duration. These images are then used to perform a new learning process, replacing the old model on the card.

4.9 Discussion of the objectives

The research proposes an efficient multimodal biometric system based on the face and the iris traits. Two contributions were proposed in this work: one in machine learning and the other in deep learning.

One of this work's objectives is to test iris and face recognition in different scenarios (for example, a person wearing a mask). This study is performed during the fusion of the modalities using the automatic matching approach. Indeed, the dataset used for unauthorized persons contained images of people wearing a mask of an authorized person's face, and the system has well classified the image.

4.10 Summary

The idea of this work in machine learning-based recognition is to propose a fast and efficient solution using a single algorithm for multimodal face and iris recognition. Furthermore, an experimental approach is adopted based on extracting the face and iris features using DWT combined with SVD. The results obtained on the IIT Delhi iris and Face94 datasets are auspicious.

Multimodal recognition using classical machine learning techniques such as DWT and SVD has given satisfactory results. However, these results are correlated to the type of data used. Indeed, the images used for iris recognition are the same ones used in deep learning, but those of the face are much simpler. The Face94 dataset contains images that have a plain green background. This will allow the model to focus on the relevant information and thus improve its results. The DWT+SVD combination was tested on the VGGFace database's images, and the results were not promising.

For deep learning, it has been shown that Transfer Learning, applied to pre-trained networks, can be a very efficient training method. It allows a network dedicated to a specific task to converge towards high classification performances for a new and more specific task. Indeed, it has been found that all trained models gave good classification performances. On the other hand, the performances drop considerably in the case of recognition (authorized/unauthorized person). This is because the choice of the threshold plays a significant role in the decision.

The results presented in this chapter have shown interest in deep learning approaches in face and iris recognition. In particular, the latest work on extracting iris and face features by DenseNet121 and FaceNet models and their fusion in an approach based on feature-level fusion opens exciting perspectives. Indeed, combining the strengths of feature extraction using deep architectures and automatic matching using multilayer perceptron gives promising results.

The results also highlighted the interest in the proposed automatic matching method compared to the standard threshold selection according to the EER point. Finally, the analysis of the results concluded that deep learning gives better results than machine learning and that it was the most adapted to the problem of this thesis; thus, it can be said that the outlined objectives have been achieved in this thesis.

Chapter 5

Case Studies, Experimental Results and Discussions

5.1 Introduction

To thrive, every company needs to provide a safe and welcoming work environment. However, finding a balance between security and freedom of movement is not always easy. An environment that is too restrictive is a hindrance to the movement of people. On the other hand, if it is too permissive, your safety is compromised.

Access control systems play an essential role in any sensitive area, allowing only the right people to enter at any given time. Multiple configurations cater to various budgets and needs, ranging from the simplest card reader to the now commonly used fingerprint terminals to the facial and iris recognition terminal.

As customer requirements become more complex, one solution can no longer meet all needs. That is why one can combine different authentication methods in access

control systems, such as ID cards, fingerprints, passwords, facial images, and QR codes, to name a few.

The use of facial and iris recognition in access control and attendance has been an inevitable trend, creating a beneficial "touchless" experience. Furthermore, with advanced Deep Learning technology and convenient features, facial and iris recognition terminals provide better security and improved access control and attendance tracking efficiency.

This section aims to highlight the importance of access control in improving the security of people, buildings, and properties. Two scenarios will be presented in which the proposed solution in the previous chapters will be tested. Within the framework of this study, all the controls are carried out with the formal authorization of the persons.

5.2 Used hardware

To use the proposed solution, two sensors are needed, one for acquiring the iris image and the other for the face. The IriTech IriShield USB MK 2120U camera is used for the iris, which is a very compact model powered by a USB cable. This sensor has been chosen since it has low power consumption and is a cross-platform model (usable under any operating system). It uses an infrared LED to illuminate the eye; thus, irises can be captured in various indoor and outdoor environments. In addition, the captured iris images comply with the ISO/IEC 19794-6 standard.

Figure 5.1: The IriTech IriShield USB MK 2120U camera

It is possible to use any camera to capture images of the face. The Logitech C920s PRO HD webcam, which offers good quality, is used for experiments.

Figure 5.2: The Logitech C920s PRO HD webcam

5.3 Solution setup

5.3.1 Environment

The implementation of a biometric access control system must consider elements specific to the human factor for the controls to work effectively. In particular, the following elements should be considered:

- Devices common to an entire population.
- Sensor: hygiene issues.
- Duration of the control.

5.3.2 Process

The macroscopic cycle of a biometric identification process can be broken down into two main stages: enrollment and control.

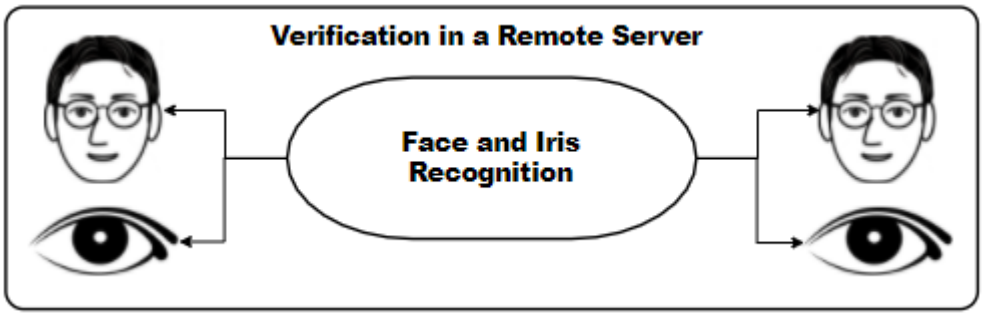
5.3.2.1 Enrollment

Person enrollment is the initial phase of creating the biometric template and storing it in conjunction with a declared identity. The physical characteristics (face and iris images) are transformed into a template representative of the person and specific to the recognition system. Additional data specific to the enrollee are recorded during this phase, such as first and last name and a personal identification number (PIN). This step is performed only once.

5.3.2.2 Control (verification / identification)

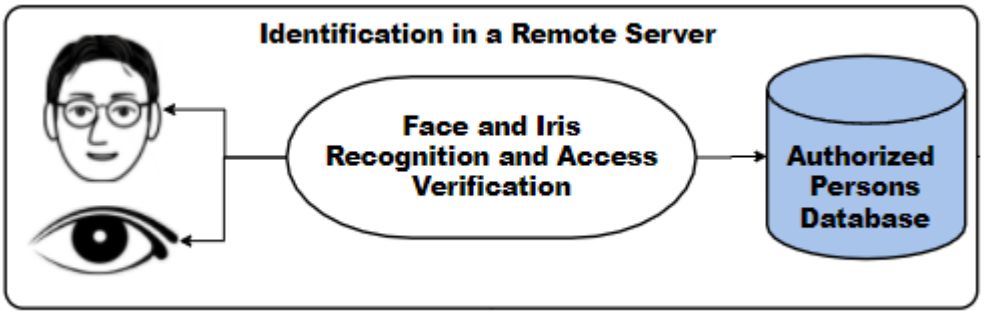
This is the action of checking a person's data to proceed with the verification of their declared identity or, in an investigation, to find out that person's identity. This step takes place every time a person comes to the system.
Verification consists of confirming the claimed identity of a person (authenticate) by checking the images of his face and iris. It is a "one-to-one" comparison in which the two biometric templates (face and iris) captured are compared to corresponding reference templates in a database.

Figure 5.3: Verification process



Identification consists of identifying a person using their physical characteristics within a previously registered population. It is a "one-to-many" comparison in which the biometric template entered is compared to all templates stored in a database.

Figure 5.4: Identification process



In addition to these two processes, the following two processes are often added:

- Refreshing or updating: the biometric system can periodically correct the reference template during a check in order to take into account changes in the person's data;
- End of life: the template and other reference data specific to the person are destroyed to take into account the person's removal from the control system.

5.4 Scenarios

Each biometric system uses specificities linked to the physical characteristic analyzed (fingerprint, iris, hand shape, etc.) and the system technology.
In order to study the feasibility of the solution and test it in the real world, it was deployed in two structures with different requirements: the first scenario in a pharmacy of a medical structure and the second in a military structure.

5.4.1 Controlled medication storage areas access

This experiment was performed in the central pharmacy of Al Asria Eye Clinic in Dubai. This pharmacy contains a compartment accessible by all pharmacists and trainees and a restricted area containing psychotropic drugs and dangerous substances.
First, the people authorized to access the restricted area were identified. Then, this specific area was secured using an electric lock which will be released if the person is authorized to visit this area.

5.4.1.1 Collection of identification data

This is the step of entering the person's identification data and, in particular, his biometric characteristics. The data entered are:

- The data on the claimed identity (name, identifier, etc.).
- Personal biometric data (iris and face). The collection of physical data is done through specialized sensors corresponding to the analyzed characteristic (IriTech IriShield sensor for the iris and Logitech C920s PRO HD Webcam for the face).

In this first use case, the proposed approach is tested in the central pharmacy of the Al Asria Eye Clinic. This structure contains two rooms: the first one contains all the drugs except psychotropic drugs and psycholeptics, which are stored in the second room, considered a restricted area. The clinic has eight (08) practitioners with the right to access the first room: five regulars and three interns. The latter are not allowed access to the restricted area. The collected dataset is shown in Figure 5.5.

