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

UNIVERSITÉ DE BATNA 2
Faculté des Mathématiques et d'Informatique
Département d'Informatique

Thèse

En vue de l'obtention du diplôme de
Doctorat en Informatique

Approche Multimodale basée sur l'Apprentissage Profond pour l'Identification Biométrique

Présentée par :

ABDESSALAM HATTAB

Les membres du jury:

Prof. RACHID SEGHIR	Université Batna 2	Président
Prof. ALI BEHLOUL	Université Batna 2	Rapporteur
Prof. KAMAL EDDINE MELKEMI	Université Batna 2	Examinateur
Prof. LARBI GUEZOULI	HNS RE2SD Batna	Examinateur
Prof. MOHAMED BENMOHAMMED	Université de Constantine 2	Examinateur

ANNÉE UNIVERSITAIRE : 2022-2023

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

UNIVERSITY OF BATNA 2
Faculty of mathematics and computer science
Department of computer science

Thesis

For obtaining the diploma of Doctorate
in Computer Science

A Multimodal Approach Based on Deep Learning for Biometric Identification

Presented By:

ABDESSALAM HATTAB

The jury members:

Prof. RACHID SEGHIR	University of Batna 2	President
Prof. ALI BEHLOUL	University of Batna 2	Reporter
Prof. KAMAL EDDINE MELKEMI	University of Batna 2	Examiner
Prof. LARBI GUEZOULI	HNS RE2SD Batna	Examiner
Prof. MOHAMED BENMOHAMMED	University of Constantine 2	Examiner

ACADEMIC YEAR: 2022-2023

ABSTRACT

Due to the growing need for user identification in various modern applications, experts highly recommend incorporating biometric technology in the application development field. Recently, many recognition systems using face and iris traits obtained remarkable performance, particularly since these biometric traits are captured from a distance without physical contact with sensors. This feature reduces the potential spread of the COVID-19 pandemic and other diseases by touch and makes biometrics more convenient and user-friendly. However, the recognition systems' performance is significantly reduced when these traits are captured under uncontrolled conditions, including occlusion, poses and illumination variation. Because the handcrafted approach used by some recognition systems extracts local features from the global image, where the image's regions affected by the uncontrolled conditions often influence the quality of extracted features.

This thesis proposed a robust face recognition system that extracts features from important facial regions, using Scale-invariant feature transform (SIFT) to identify significant facial regions and adaptive Local Ternary Pattern (ALTP) to extract features. The proposed system achieved promising results on two benchmark face datasets captured under unconstrained conditions, achieving 99.75% on the ORL dataset and 95.12% on the FERET dataset. In addition, a new face recognition system based on Deep Learning has been proposed using the pre-trained AlexNet-v2 model. The proposed system achieved excellent results on the ORL and FERET face datasets, reaching 100% on the first and 99.89% on the second. Moreover, we proposed a novel face recognition system to address the issue of illumination variation. The system used a novel model inspired by the pre-trained VGG16 model. The proposed system achieved state-of-the-art results, reaching 99.32% on the Extended Yale B dataset and 99.79% on the AR dataset.

Another contribution of this work is developing a novel iris recognition system based on Transfer Learning to achieve high accuracy rates. The proposed system used the Yolov4-tiny model to localize the iris region, while a novel Deep Convolutional Neural Network (CNN) model inspired by the pre-trained Inception-v3 model was used for features extraction. The performance of this system was evaluated on four different iris databases captured under non-cooperative conditions, where it achieved a new state-of-the-art accuracy rate reaching 99.91%, 99.60%, 99.91%, and 99.19% on the IITD, CASIA-Iris-v1, CASIA-Iris-Interval, and CASIA-Iris-Thousand, respectively.

The proposed unimodal systems achieved high accuracy compared to state-of-the-art methods. However, relying solely on a unimodal biometric trait is inadequate for high-security requirements in military and government applications.

Finally, three face-iris multimodal biometric systems have been proposed in this thesis. The first employs the fusion of images, the second utilizes feature-level fusion, and the third is based on score-level fusion. The proposed systems used Yolov4-tiny to detect the face and both iris regions. In addition, they applied a new deep CNN model inspired by the pre-trained Xception model to extract features. To evaluate the performance of the proposed systems, a two-fold cross-validation protocol is employed on the CASIA-ORL and SDUMLA-HTM multimodal benchmark databases. The results showed that our systems achieved a perfect score of 100% on both databases. Remarkably, the system utilizing score-level fusion outperformed the other systems, achieving outstanding results of 100% on the CASIA-ORL database and over 99% on the SDUMLA-HTM database, with only one sample used for training.