Figure 5.5: The collected dataset in the Al Asria Eye Clinic

5.4.1.2 Transformation into a biometric template

The iris and facial feature sensors transmit the captured data to an analysis system whose role is to transform them into a template, according to the Deep Learning algorithm proposed in the previous section.

5.4.1.3 Comparison to a reference

The calculated template must then be compared to the reference template to identify the person attempting to enter the room. The comparison involves the score

calculation that allows the comparison to be considered successful or unsuccessful.

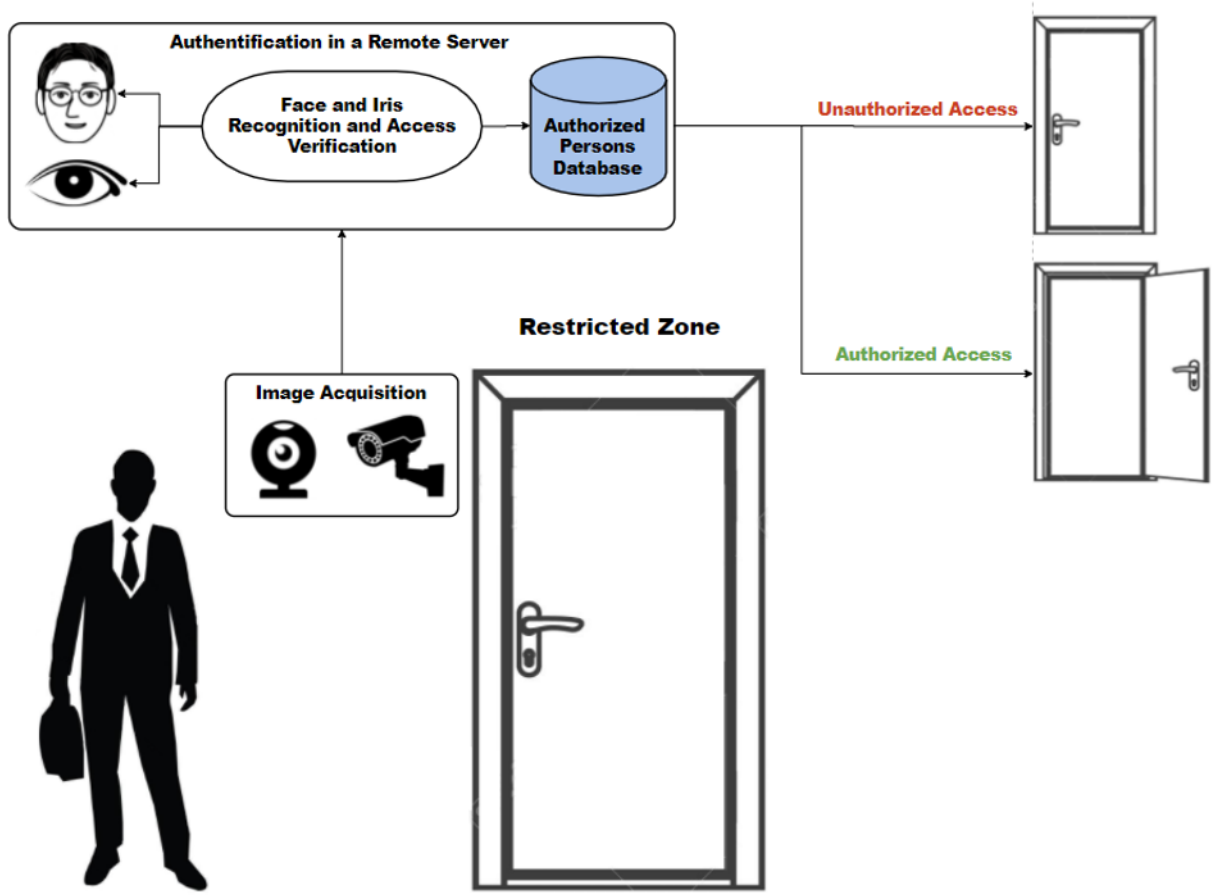
5.4.1.4 Decision making

The decision subsystem receives the result of the calculated matching score to the stored templates. Then, depending on the distance, the decision subsystem decides what actions to take (see Figure 5.6. The decision subsystem can consider the identification:

- In success and return a positive response to the user application system and open the door;
- In failure and give a negative answer to the user application system;

In case of failure, the system can offer the possibility of restarting the process at the collection stage or counting the number of failed checks of the same person and deciding to block the check for the considered person. The answer is returned to the application system that uses the biometric system.

Figure 5.6: The identification process



5.4.1.5 Results

This solution was tested on eight practitioners who had access to the first room and only five who had access to the restricted area (Figure5.7). The system allowed a faultless recognition of the eight persons for access to the first room and the five persons for the access to the restricted area (Table 5.1). The system also did not commit any error in rejecting unauthorized persons. However, when the staff registration was made without protective material (mask, visor, uniform), but they showed up with this equipment, the model did not recognize them and considered them intruders.

Figure 5.7: Architecture of the pharmacy

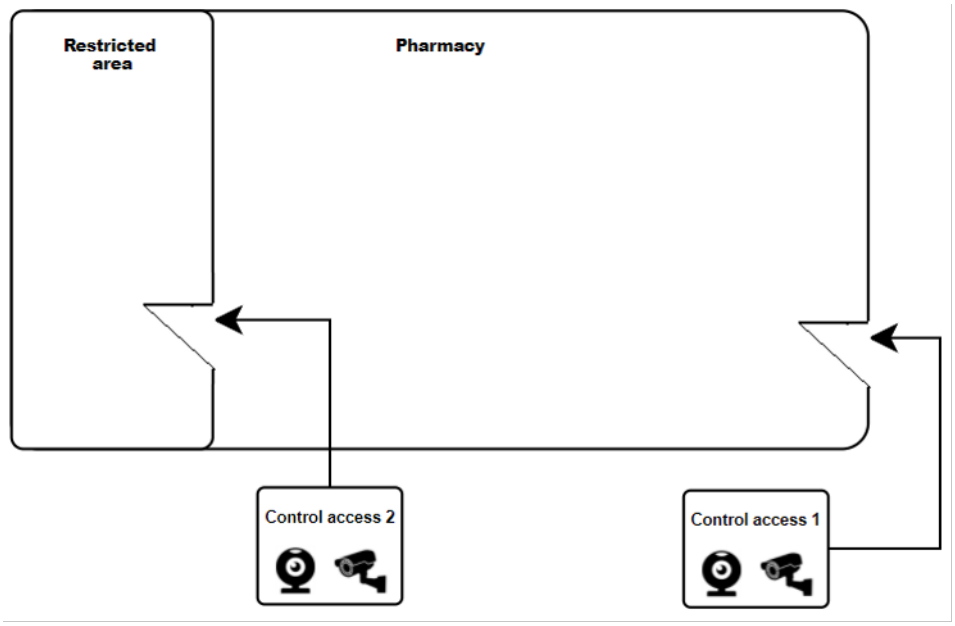


Table 5.1: The obtained results at the pharmacy

	Experiment 1	Experiment 2
Gadget on face	No	Yes, mask and visor
Success rate (%)	100	0

5.4.2 Authentication in a restricted military zone

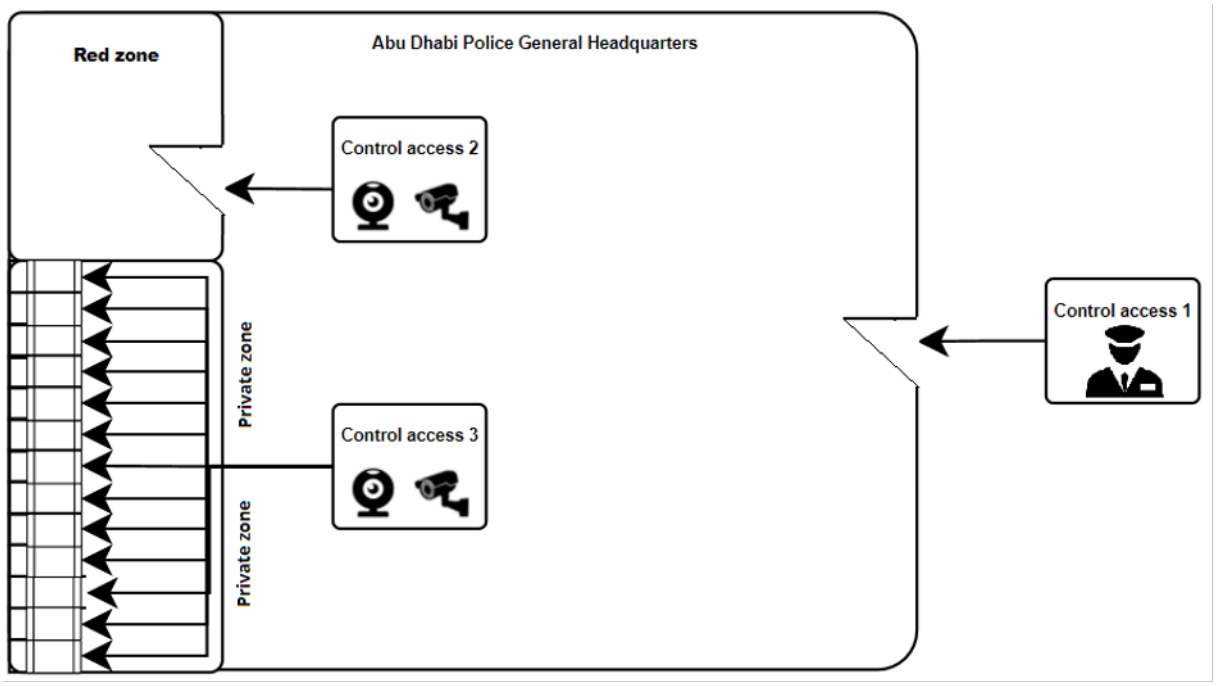
The second scenario concerns identification and verification tasks in a military structure. The study was conducted on the access control of the red zone and private areas at the Abu Dhabi Police General Headquarters (Police GHQ).

Since the GHQ police building house more than 600 people, the proposed security system has three levels (see figure). Security guards at the main entrance conduct the first security check. The purpose of the controllers is to orient visitors and check their identities.

As shown in Figure 5.8, the security system deployed in this structure has two objectives. First, control access to the red zone, which contains the weapons, by allowing only people who have the right to access it (identification task). Second, control the private zone of the police officers by giving them access only to their private space

(verification task).

Figure 5.8: Architecture of the military structure



5.4.2.1 Collection of identification data

This study was conducted on twenty police officers. Images were collected from the faces and irises of these people using the same sensors used in the clinic's pharmacy. The collected images are shown in Figure 5.9.

Figure 5.9: The collected dataset in the Abu Dhabi Police General Headquarters

5.4.2.2 Transformation into a biometric template

This step is conducted in the same way as in the clinic's pharmacy.

5.4.2.3 Comparison to a reference

Again, the same approach of computing similarity between feature vectors is used here.

5.4.2.4 Decision making

The decision depends on the target task; if it is an identification (access to the red zone), the subsystem responds with identification or not, and action on the door (opening or not). On the other hand, suppose it is an identity verification (access to the private area). In that case, the subsystem checks whether it is the person who owns the area in question and grants access or not, depending on the result.

5.4.2.5 Results

To extend our experiments, we performed acquisitions with different cameras to evaluate the impact of the sensor on transmission speed, processing speed, and performance. We used the IriShield sensor, the Logitech webcam, and a smartphone camera (Samsung Galaxy S20) for the iris. We used the Logitech webcam and the smartphone's camera for the face.

Table 5.2 summarizes the obtained results. The first observation is that the performances of the verification and identification are very satisfactory using the IriTech IriShield USB MK 2120U camera and the Logitech Webcam sensors. Out of the twenty police officers, no errors were recorded in both task. However, if a smartphone camera or the Logitech is used instead of IriShield, the iris is not recognized, and thus the fusion gives awful results.

On the other hand, the image transmission time using the Logitech camera or the IriShield sensor is faster than the smartphone. This can be explained by the fact that the size of the image obtained using the IriShield is 640 x 480 pixels, and that of the Logitech is 970x728 pixels, while that obtained by the smartphone is 1920x1080.

Table 5.2: The obtained results at the Police GHQ

	Experiment 1	Experiment 2	Experiment 3
Face Camera	Logitech	Smartphone	Logitech
Iris Camera	IriShield	Smartphone	Logitech
Face image transmission time	∼2 s	∼5 s	∼2 s
Iris image transmission time	∼1.5 s	∼5 s	∼2 s
Success rate (%)	100	0	0
Process time	∼1.2 s	∼1.2 s	∼1.2 s

5.4.3 Advantages of the proposed solution

The main advantage of this solution is that it allows quick identification of the persons authorized to access the different rooms and thus protects the pharmacy against

any abuse of medication use and the red zone in the military zone against the unauthorized use of weapons.

This system also achieved 100% recognition rates in the identification task (access to the first room of the pharmacy by the eight practitioners, access to the restricted area by the five practitioners, and access to the red zone by the twenty police officers) and the verification task (access by each of the twenty police officers to their private area).

The model also showed great robustness to light variations (see Figure 5.10). Images were tested on different light intensities, and the results were good. This is because when training the model, data augmentation methods were used by varying the illumination of the images.

Figure 5.10: Variation of the brightness of the images

5.4.4 Limitations of the proposed solution

During the development and deployment of the solution, several problems related to the data and the deployment environments were encountered.

Indeed, despite the advantages of biometric systems over traditional systems, their

use is still limited to specific applications (such as biometric passports, automatic time registration systems, and access control systems). Moreover, these systems suffer from several limitations that can considerably degrade their interest.

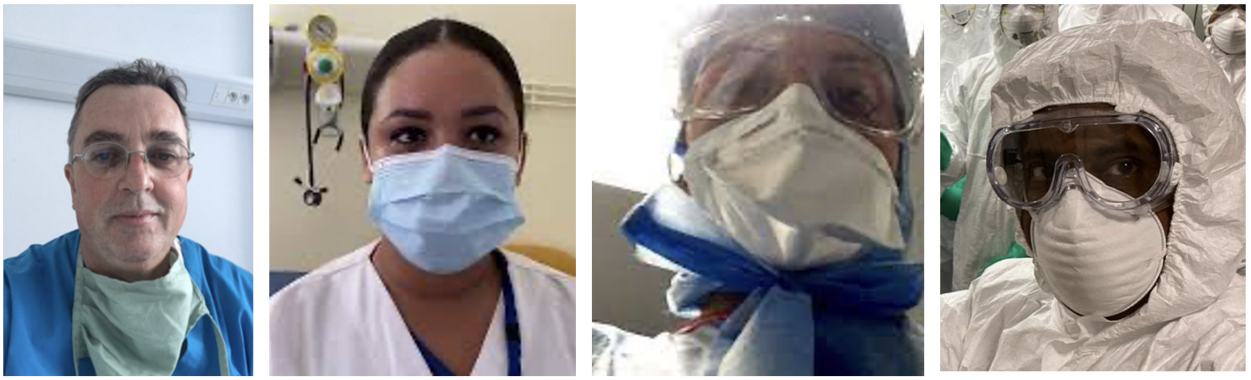
5.4.4.1 Limitations related to data

Despite the excellent performance and efficiency of the proposed solutions, they suffer from some limitations related mainly to the data. For example, unlike traditional authentication systems, biometric-based authentication systems can become less accurate due to several factors: variability during capture (i.e., acquisition noise, use of multiple acquisition sensors, etc.), intra-class variability (i.e., variability of biometric data for an individual), and inter-class similarity (i.e., the similarity of biometric data for multiple individuals). Another limitation of biometrics is the limitation of use. Depending on the modality used, biometric data acquisition is performed with or without contacting the biometric sensor (in the case of the iris and the face, respectively). This contact is a concern for some users of hygiene and physical intrusion.

a. **Limitations noticed in the first scenario**

   The variation between the enrolled and captured images during the authentication has misled our system. Indeed, to add them to the database, images of practitioners were taken at rest without visors, goggles, and masks. However, while working, they wore these safety measures (see Figure 5.11), which made the recognition task very difficult or impossible. As a result, the system's performance decreased considerably since the personnel were not recognized and were not allowed to access the room. To solve this problem, the recognition model should be trained on images of people wearing masks.

Figure 5.11: From left to right, a doctor, a doctor wearing a mask, a doctor wearing a mask and classes, a doctor wearing a full protection kit



Another problem has been identified in this model. The solution has two separate identification systems (control access 1 and 2) and thus two processing servers. However, to solve this problem, it would be possible to use a single server with access rights management that gives access to the first room for the eight practitioners and access to the second room only for the five permanent practitioners.

b. **Limitations noticed in the second scenario**

In this system, a test was made by changing the iris sensor with an ordinary camera (see figure 5.12). Unfortunately, the system could not identify the persons because it did not learn from the same images. To remedy this problem, the model should either be retrained on images taken by an ordinary camera or use iris images captured with a sensor similar to the IriTech IriShield USB MK 2120U camera.

Figure 5.12: Iris image captured using a phone camera

In the same modality, the recognition is tested on people wearing colored lenses. Again, the system did not recognize these people and considered them as intruders.

5.4.4.2 Limitations related to the deployment environment

Despite the success of the experiments conducted in both structures (the clinic and the police), the framework was confronted with several obstacles and problems related to several factors, such as:

a. **Accessibility**

This is the first problem faced when accessing the red zone of the police headquarters. In order to enter this zone, it was necessary to submit a request to the chief of police and wait for his return, which sometimes took weeks. Moreover, after the signature of a confidentiality contract, the authorization access was not permanent but limited in time, to single-use, at a precise date, and conditioned by the presence of a staff designated by the director. Therefore, a new request had to be submitted to reaccess the site.

It should also be noted that even though the authorization is granted, it was not permitted to access the whole structure. It was not allowed to connect the solution to the local network or make any modifications to the installations, walls, or doors. So, USB cables were used to connect the cameras to the server.

b. **Staff availability**

The second challenge is the continuous presence of staff. Because of the COVID-19 health crisis, the medical staff was always busy, and it was tough to find free time to test the recognition.

For the police, another type of problem was encountered; the staff was working in rotation with a hybrid system (online/office). Since the authorized presence was limited in time, meeting with the people who agreed and were allowed to

participate in this study was complicated.

c. **Domain experience**

When designing the solutions, no prior knowledge of the chosen domains (healthcare and military) was available, so the first step was understanding each structure's needs. Specifying requirements was challenging and slow given the unavailability of experts and contacts in the clinic and police headquarters.

d. **Funding**

Training the models, purchasing the cameras and various computer components, and deploying the solutions imply cost. To perform the training of the Deep Learning architectures, the personal computer in our possession and the free cloud resources like Google Colab were not sufficient, and renting GPU resources to carry out this task was the only solution. Unfortunately, the university or the host structures (the clinic and the police) did not offer any funding.