Keywords: *Multimodal biometric systems, Face recognition, Iris recognition, Deep Learning, Transfer Learning, Convolutional Neural Network.*

ملخص

نظرًا للحاجة المتزايدة لتحديد هوية المستخدم في مختلف التطبيقات الحديثة، يوصي الخبراء بشدة بدمج تكنولوجيا القياسات الحيوية في مجال تطوير التطبيقات. مؤخرًا، حصلت العديد من أنظمة التعرُّف التي تستخدم سمة الوجه وقزحية العين على أداء رائع، لا سيما وأن هذه السمات الحيوية يتم التقاطها من مسافة بعيدة دون الاتصال الجسدي بأجهزة الاستشعار، حيث تساهم هذه الميزة في تقليل الانتشار المحتمل لوباء COVID-19 والأمراض الأخرى التي تنتقل باللمس كما تجعل القياسات الحيوية أكثر ملاءمة وسهولة في الاستخدام. ومع ذلك، ينخفض أداء أنظمة التعرُّف بشكل كبير عندما يتم التقاط هذه السمات في ظل ظروف غير خاضعة للرقابة، بما في ذلك العوائق التي تحجب السمة وتباين الإضاءة ووضعية السمة. لأن النهج اليدوي المستخدم من قبل بعض أنظمة التعرُّف يستخرج السمات المحلية من كامل الصورة، حيث غالبًا ما تؤثر مناطق الصورة المتأثرة بالظروف الغير خاضعة للرقابة على جودة الميزات المستخرجة.

اقترحت هذه الأطروحة نظامًا قويًا للتعرُّف على الوجوه يستخرج الميزات من مناطق الوجه المهمة، حيث يستخدم النظام المقترح تقنية تحويل الميزات الغير المرتبطة بمقياس (SIFT) لتحديد مناطق الوجه المهمة، كما يستخدم خوارزمية الأنماط الثلاثية المحلية التكيفية (ALTP) لاستخراج الميزات. حقق النظام المقترح نتائج واعدة على مجموعتي بيانات وجه معياريتين تم التقاطهما في ظل ظروف غير مقيدة، حيث حقق دقة تعرف تساوي 99.75٪ على مجموعة بيانات ORL و95.12٪ على مجموعة بيانات FERET. بالإضافة إلى ذلك، تم اقتراح نظام جديد للتعرُّف على الوجوه يعتمد على التعلم العميق، حيث استخدم النظام المقترح نموذج AlexNet-v2 تم تدريبه مسبقًا. حقق النظام المقترح معدل دقة تعرُّف ممتاز في مجموعتي بيانات ORL وFERET، وصل إلى 100٪ في الأولى و99.89٪ في الثانية. علاوة على ذلك، اقترحنا نظامًا جديدًا للتعرُّف على الوجوه لمعالجة مشكلة اختلاف الإضاءة. استخدم النظام نموذجا جديدًا مستوحى من نموذج VGG16 المدرب مسبقًا. حقق النظام المقترح نتائج تفوقت على أحدث الأنظمة في مجال التعرُّف على الوجه في ضل تغيير الإضاءة، حيث حقق نظامنا المقترح 99.32٪ على مجموعة البيانات Extended Yale B و 99.79٪ على مجموعة البيانات AR.

كذلك، ساهمت هذه الأطروحة في تطوير نظام جديد للتعرُّف على قزحية العين، حيث اعتمد النظام المقترح على نقل التعلم لتحقيق معدلات دقة عالية. استخدم النظام المقترح نموذج Yolov4-tiny لتحديد منطقة قزحية العين، كما استخدم نموذج جديد لشبكة عصبية تلافيفيه (CNN) مستوحى من نموذج Inception-v3 المدرب مسبقًا لاستخراج الميزات. تم تقييم أداء هذا النظام على أربع قواعد بيانات مختلفة لقزحية العين تم التقاطها في ظل ظروف غير تعاونية. حقق النظام المقترح معدل دقة جديد وصل إلى 99.91٪، 99.60٪، 99.91٪ و99.19٪ على IITD، CASIA-Iris-v1، CASIA-Iris-Interval و CASIA-Iris-Thousand على التوالي.