5.5 Summary

This chapter deployed the deep learning-based identification/verification solution in two different structures (Al Asria Eye Clinic and the Abu Dhabi Police General Headquarters). The results were very satisfactory when the images' quality and nature were similar to those used during the training.

Our solution suffers from some limitations when there is a significant variation in the data (using another sensor that gives very different images from the originals, using colored lenses, masking a large part of the face, etc.).

This system can be improved by adding other images during training and by using an access management application on its side.

Chapter 6

Conclusions, Recommendations and Future work

6.1 Conclusions

The proposed work in this thesis focused on using machine and deep learning for multimodal biometrics identification. The definition of the context of the thesis's subject was first given by introducing basic notions about biometric systems in general and the tools to evaluate them.

Then, the different aspects of multimodal biometric systems through their architecture, information sources, and fusion levels were given. After that, a bibliographic study on multimodal systems was also performed.

Finally, the proposed frameworks using machine and deep learning were explained, and the experiments on benchmarks and two case studies were given.

The difficulty in extracting facial and iris features is mainly due to variations in lighting conditions and the difference between the images acquired during person registration and access control. For this reason, we proposed two different platforms,

one based on machine learning, where the user manages the characterization step, and the second one is automatic, where the deep learning model performs the characterization.
The fusion phase combines the face and iris results after the feature extraction stage. Given the many multimodal problems and the many multimodal fusion methods associated, it is particularly difficult to select an efficient fusion strategy when faced with a new multimodal problem. Therefore, different fusion techniques have been tested to adopt the best strategy for our problem.

The first part of this thesis presents a new recognition method using machine learning based on the face and the iris. The advantage of this framework is its simplicity since it uses a unique face and iris feature extraction method. This work's contribution consists of the proposal of a face and iris image processing pipeline for multimodal recognition. The images are preprocessed using the contrast enhancement, and then each image is divided into 128 blocks, each of which is characterized using DWT. A two-level wavelet decomposition carries out DWT extraction by the Haar window. The SVD has been applied on low-frequency LL sub-bands of the input image to extract the singular values. Once the singular values are calculated using the SVD formula, the result will be appended at the end of the corresponding feature vector. The decision-level and fusion-level fusion matching are performed during the recognition process using the Euclidean distance. From the obtained results from experiments and in comparison with state of the art, it can be concluded that:

- techniques using Euclidean distance give better results than techniques using a Support vector machine (SVM)
- Fuzzy k-nearest neighbor (FK-NN) also gives good results, which means that, for the matching, techniques based on distance give better results than techniques based on classical machine learning.

In the second part of the thesis, a feature-level fusion of the multimodal iris and face recognition system is presented.

Ten different architectures are tested: VGG16, Resnet50, DensenNet121, MobileNet, and InceptionV3 architectures are used for iris recognition. In addition, the VGGFace, FaceNet, InceptionV3, Resnet50, and OpenFace architectures are used for the face. These models were trained on the ImageNet dataset and then finetuned using our datasets. We have selected the best models from the experiment results: DenseNet121 for the iris and FaceNet for the face. These architectures will be used for extracting features from images. The training set results will be stored in a feature base that will be used later to find the matching.

Then, a new automatic matching method based on the multilayer perceptron is proposed. This technique classifies a person as being authorized or not. If the individual belongs to the authorized class, it will be assigned to the person's identity closest to it in the features database. Otherwise, it will be treated as unauthorized.

Results are very promising, and the following conclusions can be drawn:

- A general observation is that the DenseNet-121 model and the FaceNEt model learned well from the data to perform their respective tasks
- the proposed model is insensitive to light variation as we have used data augmentation techniques that allow varying the size and brightness of the training images
- feature-level fusion is more effective than the decision-level fusion
- transfer learning using pre-trained models on large datasets such as imagenet allows the model to be more robust and to generalize better
- the automatic matching gives better results than non-automatic where the user should specify a threshold that represents the best compromise between the Accuracy, FAR, and FRR.

To validate the second platform based on deep learning, we deployed it in two different institutions: Al-Asria Eye Clinic's central pharmacy in Dubai and the Abu Dhabi Police General Headquarters (Police GHQ).
The main outcome and advantage of this platform solution are that it allows quick identification of the persons authorized to access the different rooms and thus protects the pharmacy against any abuse of medication use and the red zone in the military zone against the unauthorized use of weapons. The results were very satisfactory, and the platform provided good robustness under normal environmental conditions.

6.2 Research Limitations and Recommendations

The works carried out within the framework of this thesis have proved their effectiveness for multimodal recognition. However, since no model can claim to be perfect, as such, the approaches presented in this study have limitations, some of which are given below:

- The facial dataset used by traditional machine learning methods is simpler than that used with machine learning because it contains a solid green background. On the other hand, the dataset used by deep learning is more complex and contains more data. Unfortunately, only a subset of this dataset has been used due to limited computing resources.
- It was challenging to choose the most appropriate architecture for this problem during this study. The choices were made based on the most used techniques in state of the art.
- Although the deep learning approach performs well under normal conditions, it suffers when another sensor captures the iris image or when a mask or protective equipment covers the face image during identification/verification.

6.3 Future Work

This thesis opens the way to new research. In perspective, it will be interesting to consider exploring other biometric signatures.

The second perspective concerns testing and validating the proposed approaches' effectiveness using larger datasets containing more complex images.

The third line of research concerns AutoML. Indeed, there are many AI, Machine Learning, and deep learning algorithms; therefore, many options make choosing a suitable algorithm for a powerful and robust biometric system challenging. AutoML intelligently searches for the ideal combination of data processing, algorithm, and parameters to select the most optimal option among this infinite number of choices. This selection process is based, among other things, on genetic algorithms based on the old concept of "survival of the fittest".

Finally, while deep learning can achieve excellent results in practice, end-users ' interpretability remains a significant issue. The representations learned by these statistical models are complex for humans to use and only convey limited information.

Interpretability of models to enable specialists to understand model predictions is a prerequisite for collaboration between end-users and machines. On the one hand, this requires associating the network's representations with semantic concepts that humans can manipulate. On the other hand, it makes the decision-making process explainable by transparentizing the most influential factors.

References

A. A., D. C.M., and D. P.K. Wavelet fundamentals. In M. Birkhäuser, Boston, editor, *Wavelets and Subbands*. Applied and Numerical Harmonic Analysis, 2002. ISBN 978-1-4612-6618-1. URL https://doi.org/10.1007/978-1-4612-0113-7_2.

C. C. Aggarwal, A. Hinneburg, and D. A. Keim. On the surprising behavior of distance metrics in high dimensional space. In J. Van den Bussche and V. Vianu, editors, *Database Theory — ICDT 2001*, pages 420–434, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg. ISBN 978-3-540-44503-6.

N. Ahmed, T. Natarajan, and K. R. Rao. Discrete cosine transform. *IEEE Transactions on Computers*, C-23(1):90–93, 1974.

P. Akulwar and N. A. Vijapur. Secured multi modal biometric system : A review. In *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, pages 396–403, 2019. doi: 10.1109/I-SMAC47947.2019.9032628.

A. Al-Waisy, R. Qahwaji, and et. al. A multi-biometric iris recognition system based on a deep learning approach. *Pattern Anal Applic*, 21:783–802, 2018. doi: https://doi.org/10.1007/s10044-017-0656-1.

A. S. Al-Waisy, R. Qahwaji, S. Ipson, and S. Al-Fahdawi. A multimodal biometrie system for personal identification based on deep learning approaches. In *2017 Seventh International Conference on Emerging Security Technologies (EST)*, pages 163–168, 2017a. doi: https://doi.org/10.1109/EST.2017.8090417.

A. S. Al-Waisy, R. Qahwaji, S. Ipson, and S. Al-Fahdawi. A multimodal biometrie system for personal identification based on deep learning approaches. In *2017 Seventh International Conference on Emerging Security Technologies (EST)*, pages 163–168, 2017b. doi: 10.1109/EST.2017.8090417.

S. Alam. Cybersecurity: Past, present and future, 2022. URL https://arxiv.org/abs/2207.01227.

N. Alay and H. H. Al-Baity. Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits. *Sensors*, 20 (19), 2020a. ISSN 1424-8220. doi: https://doi.org/10.3390/s20195523. URL https://www.mdpi.com/1424-8220/20/19/5523.

N. Alay and H. H. Al-Baity. Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits. *Sensors*, 20(19), 2020b. ISSN 1424-8220. doi: 10.3390/s20195523. URL https://www.mdpi.com/1424-8220/20/19/5523.

F. Altenberger and C. Lenz. A non-technical survey on deep convolutional neural network architectures. *CoRR*, abs/1803.02129, 2018. URL http://arxiv.org/abs/1803.02129.

B. Ammour, T. Bouden, and L. Boubchir. Face-iris multimodal biometric system based on hybrid level fusion. In *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*, pages 1–5, 2018.

B. Ammour, L. Boubchir, T. Bouden, and M. Ramdani. Face–iris multimodal biometric identification system. *Electronics*, 9(1):85, Jan 2020. ISSN 2079-9292. doi: https://doi.org/10.3390/electronics9010085. URL http://dx.doi.org/10.3390/electronics9010085.

B. Amos, B. Ludwiczuk, and M. Satyanarayanan. Openface: A general-purpose face recognition library with mobile applications. 2016.

J. Anil K., R. Arun A., and N. Karthik. *Introduction to Biometrics*. Springer, Boston, MA, 2011. ISBN 978-0-387-77325-4. URL https://doi.org/10.1007/978-0-387-77326-1.

S. Arora, M. P. S. Bhatia, and H. Kukreja. A multimodal biometric system for secure user identification based on deep learning. In X.-S. Yang, R. S. Sherratt, N. Dey, and A. Joshi, editors, *Proceedings of Fifth International Congress on Information and Communication Technology*, pages 95–103, Singapore, 2021. Springer Singapore. ISBN 978-981-15-5856-6.

Y. Bouzouina and L. Hamami. Multimodal biometric: Iris and face recognition based on feature selection of iris with ga and scores level fusion with svm. In *2017 2nd International Conference on Bio-engineering for Smart Technologies (BioSMART)*, pages 1–7, 2017.

R. Brunelli and D. Falavigna. Person identification using multiple cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10):955–966, 1995.

S. C., Z. H., D. Feng, and e. al. Survey of information security. *Science in China Series F: Information Sciences*, 50:273–298, 2007. URL https://doi.org/10.1007/s11432-007-0037-2.

D. Campo, O. L. Quintero, and M. Bastidas. Multiresolution analysis (discrete wavelet transform) through daubechies family for emotion recognition in speech. *Journal of Physics: Conference Series*, 705:012034, apr 2016. doi: 10.1088/1742-6596/705/1/012034. URL https://doi.org/10.1088/1742-6596/705/1/012034.

W. Cathy. How biometrics keep sensitive information secure. *Nursing*, 35(5):76, 2005.

K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In *Proceedings of the British Machine Vision Conference*. BMVA Press, 2014.

K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *CoRR*, abs/1409.1259, 2014. URL http://arxiv.org/abs/1409.1259.

F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1800–1807, 2017. doi: 10.1109/CVPR.2017.195.

N. Damer. *Application-driven Advances in Multi-biometric Fusion*. PhD thesis, 04 2018.

S. Dargan and M. Kumar. A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities. *Expert Systems with Applications*, 143:113114, 2020. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2019.113114. URL https://www.sciencedirect.com/science/article/pii/S0957417419308310.

R. Dechter. Learning while searching in constraint-satisfaction problems. In *AAAI-86 Proceedings*, pages 178–183, 1986.

M. Elhoseny, E. Essa, A. Elkhateb, A. E. Hassanien, and A. Hamad. Cascade multi-modal biometric system using fingerprint and iris patterns. In A. E. Hassanien, K. Shaalan, T. Gaber, and M. F. Tolba, editors, *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2017*, pages 590–599, Cham, 2018. Springer International Publishing. ISBN 978-3-319-64861-3.

M. Eskandari. Optimum scheme selection for face–iris biometric. *IET Biometrics*, 6:334–341(7), September 2017. ISSN 2047-4938. URL https://digital-library.theiet.org/content/journals/10.1049/iet-bmt.2016.0060.

M. Eskandari and Önsen Toygar. Selection of optimized features and weights on face-iris fusion using distance images. *Computer Vision and Image Understanding*, 137:63 – 75, 2015. ISSN 1077-3142. doi: https://doi.org/10.1016/

j.cviu.2015.02.011. URL http://www.sciencedirect.com/science/article/pii/S1077314215000454.

G. Feng, K. Dong, D. Hu, and D. Zhang. When faces are combined with palmprints: A novel biometric fusion strategy. In D. Zhang and A. K. Jain, editors, *Biometric Authentication*, pages 701–707, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg. ISBN 978-3-540-25948-0.

J. Fierrez-Aguilar. *Adapted Fusion Schemes for Multimodal Biometric Authentication. University Polytecnique of Madrid*. PhD thesis, 2006.

T. L. Fine, S. L. Lauritzen, M. Jordan, J. Lawless, and V. Nair. *Feedforward Neural Network Methodology*. Springer-Verlag, Berlin, Heidelberg, 1st edition, 1999. ISBN 0387987452.

Y. Fu, G. Guo, and T. S. Huang. Age synthesis and estimation via faces: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):1955–1976, 2010. doi: 10.1109/TPAMI.2010.36.

B. Furht, editor. *Discrete Wavelet Transform (DWT)*, pages 188–188. Springer US, Boston, MA, 2008. ISBN 978-0-387-78414-4. doi: https://doi.org/10.1007/978-0-387-78414-4_305. URL https://doi.org/10.1007/978-0-387-78414-4_305.

L. Guarnera, O. Giudice, F. Guarnera, A. Ortis, G. Puglisi, A. Paratore, L. M. Q. Bui, M. Fontani, D. A. Coccomini, R. Caldelli, F. Falchi, C. Gennaro, N. Messina, G. Amato, G. Perelli, S. Concas, C. Cuccu, G. Orrù, G. L. Marcialis, and S. Battiato. The face deepfake detection challenge. *Journal of Imaging*, 8 (10), 2022. ISSN 2313-433X. doi: 10.3390/jimaging8100263. URL https://www.mdpi.com/2313-433X/8/10/263.

S. Guennouni, A. Mansouri, and A. Ahaitouf. Biometric systems and their applications. In G. L. Giudice and A. Catalá, editors, *Visual Impairment and Blindness*,

chapter 20. IntechOpen, Rijeka, 2020. doi: 10.5772/intechopen.84845. URL https://doi.org/10.5772/intechopen.84845.

K. Gunasekaran, J. Raja, and R. Pitchai. Deep multimodal biometric recognition using contourlet derivative weighted rank fusion with human face, fingerprint and iris images. *Automatika*, 60(3):253–265, 2019. doi: https://doi.org/10.1080/00051144.2019.1565681. URL https://doi.org/10.1080/00051144.2019.1565681.

N. Gómez Blas, L. F. de Mingo López, A. Arteta Albert, and J. Martínez Llamas. Image classification with convolutional neural networks using gulf of maine humpback whale catalog. *Electronics*, 9(5), 2020. ISSN 2079-9292. doi: 10.3390/electronics9050731. URL https://www.mdpi.com/2079-9292/9/5/731.

M. Hardt, B. Recht, and Y. Singer. Train faster, generalize better: Stability of stochastic gradient descent. In *International Conference on Machine Learning*, pages 1225–1234, 2016.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90.

S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9 (8):1735–1780, 11 1997. ISSN 0899-7667. doi: 10.1162/neco.1997.9.8.1735. URL https://doi.org/10.1162/neco.1997.9.8.1735.

D. Hond and L. Spacek. Distinctive descriptions for face processing. In A. F. Clark, editor, *BMVC*. British Machine Vision Association, 1997. ISBN 0-952-18987-9. URL http://dblp.uni-trier.de/db/conf/bmvc/bmvc1997.html#HondS97.

P. C. Hough V. Method and means for recognizing complex patterns, U.S. Patent 3069654A. 1962. URL https://www.freepatentsonline.com/3069654.html.

G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger. Deep networks with stochastic depth. In B. Leibe, J. Matas, N. Sebe, and M. Welling, editors, *Computer Vision – ECCV 2016*, pages 646–661, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46493-0.

G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, 2017. doi: 10.1109/CVPR.2017.243.

G. Huo, Y. Liu, X. Zhu, H. Dong, and F. He. Face–iris multimodal biometric scheme based on feature level fusion. *Journal of Electronic Imaging*, 24(6): 1 – 10, 2015. doi: https://doi.org/10.1117/1.JEI.24.6.063020. URL https://doi.org/10.1117/1.JEI.24.6.063020.

S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In F. Bach and D. Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR. URL http://proceedings.mlr.press/v37/ioffe15.html.

A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270 – 2285, 2005a. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2005.01.012. URL http://www.sciencedirect.com/science/article/pii/S0031320305000592.

A. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, 2005b. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2005.01.012. URL https://www.sciencedirect.com/science/article/pii/S0031320305000592.

A. K. Jain and A. Ross. Multibiometric systems. *Commun. ACM*, 47(1):34–40,

Jan. 2004. ISSN 0001-0782. doi: https://doi.org/10.1145/962081.962102. URL https://doi.org/10.1145/962081.962102.

A. K. Jain, R. Bolle, and S. Pankanti. *Biometrics*. Springer US, 2006. ISBN 978-0-387-28539-9. doi: https://doi.org/10.1007/978-0-387-32659-7. URL https://www.springer.com/gp/book/9780387285399.

C. Jamdar and A. Boke. Review paper on person identification system using multi-model biometric based on face. *Int. J. Sci. Eng. Technol. Res.*, 6(3):626–629, 2017.

Jiancheng Zou, R. K. Ward, and Dongxu Qi. A new digital image scrambling method based on fibonacci numbers. In *2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No.04CH37512)*, volume 3, pages III–965, 2004. doi: https://doi.org/10.1109/ISCAS.2004.1328909.

W. Kabir, M. O. Ahmad, and M. N. S. Swamy. Normalization and weighting techniques based on genuine-impostor score fusion in multi-biometric systems. *IEEE Transactions on Information Forensics and Security*, 13(8):1989–2000, 2018.