حققت الأنظمة أحادية الوسائط المقترحة دقة عالية مقارنة بأحدث الأساليب المقترحة للتعرُّف على الوجه وقزحية العين. ومع ذلك، فإن الاعتماد فقط على سمة القياسات الحيوية أحادية الوسائط غير كافٍ للتطبيقات التي تتطلب أمانًا عالي المستوى، مثل التطبيقات العسكرية والحكومية الحساسة.

أخيرا، اقترحنا ثلاثة أنظمة متعددة المقاييس الحيوية تجمع بين الوجه وقزحية العين. الأول يستخدم الانصهار على مستوى الصورة، والثاني يستخدم الانصهار على مستوى الميزة، والثالث يعتمد على الانصهار على مستوى الدرجات. استخدمت الأنظمة المقترحة Yolov4-tiny لاكتشاف الوجه وكلا منطقتي القزحية. بالإضافة إلى ذلك استخدمنا نموذج CNN عميق جديد مستوحى من نموذج Xception المدرب مسبقًا لاستخراج الميزات. لتقييم أداء الأنظمة المقترحة، تم استخدام بروتوكول التحقق المتقاطع الثنائي في قاعدتي بيانات معياريتين متعددتي السمات الحيوية CASIA-ORL وSDUMLA-HTM. أظهرت النتائج أن أنظمتنا حققت معدلات دقة مثالية وصلت لـ 100٪ في قاعدتي البيانات. لاحظنا، تفوق النظام الذي يستخدم الانصهار على مستوى الدرجات مقارنة بالأنظمة الأخرى، حيث حقق دقة تعرُّف وصلت لـ 100٪ على قاعدة بيانات CASIA-ORL وأكثر من 99٪ على قاعدة بيانات SDUMLA-HTM، بالرغم من استخدامنا لعينة واحدة فقط للتدريب.

الكلمات المفتاحية: *أنظمة متعددة المقاييس الحيوية، التعرُّف على الوجه، التعرُّف على قزحية العين، التعلم العميق، نقل التعلم، الشبكة العصبة التلافيفيه.*

RÉSUMÉ

En raison du besoin croissant d'identification des utilisateurs dans diverses applications modernes, les experts recommandent fortement d'intégrer la technologie biométrique dans le domaine du développement d'applications. Récemment, de nombreux systèmes de reconnaissance utilisant les traits du visage et de l'iris ont obtenu des performances remarquables, d'autant plus que ces traits biométriques sont capturés à distance sans contact physique avec des capteurs. Cette caractéristique réduit la propagation potentielle de la pandémie COVID-19 et d'autres maladies par le toucher et rend la biométrie plus pratique et conviviale. Cependant, les performances des systèmes de reconnaissance sont considérablement réduites lorsque ces traits sont capturés dans des conditions non contrôlées, notamment l'occlusion, les poses et les variations d'éclairage. Parce que les méthodes classiques utilisées par certains systèmes de reconnaissance extrait des caractéristiques locales de l'image globale, où les caractéristiques extraites sont souvent influencées par les facteurs environnementaux présents lors de la capture d'image.

Cette thèse propose un système de reconnaissance faciale robuste qui extrait les caractéristiques des régions faciales importantes, en utilisant la transformation de de caractéristiques visuelles invariante à l'échelle (SIFT) pour identifier les régions faciales significatives et le motif ternaire local adaptatif (ALTP) pour extraire les caractéristiques. Le système proposé a obtenu des résultats prometteurs sur deux ensembles de données de visages capturés dans des conditions incontrôlées, atteignant 99.75% sur l'ensemble de données ORL et 95.12% sur l'ensemble de données FERET. De plus, un nouveau système de reconnaissance faciale basé sur l'apprentissage profond a été proposé en utilisant le modèle AlexNet-v2 pré-entraîné. Le système proposé a obtenu d'excellents résultats sur les ensembles de données de visages ORL et FERET, atteignant 100% sur le premier et 99.89% sur le second. En outre, nous avons proposé un nouveau système de reconnaissance des visages pour traiter le problème de la variation d'éclairage. Le système utilise un nouveau modèle inspiré du modèle VGG16 pré-entraîné. Le système proposé a obtenu des résultats de pointe, atteignant 99.32% sur l'ensemble de données Extended Yale B et 99.79% sur l'ensemble de données AR.

Une autre contribution de ce travail est le développement d'un nouveau système de reconnaissance de l'iris basé sur l'apprentissage par transfert pour atteindre des taux de précision élevés. Le système proposé utilise le modèle Yolov4-tiny pour localiser la région de l'iris, tandis qu'un nouveau modèle de Réseau neuronal convolutif profond (CNN) inspiré du modèle Inception-v3 pré-entraîné a été utilisé pour l'extraction des caractéristiques. Les performances de ce système ont été évaluées sur quatre bases de données d'iris différentes capturées dans des conditions non coopératives, où il a atteint un nouveau taux de précision de pointe atteignant 99.91%, 99.60%, 99.91% et 99.19% sur l'IITD, CASIA-Iris-v1, CASIA-Iris-Interval, et CASIA-Iris-Thousand, respectivement.

Les systèmes unimodaux proposés ont atteint une précision élevée par rapport aux méthodes les plus récentes. Toutefois, le fait de s'appuyer uniquement sur un trait biométrique unimodal est inadéquat pour les applications nécessitant un niveau de sécurité élevé, telles que les applications militaires et gouvernementales sensibles.

Enfin, trois systèmes biométriques multimodaux visage-iris ont été proposés dans cette thèse. Le premier utilise la fusion au niveau de l'image, le deuxième utilise la fusion au niveau des caractéristiques et le troisième est basé sur la fusion au niveau du score. Les systèmes proposés utilisent Yolov4-tiny pour détecter le visage et les deux régions de l'iris. De plus, ils ont appliqué un nouveau modèle CNN profond inspiré du modèle Xception pré-entraîné pour extraire les car-

actéristiques. Pour évaluer les performances des systèmes proposés, un protocole de validation croisée double est utilisé sur les bases de données multimodales CASIA-ORL et SDUMLA-HTM. Les résultats ont montré que nos systèmes ont atteint un score parfait de 100% sur les deux bases de données. Remarquablement, le système utilisant la fusion au niveau du score a surpassé les autres systèmes, obtenant des résultats exceptionnels de 100% sur la base de données CASIA-ORL et de plus de 99% sur la base de données SDUMLA-HTM, avec un seul échantillon utilisé pour l'entraînement.

Mots-clés : *Système biométrique multimodal, reconnaissance du visage, reconnaissance de l'iris, apprentissage profond, apprentissage par transfert, réseau neuronal convolutif.*

ACKNOWLEDGEMENTS

Writing a PhD thesis is a long and arduous journey and I am deeply grateful to all those who have supported me along the way. .First of all, I am deeply grateful to my GOD for empowering me in all my life stations, including this journey.

I would like to express my sincere gratitude to my advisor, Prof. Ali BEHLOUL, for its invaluable guidance, support, and encouragement throughout my research.

I am so thankful to the jury members: Prof. Rachid SEGHIR, Prof. Kamal Eddine MELKEMI, Prof. Larbi GUEZOULI , and Prof. Mohamed BENMOHAMMED for accepting the request to join us and contribute to evaluating the presented work based on their rich experiences.

I would like to express my heartfelt gratitude to my family, who have always been there for me, providing love, support, and encouragement throughout my academic journey.

Finally, I am also thankful to my colleagues and friends who have provided me with support, encouragement, and motivation during my PhD journey. Their camaraderie and friendship have made this experience truly unforgettable.

This work would not have been possible without the support and help of all of these individuals, and I am deeply grateful to each and every one of them.

AUTHOR'S DECLARATION

We declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programs, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such

EMAIL :

SIGNED: ..... DATE: .....

TABLE OF CONTENTS

List of Tables

Table **Page**

List of Figures

Figure **Page**

List of Abbreviations

Symbol	Definition
ALTP	Adaptive Local Ternary Patterns/
AFIS	Applied Automated Fingerprint Identification Systems/
AI	Artificial Intelligence
ANN	Artificial Neural Networks
CHT	Circular Hough Transform
CR	Collaborative Representation
CG-RF	Complete Gabor Filter with Random Forest
CLAHE	Contrast Limited Adaptive Histogram Equalization
CS	Contrast Stretching
CNN	Convolutional Neural Network
CSPNet	Cross Stage Partial Network
DA	Data Augmentation
DCT	Discrete Cosine Transform
DBN	Deep Belief Network
DL	Deep Learning
DNN	Deep Neural Network
DNA	DeoxyriboNucleic Acid
DC	Directional Coding
DWPT	Discrete Wavelet Packet Transform