Z. Kalal, K. Mikolajczyk, and J. Matas. Face-tld: Tracking-learning-detection applied to faces. In *2010 IEEE International Conference on Image Processing*, pages 3789–3792, 2010. doi: 10.1109/ICIP.2010.5653525.

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. *Sensors*, 18(7), 2018. ISSN 1424-8220. doi: https://doi.org/10.3390/s18072296. URL https://www.mdpi.com/1424-8220/18/7/2296.

D. E. King. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.*, 10:1755–1758, Dec. 2009. ISSN 1532-4435.

J. Kittler and K. Messer. Fusion of multiple experts in multimodal biometric personal identity verification systems. In *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, pages 3–12, 2002.

A. Krenker, J. Bester, and A. Kos. Introduction to the artificial neural networks. In K. Suzuki, editor, *Artificial Neural Networks*, chapter 1. IntechOpen, Rijeka, 2011. doi: 10.5772/15751. URL https://doi.org/10.5772/15751.

A. Krizhevsky. Learning multiple layers of features from tiny images. Technical report, 2009.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

A. Kumar and A. Passi. Comparison and combination of iris matchers for reliable personal authentication. *Pattern Recognition*, 43(3):1016–1026, 2010.

N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *2009 IEEE 12th International Conference on Computer Vision*, pages 365–372, 2009. doi: 10.1109/ICCV.2009.5459250.

O. C. Kurban, T. Yildirim, and A. Bilgiç. A multi-biometric recognition system based on deep features of face and gesture energy image. In *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 361–364, 2017. doi: 10.1109/INISTA.2017.8001186.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. doi: 10.1038/nature14539. URL https://doi.org/10.1038/nature14539.

C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu. Deeply-Supervised Nets. In G. Lebanon and S. V. N. Vishwanathan, editors, *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, volume 38 of *Proceedings of Machine Learning Research*, pages 562–570, San Diego, California, USA, 09–12 May 2015. PMLR. URL http://proceedings.mlr.press/v38/lee15a.html.

M. Leese. Fixing state vision: Interoperability, biometrics, and identity management in the eu. *Geopolitics*, 27(1):113–133, 2022. doi: 10.1080/14650045.2020.1830764. URL https://doi.org/10.1080/14650045.2020.1830764.

M. Leghari, S. Memon, L. D. Dhomeja, A. H. Jalbani, and A. A. Chandio. Deep feature fusion of fingerprint and online signature for multimodal biometrics. *Computers*, 10(2), 2021. ISSN 2073-431X. doi: 10.3390/computers10020021. URL https://www.mdpi.com/2073-431X/10/2/21.

Lin Hong and Anil Jain. Integrating faces and fingerprints for personal identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1295–1307, 1998.

C. Liu, F. Yin, Q. Wang, and D. Wang. Icdar 2011 chinese handwriting recognition competition. In *2011 International Conference on Document Analysis and Recognition*, pages 1464–1469, 2011. doi: 10.1109/ICDAR.2011.291.

Y. Liu, S. Piramanayagam, S. T. Monteiro, and E. Saber. Dense semantic labeling of very-high-resolution aerial imagery and lidar with fully-convolutional neural networks and higher-order crfs. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1561–1570, 2017. doi: 10.1109/CVPRW.2017.200.

R. M. Jomaa, H. Mathkour, Y. Bazi, and M. S. Islam. End-to-end deep learning fusion of fingerprint and electrocardiogram signals for presentation attack detection. *Sensors*, 20(7), 2020. ISSN 1424-8220. doi: 10.3390/s20072085. URL https://www.mdpi.com/1424-8220/20/7/2085.

M. V. Malakooti and M. R. N. Dobuneh. A lossless digital encryption system for multimedia using orthogonal transforms. In *2012 Second International Conference on Digital Information and Communication Technology and it's Applications (DICTAP)*, pages 240–244, 2012.

M. V. Malakooti and M. Khederzdeh. A lossless secure data embedding in image using dct and randomize key generator. In *2012 Second International Conference on Digital Information and Communication Technology and it's Applications (DICTAP)*, pages 236–239, 2012. doi: 10.1109/DICTAP.2012.6215381.

M. V. Malakooti and N. Mansourzadeh. A robust and lossless information embedding in image based on dct and scrambling algorithms. In *The International Conference on Computing Technology and Information Management (ICCTIM2014)*, pages 406–410. The Society of Digital Information and Wireless Communication, 2014a.

M. V. Malakooti and N. Mansourzadeh. A robust information security model for cloud computing based on the scrambling algorithm and multi-level encryption. In *The International Conference on Computing Technology and Information Management (ICCTIM2014)*, pages 411–416, 2014b.

M. V. Malakooti, A. P. Tafti, and H. R. Naji. An efficient algorithm for human cell detection in electron microscope images based on cluster analysis and vector quantization techniques. In *2012 Second International Conference on Digital Information and Communication Technology and it's Applications (DICTAP)*, pages 125–129, 2012. doi: https://doi.org/10.1109/DICTAP.2012.6215358.

M. V. Malakooti, V. Saffari, and T. Zeki. A novel method for secure image delivery

over mobile networks based on orthogonal transforms and scrambling algorithms. 2013.

D. Maltoni, D. Maio, A. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer-Verlag London, 2009. ISBN 978-1-4471-6106-6. doi: https://doi.org/10.1007/978-1-84882-254-2. URL https://www.springer.com/gp/book/9781848822535.

A. Matin, F. Mahmud, T. Ahmed, and M. S. Ejaz. Weighted score level fusion of iris and face to identify an individual. In *IEEE proceeding of the International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pages 1–4, 2017.

E. Micheli-Tzanakou and K. N. Plataniotis. *Biometrics: Terms and Definitions*, pages 142–147. Springer US, Boston, MA, 2011. ISBN 978-1-4419-5906-5. doi: 10.1007/978-1-4419-5906-5_730. URL https://doi.org/10.1007/978-1-4419-5906-5_730.

L. Min, L. Ting, and H. Yu-jie. Arnold transform based image scrambling method. In *Proceedings of 3rd International Conference on Multimedia Technology(ICMT-13)*, pages 1302–1309. Atlantis Press, 2013/11. ISBN 978-90-78677-89-5. doi: https://doi.org/10.2991/icmt-13.2013.160. URL https://doi.org/10.2991/icmt-13.2013.160.

G. Montavon, W. Samek, and K.-R. Müller. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, 73:1–15, 2018. ISSN 1051-2004. doi: https://doi.org/10.1016/j.dsp.2017.10.011. URL https://www.sciencedirect.com/science/article/pii/S1051200417302385.

N. Morizet and J. Gilles. A new adaptive combination approach to score level fusion for face and iris biometrics combining wavelets and statistical moments. In G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, F. Porikli, J. Peters,

J. Klosowski, L. Arns, Y. K. Chun, T.-M. Rhyne, and L. Monroe, editors, *Advances in Visual Computing*, pages 661–671, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. ISBN 978-3-540-89646-3.

B. Mróz-Gorgoń, W. Wodo, A. Andrych, K. Caban-Piaskowska, and C. Kozyra. Biometrics innovation and payment sector perception. *Sustainability*, 14(15), 2022. ISSN 2071-1050. doi: 10.3390/su14159424. URL https://www.mdpi.com/2071-1050/14/15/9424.

Multimedia-University. Mmu database (onlibne). In *Artificial Neural Networks*. IntechOpen, Rijeka, Last Accessed (February 2021). URL https://www.kaggle.com/naureenmohammad/mmu-iris-dataset.

H. R. Nemati, H. R. Nemati, and L. Yang. *Applied Cryptography for Cyber Security and Defense: Information Encryption and Cyphering*. IGI Global, USA, 1st edition, 2010. ISBN 161520783X.

C. Olah, A. Mordvintsev, and L. Schubert. Feature visualization. *Distill*, 2017. doi: 10.23915/distill.00007. https://distill.pub/2017/feature-visualization.

C. Olah, A. Satyanarayan, I. Johnson, S. Carter, L. Schubert, K. Ye, and A. Mordvintsev. The building blocks of interpretability. *Distill*, 2018. doi: 10.23915/distill.00010. https://distill.pub/2018/building-blocks.

M. O. Oloyede and G. P. Hancke. Unimodal and multimodal biometric sensing systems: A review. *IEEE Access*, 4:7532–7555, 2016. doi: 10.1109/ACCESS.2016.2614720.

I. Omara, G. Xiao, M. Amrani, Z. Yan, and W. Zuo. Deep features for efficient multi-biometric recognition with face and ear images . In C. M. Falco and X. Jiang, editors, *Ninth International Conference on Digital Image Processing (ICDIP 2017)*, volume 10420, pages 68 – 73. International Society for Optics and Photonics, SPIE, 2017. doi: https://doi.org/10.1117/12.2281694. URL https://doi.org/10.1117/12.2281694.

C. Paar and J. Pelzl. *Understanding Cryptography: A Textbook for Students and Practitioners*. Springer Publishing Company, Incorporated, 1st edition, 2009. ISBN 3642041000.

O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In M. W. J. Xianghua Xie and G. K. L. Tam, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 41.1–41.12. BMVA Press, September 2015a. ISBN 1-901725-53-7. doi: 10.5244/C.29.41. URL https://dx.doi.org/10.5244/C.29.41.