DWT	Discrete Wavelet Transform
ECG	Electrocardiograph
EMiCoAReNet	Emerging Mixed Convolutional and Adaptive Residual Network
FFT	Fast Fourier Transform
FBI	Federal Bureau of Investigation
FDA	Fisher Discriminant Analysis
FLDA	Fisher Linear Discriminant Analysis
FPS	Frames Per Second
FC	Fully Connected
FLBP	Fuzzy Local Binary Pattern
GM	Gabor Magnitude
GW	Gabor Wavelets
GIC	Gamma Intensity Correction
GAN	Generative Adversarial Network
GAP	Global Average Polling
HOG	Histogram of Oriented Gradients
HF	Homomorphic Filter
HRPSM CNN	Hybrid Robust Point Set Matching Convolutional Neural Network
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IKLDA	Improved Kernel Linear Discriminant Analysis
IAFIS	Integrated Automated Fingerprint Identification System
JSLAR	Joint Sparse Locality-Aware Regression
KCFT	K Class Feature Transfer
k-NN	k-Nearest-Neighbor
LoG	Laplacian of Gaussian

LDA	Linear Discriminant Analysis
LCSRC	Linear Collaborative Discriminant Regression Classification
LBP	Local Binary Pattern
LGPP	Local Gradient Probabilistic Pattern
LTP	Local Ternary Pattern
LSTM	Long Short-Term Memory
MR	Magnitude Responses
M-CLAHE	Modified Contrast Limited Adaptive Histogram Equalization
NIST	National Institute of Standards and Technology
NN	Nearest Neighbor
POEM	Oriented Edge Magnitudes
OGPCI	Oriented Gabor Phase Congruency Image
PSO	Particle Swarm Optimization
PCA	Principal Component Analysis
HPC	Performance Computing
PINs	Personal Identification Numbers
PNNs	Probabilistic Neural Networks
PR	Product Rule
RNN	Recurrent Neural Networks
RR	Reflectance Ratio
RRHE	Reflectance Ratio and Histogram Equalization
ROI	Region of Interest
RGHF	Relative Gradient Histogram Features
RSLDA	Robust Sparse Linear Discriminant Analysis
SD	Standard Deviation

SIFT	Scale Invariant Feature Transform
SOM	Self Organizing Maps
SVD	Singular Value Decomposition
SURF	Speeded-Up Robust Features
SR	Sum Rule
SVM	Support Vector Machines
VGG	Visual Geometry Group
IWT	Wavelet Transform
YOLO	You Only Look Once

CHAPTER 1

INTRODUCTION

"The good thing about biometrics is that people are relieved from the responsibility of designing and remembering a strong password – you don't have to remember your fingerprint, it's a part of who you are."

Mikhail Gofman, Multimodal Biometrics Expert.

1.1 Research background and problem statement

The increasing reliance on digital systems and the internet's growth have created a high demand for secure personal identification in our digital world. With the proliferation of military and government information systems, online transactions, and critical assets, there is a growing need for robust methods of identifying users who access this information and resources. Biometrics is one of the most promising solutions for addressing this need.

Biometrics is a widespread technique that exploits physical or behavioral traits, such as the face, iris, walking style, or keystroke patterns, to identify or confirm a person's identity. The principal goal of biometrics is to provide a more accurate and reliable personal authentication and identification technique than traditional methods, such as PINs, passwords, and ID cards.

The selection of biometric traits is the most crucial factor in the success of biometric systems building. Physical traits are more permanent and less vulnerable to spoofing than behavioral traits, which can be simply imitated and severely influenced by various psychological factors

and physical conditions. With the COVID-19 pandemic and the concern of disease transmission through touch, users have grown wary of touch-based biometric systems, which motivated researchers to strive to develop biometric technologies that capture biometric traits without physical contacts, like facial and iris recognition.

The face is a widely utilized biometric trait for human identification and is used in various fields such as attendance management, video surveillance, healthcare, and authentication on mobile phones. The popularity of face recognition is attributed to two main factors. The first is that it is a non-cooperative identification system where the user's face can be captured without specific actions or poses. The second one, it is a natural way to recognize persons, doesn't compromise the individual's privacy since faces are not covered in daily life. Recently, many systems have been proposed to improve face recognition performance [1–3] . However, the illumination variation, head poses, facial expressions, occlusion, and low-light conditions, have impeded the accuracy of the proposed systems.