O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015b.

C. Raghavendra, A. Kumaravel, and S. Sivasubramanian. Iris technology: A review on iris based biometric systems for unique human identification. In *2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET)*, pages 1–6, 2017. doi: 10.1109/ICAMMAET.2017.8186679.

J. Ramesh and P. Ramesh. Contrast enhancement algorithm for colour images. *International Journal Magazine of Engineering, Technology, Management and Research*, 3(12):354–358, 2016.

U. H. Rao and U. Nayak. *Physical Security and Biometrics*, pages 293–306. Apress, Berkeley, CA, 2014. ISBN 978-1-4302-6383-8. doi: 10.1007/978-1-4302-6383-8_14. URL https://doi.org/10.1007/978-1-4302-6383-8_14.

A. Rattani and M. Tistarelli. Robust multi-modal and multi-unit feature level fusion of face and iris biometrics. In M. Tistarelli and M. S. Nixon, editors, *Advances in Biometrics*, pages 960–969, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg. ISBN 978-3-642-01793-3.

A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recogn. Lett.*, 24 (13):2115–2125, Sept. 2003. ISSN 0167-8655. doi: https://doi.org/10.1016/

S0167-8655(03)00079-5. URL https://doi.org/10.1016/S0167-8655(03)00079-5.

A. Ross, S. Banerjee, C. Chen, A. Chowdhury, V. Mirjalili, R. Sharma, T. Swearingen, and S. Yadav. Some research problems in biometrics: The future beckons. *CoRR*, abs/1905.04717, 2019. URL http://arxiv.org/abs/1905.04717.

A. A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics*. Springer US, 2006. ISBN 978-1-4419-3547-2. doi: https://doi.org/10.1007/0-387-33123-9. URL https://www.springer.com/gp/book/9780387222967.

D. Rumelhart, G. Hinton, and R. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986. doi: 10.1038/323533a0. URL https://doi.org/10.1038/323533a0.

O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and F.-F. Li. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.*, 115(3):211–252, 2015. URL http://dblp.uni-trier.de/db/journals/ijcv/ijcv115.html#RussakovskyDSKS15.

E. Salveggio, S. Lovaas, D. R. Lease, and R. Guess. *Biometric Authentication*, chapter 29, pages 29.1–29.28. John Wiley Sons, Ltd, 2012. ISBN 9781118851678. doi: https://doi.org/10.1002/9781118851678.ch29. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118851678.ch29.

J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015. doi: 10.1016/j.neunet.2014.09.003. Published online 2014; based on TR arXiv:1404.7828 [cs.NE].

F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2015. doi: 10.1109/cvpr.2015.7298682. URL http://dx.doi.org/10.1109/CVPR.2015.7298682.

M. Scott, T. Acton, and M. Hughes. An assessment of biometric identities as a standard for e-government services. *International Journal of Services and Standards*, 1(3):271–286, 2005. doi: 10.1504/IJSS.2005.005800. URL https://www.inderscienceonline.com/doi/abs/10.1504/IJSS.2005.005800.

M. Y. Shams, S. H. Sarhan, and A. S. Tolba. Adaptive deep learning vector quantisation for multimodal authentication. *J. Inf. Hiding Multim. Signal Process.*, 8(3):702–722, 2017. URL http://bit.kuas.edu.tw/%7Ejihmsp/2017/vol8/JIH-MSP-2017-03-020.pdf.

Z. Shao, H. Zhu, X. Tan, Y. Hao, and L. Ma. Deep multi-center learning for face alignment. *Neurocomputing*, 396:477–486, 2020. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2018.11.108. URL https://www.sciencedirect.com/science/article/pii/S0925231219304515.

W. M. Shbair, T. Cholez, J. Francois, and I. Chrisment. A survey of https traffic and services identification approaches, 2020.

K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

S. Soleymani, A. Dabouei, H. Kazemi, J. Dawson, and N. M. Nasrabadi. Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification, 2018.

S. Soleymani, A. Torfi, J. Dawson, and N. M. Nasrabadi. Generalized bilinear deep convolutional neural networks for multimodal biometric identification. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 763–767, 2018. doi: https://doi.org/10.1109/ICIP.2018.8451532.

N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014. URL http://jmlr.org/papers/v15/srivastava14a.html.

J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel. The german traffic sign recognition benchmark: A multi-class classification competition. In *The 2011 International Joint Conference on Neural Networks*, pages 1453–1460, 2011. doi: 10.1109/IJCNN.2011.6033395.

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016. doi: 10.1109/CVPR.2016.308.

C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 4278–4284. AAAI Press, 2017.

Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014. doi: 10.1109/CVPR.2014.220.

V. Talreja, M. C. Valenti, and N. M. Nasrabadi. Multibiometric secure system based on deep learning, 2017.

L. Tiong, S. Kim, and Y. Ro. Implementation of multimodal biometric recognition via multi-feature deep learning networks and feature fusion. *Multimed Tools Appl*, 78:22743–22772, 2019. doi: https://doi.org/10.1007/s11042-019-7618-0.

C. Tissé. *Contributing to the biometric verification of individuals by iris recognition. University of Montpellier II.* PhD thesis, 2003.

C. Touzet. *LES RESEAUX DE NEURONES ARTIFICIELS, INTRODUCTION AU CONNEXIONNISME*. Collection de l'EERIE. EC2, 1992. URL https://hal-amu.archives-ouvertes.fr/hal-01338010.

S. Umer, A. Sardar, B. C. Dhara, R. K. Rout, and H. M. Pandey. Person identification using fusion of iris and periocular deep features. *Neural Networks*, 122:407–419, 2020. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2019.11.009. URL https://www.sciencedirect.com/science/article/pii/S089360801930348X.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017. URL http://arxiv.org/abs/1706.03762.

A. Veit, M. Wilber, and S. Belongie. Residual networks behave like ensembles of relatively shallow networks. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 550–558, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.

S. Veluchamy and L. Karlmarx. System for multimodal biometric recognition based on finger knuckle and finger vein using feature-level fusion and k-support vector machine classifier. *IET Biometrics*, 6(3):232–242, 2017. doi: https://doi.org/10.1049/iet-bmt.2016.0112. URL https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-bmt.2016.0112.

R. Von Solms and J. Van Niekerk. From information security to cyber security. *Comput. Secur.*, 38:97–102, Oct. 2013. ISSN 0167-4048. doi: https://doi.org/10.1016/j.cose.2013.04.004. URL https://doi.org/10.1016/j.cose.2013.04.004.

C. Wang, Y. Zhang, and X. Zhou. Robust image watermarking algorithm based on asift against geometric attacks. *Applied Sciences*, 8(3), 2018. ISSN 2076-3417. doi: 10.3390/app8030410. URL https://www.mdpi.com/2076-3417/8/3/410.

Y. Wang, L. Zhang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras. Face relighting from a single image under arbitrary unknown lighting conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):1968–1984, 2009. doi: 10.1109/TPAMI.2008.244.

Y. Wu, P. Qian, and X. Zhang. Two-level wavelet-based convolutional neural network for image deblurring. *IEEE Access*, 9:45853–45863, 2021. doi: 10.1109/ACCESS.2021.3067055.

L. Xu, X. Gou, Z. Li, and J. Li. A novel chaotic image encryption algorithm using block scrambling and dynamic index based diffusion. *Optics and Lasers in Engineering*, 91:41–52, 2017. ISSN 0143-8166. doi: https://doi.org/10.1016/j.optlaseng.2016.10.012. URL https://www.sciencedirect.com/science/article/pii/S0143816616302858.

J. Yang, W. Sun, N. Liu, Y. Chen, Y. Wang, and S. Han. A novel multimodal biometrics recognition model based on stacked elm and cca methods. *Symmetry*, 10(4), 2018. ISSN 2073-8994. doi: https://doi.org/10.3390/sym10040096. URL https://www.mdpi.com/2073-8994/10/4/96.

M. B. Yassein, S. Aljawarneh, E. Qawasmeh, W. Mardini, and Y. Khamayseh. Comprehensive study of symmetric key and asymmetric key encryption algorithms. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–7, 2017. doi: https://doi.org/10.1109/ICEngTechnol.2017.8308215.

L. Zhang, X. Tian, and S. Xia. A scrambling algorithm of image encryption based on rubik's cube rotation and logistic sequence. In *2011 International Conference on Multimedia and Signal Processing*, volume 1, pages 312–315, 2011. doi: 10.1109/CMSP.2011.69.

W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, Dec. 2003. ISSN 0360-0300. doi: 10.1145/954339.954342. URL https://doi.org/10.1145/954339.954342.

Appendices

Appendix A

Parts of the code of the deep learning solution

In this thesis, for the training of deep models, two databases were used: the VGGFace dataset [Parkhi et al., 2015b] for the face and the IIT Delhi dataset [Kumar and Passi, 2010] for the iris. To train the proposed models, total fine-tuning was performed using pre-trained deep learning architectures on the imagenet [Russakovsky et al., 2015] dataset.

The first step to performing a fine-tuning using Keras is to initialize the deep architectures with the ImageNet weights. The Python code used for the Iris recognition is as follows:

```
#loading packages
import pandas as pd
import numpy as np
import pickle
from matplotlib import pyplot as plt
```

```
#loading keras packages
from keras.applications.vgg16 import VGG16
from keras.applications.vgg16 import preprocess_input
from keras.applications.resnet50 import ResNet50
from keras.applications.inception_v3 import InceptionV3
from keras.applications.densenet  import DenseNet121
from keras.applications.mobilenet  import MobileNet
from keras.models import Sequential
from keras.layers import Flatten

#loading models with imagenet weigths
width=224 # Default mage width
height=224 # Default image width
channels=3 # Default image channels
num_classes=224 # Number of classes (individuals)
model1=VGG16(weights='imagenet',include_top=False,input_shape=
                                    (width,height,channels))
model2=ResNet50(weights='imagenet',include_top=False,input_shape=
                                    (width,height,channels))
model3=InceptionV3(weights='imagenet',include_top=False,input_shape=
                                    (width,height,channels))
model4=DenseNet121(weights='imagenet',include_top=False,input_shape=
                                    (width,height,channels))
model5=MobileNet(weights='imagenet',include_top=False,input_shape=
                                    (width,height,channels))
```

Since total fine-tuning has been adopted for the transfer of learning, the next step is to make all layers trainable. The code is as follows:

```
for layer in model1.layers:
```

```
  layer.trainable = True
for layer in model2.layers:
  layer.trainable = True
for layer in model3.layers:
  layer.trainable = True
for layer in model4.layers:
  layer.trainable = True
for layer in model5.layers:
  layer.trainable = True
```

Once the architecture is defined, the final decision layers should be added (fully connected). For this, a Global Average pooling layer is applied to the model's output, followed by a Dense layer of 512 units and finally an output layer of 224 neurons, representing the number of classes. Furthermore, the Softmax function is used because this problem is multi-class (not a binary classification). The code is as follows:

```
# Usefull packages
from keras.models import Sequential
from keras.layers import Dense, Dropout, Activation, GlobalAveragePooling2D
from keras.layers import Conv2D, MaxPooling2D, ZeroPadding2D, Flatten
from keras.layers.normalization import BatchNormalization
from keras.models import Model
# Fully connected layers function
def add_layer(bottom_model, num_classes):
  top_model = bottom_model.output
  top_model = GlobalAveragePooling2D()(top_model)
  top_model = Dense(512 , activation = 'relu')(top_model)
  top_model = Dense(num_classes , activation = 'softmax')(top_model)
  return top_model
```

At this point, the previously defined function (add_layer) is called to create output layers for each model, the model is grouped (the basic architecture with output layers), and the inputs and output of the model are specified as follow:

```
FC1 = add_layer(model1, num_classes)
FC2 = add_layer(model2, num_classes)
FC3 = add_layer(model3, num_classes)
FC4 = add_layer(model4, num_classes)
FC5 = add_layer(model5, num_classes)

model1 = Model(inputs = model1.input, outputs = FC1)
model2 = Model(inputs = model2.input, outputs = FC2)
model3 = Model(inputs = model3.input, outputs = FC3)
model4 = Model(inputs = model4.input, outputs = FC4)
model5 = Model(inputs = model5.input, outputs = FC5)
```

The next step is crucial and concerns the preparation of the data. Indeed, the training and validation images are loaded and processed. The processing consists in augmenting the data and putting them in a data generator with batch size 64.
Data augmentation is a technique for artificially creating transformed versions of images in the training dataset that belong to the same class as the original image. Several transformations such as shifts, flips, zooms are applied. Data augmentation will allow the models to be more robust and perform better. The code is as follows:

```
from keras.preprocessing.image import ImageDataGenerator

Train_Data_Dir = "Iris_Data/Iris_Train/"
Valid_Data_Dir = "Iris_Data/Iris_Valid/"

train_datagen=ImageDataGenerator(rescale=1./255,
```

```
                                    rotation_range=10,
                                    width_shift_range=0.2,
                                    height_shift_range=0.2,
                                    shear_range=0.2,
                                    zoom_range=0.2,
                                    horizontal_flip=True,
                                    fill_mode="nearest",
                                    featurewise_center=True,
                                    featurewise_std_normalization=True,)

valid_datagen=ImageDataGenerator(rescale=1./255,
                                    rotation_range=10,
                                    width_shift_range=0.2,
                                    height_shift_range=0.2,
                                    shear_range=0.2,
                                    zoom_range=0.2,
                                    horizontal_flip=True,
                                    fill_mode="nearest",
                                    featurewise_center=True,
                                    featurewise_std_normalization=True,)

batch_size = 64

train_generator = train_datagen.flow_from_directory(
    Train_Data_Dir,
    target_size=(width, height),
    batch_size=batch_size,
    class_mode='categorical')
```

```
validation_generator = valid_datagen.flow_from_directory(
    Valid_Data_Dir,
    target_size=(width, height),
    batch_size=batch_size,
    class_mode='categorical')
```

The last step is the compilation of the models and the training process. The number of epochs, the optimizer, and callbacks must be specified in this step. Let us take the example of the VGG16 model training:

```
from keras.callbacks import ModelCheckpoint
checkpoint = ModelCheckpoint("Saved_models/Iris_VGG16.h5",
    monitor='val_acc', verbose=1, save_best_only=True, mode='auto')
num_Epochs = 100

model1.compile(loss='categorical_crossentropy',
    optimizer='SGD', metrics=['acc'])

history  = model1.fit_generator(
        train_generator,
        epochs=num_Epochs,
        validation_data=validation_generator)
```

Once the training is finished, the best model will be saved in the file of type h5. This file contains the architecture with the model weights. Once the face and iris models are trained, they will be used according to the chosen combination mode (feature level fusion or decision level fusion). In the first case, the models are used as feature extractors. Here an example of using the trained Inception_V3 Face model as feature extractor:

```
from keras.models import Model
from keras.models import load_model

model = load_model("All_Models_h5/Face_InceptionV3.h5")

feature_extractor = Model(
    inputs=model.inputs,
    outputs=[layer.output for layer in model.layers],)
```

The "outputs" file contains all features at each convolution layer. Since we deal with late fusion, we use the last layer, which is the 19th in Inception_V3.

```
import os
import cv2
import numpy as np

Data_Dir = "/Face_Train/"

Features_Data=np.zeros((1,512))
Features_Mean_Data=np.zeros((1,512))

Label = []
l = 1

for i in os.listdir(Data_Dir):
  for j in os.listdir(Data_Dir+i):
    Image = cv2.imread(Data_Dir+i+"/"+j)/255
    T = np.expand_dims(Image, axis = 0)
    features = feature_extractor(T)
    vec1 = features[19]
```

```
    Features_Data=np.concatenate((Features_Data,vec1),axis=0)
  Features_Data = np.delete(Features_Data,0,0)
  Features_Data = np.array(Features_Data)
  F=np.mean(Features_Data, axis = 0)
  FF=np.expand_dims(F,axis = 0)
  Features_Mean_Data=np.concatenate((Features_Mean_Data,FF), axis = 0)
  Label.append(l)
  l=l+1

Features_Mean_Data = np.delete(Features_Mean_Data,0,0)
Features_Mean_Data = np.array(Features_Mean_Data)
Label = np.array(Label)

np.save('Extracted_Data/Face_InceptionV3/Features_Mean_Face_InceptionV3.npy',
                                Features_Data)
np.save('Extracted_Data/Face_InceptionV3/Labels_Mean_Face_InceptionV3.npy',
                            Label)
```

At the end of this step, two feature vectors will be generated (one for the face and the second for the iris), combined to make the matching.

```
import os
import cv2
import numpy as np
from sklearn.neighbors import NearestNeighbors
from sklearn.neighbors import DistanceMetric

# use of Euclidean distance
dist = DistanceMetric.get_metric('euclidean')
```

```
# load registration features
Iris_Features_DenseNet121 = np.load('Extracted_Data/Iris/
        Features_Iris_DenseNet121.npy')
Iris_Hypo_DenseNet121 = NearestNeighbors(n_neighbors=1,
             metric='euclidean').fit(Iris_Features_DenseNet121)
Face_Features_Facenet = np.load('Extracted_Data/Face/
        Features_Face_Facenet.npy')
Face_Hypo_Facenet = NearestNeighbors(n_neighbors=1,
             metric='euclidean').fit(Face_Features_Facenet)

Iris_Label = np.load('Extracted_Data/Iris/Labels_Iris.npy')
Face_Label = np.load('Extracted_Data/Face/Labels_Face.npy')

# load test features
Iris_Features_DenseNet121_Test = np.load('Extracted_Data/Iris/
        Features_Iris_DenseNet121_Test.npy')
Face_Features_acenet_Test = np.load('Extracted_Data/Face/
        Features_Face_Facenet_Test.npy')

# distance calculation
Iris_distances_DenseNet121, Iris_indices_DenseNet121 =
    Iris_Hypo_DenseNet121.kneighbors(Iris_Features_DenseNet121_Test)
Face_distances_Facenet, Face_indices_Facenet =
    Face_Hypo_Facenet.kneighbors(Face_Features_Facenet_Test)
```

Part of the matching code where the distance is compared to the decision threshold:

```
# Fusion between Iris_DenseNet121 && Facenet based on Threshold
if Iris_distances_DenseNet121[i]<=Threshold_Iris_DenseNet121 and Face_distances_Fa
P_DenseNet121_Facenet = Face_Label[Face_indices_model_Facenet[i]]
```

```
else :
P_DenseNet121_Facenet = 0
```

In the second case (decision level fusion), each model produces a classification, and the results are combined.