Currently, iris recognition is considered one of the most reliable and distinctive biometric modalities used for personal identification. Iris recognition is widely used in many applications, such as unlocking mobile devices, law enforcement, and verifying passports. The iris patterns remain stable over a person's life, can be captured without physical contact, and has a rich and unique texture that can distinguish even between twins [4]. Lately, several research papers have tried to build robust iris recognition systems [5–7]. However, the performance of iris recognition systems is impacted by non-cooperative conditions, such as reflection, blur, partial closure of eyes, and pupil dilation/constriction due to light conditions. Furthermore, certain eye diseases are a real challenge for iris recognition [8].

Therefore, relying solely on a unimodal biometric trait like the face or iris is insufficient for high-security needs as in military and government applications. Recently, there has been a surge in the popularity of multimodal systems which utilize multiple biometric traits. This popularity is primarily due to their ability to overcome unimodal systems' performance and reliability limitations, like intra-class variation, inter-class similarity, susceptibility to spoofing, and non-universality issues [9, 10].

The widespread use of face and iris traits for personal identification in daily life. Additionally, the complementary between the face and the iris traits, in which the iris is the most precise and the face the least invasive. Furthermore, the advancements in high-resolution cameras can capture both traits at a distance with one device. All these points to the potential success of combining these two modalities.

Several multimodal biometric systems fused the face, and the iris traits have been proposed in recent years [11–13]. However, the proposed multimodal systems are affected by the non-constrained condition. Additionally, they are far from perfect in terms of accuracy if one of the traits is missed or unavailable. These challenges force us to develop effective multimodal systems for applications that require high security, like military and government applications.

Recently, Deep Convolutional Neural Networks (CNNs) have seen significant success in the field of computer vision [14, 15], being applied to a variety of domains, such as object detection [16] and image recognition [17]. The prior successes of Deep CNN in object detection and recognition motivate us to believe that utilizing Deep CNN in biometrics will achieve high success.

1.2 Thesis Objective and Contribution

The objectives of this Ph.D. thesis are focused on advancing the current state of biometric technology and developing robust, effective, and reliable solutions for biometric identification. This dissertation aims to build an effective face recognition system that can handle the limitations of unconstrained environments. In addition, it aims to propose a robust iris recognition system that can detect and recognize iris images captured under non-cooperative conditions. The main goal is to design and implement a robust face-iris multimodal biometric system based on deep convolutional neural networks. This system aims to improve the accuracy and robustness of biometric identification and use it in applications requiring high-security levels of reliability. The ultimate goal is to evaluate and compare the proposed systems' performance with state-of-the-art methods and prove their robustness against various factors such as poses, illumination variation, and occlusion. The first contribution part of this thesis focuses on the issue of unimodal biometric systems utilizing face and iris traits. The second part addresses the challenge of building an effective multimodal biometric system that fuses face and iris traits. The significant contributions of this dissertation can be summarized as follows:

I. An efficient face recognition system based on handcraft methods is proposed. It applied the Scale-Invariant Feature Transform (SIFT [18] method to localize the important regions containing discriminant information. Then, the proposed system extracts feature using the Adaptive Local Ternary Patterns (ALTP [19]) and apply the KNN in the classification task.

II. An effective face recognition system based on Deep Learning is presented. It used the pre-trained AlexNet-v2 [20] model on the ImageNet database to extract features. Then, it used the PCA to get the relevant descriptors and applied a robust SVM algorithm called Linear Support Vector Classification (LinearSVC) in the classification stage. Utilizing this proposed system enhances the performance of face identification.

III. A new robust face recognition system under illumination variation is proposed. It used four convolutional blocks from the VGG16 [17] model pre-trained on the ImageNet database to extract features. Then, it used a Global Average Polling (GAP) for the dimensionality reduction and the PCA to keep the relevant features. The system was evaluated under challenged benchmark databases and obtained a level of accuracy comparable to state-of-the-art methods.

IV. A new iris recognition system based on Transfer Learning is introduced. The proposed system used Yolov4-tiny [21] to detect the iris region. Then, it extracts features from the detected region without iris segmenting and normalizing. In addition, we proposed an effective deep CNN model for iris recognition that saves training time and does not require massive data. The proposed system advanced the state-of-the-art iris recognition accuracy on all the benchmark datasets used in the evaluation process.

V. Three effective Face-Iris multimodal biometric systems based on Deep Learning are proposed. We propose a new multimodal biometric system that uses image-level fusion to fuse the face and iris traits . To our knowledge, this is the first proposed face-iris multimodal system that utilizes Deep Learning and applies image-level fusion. In addition, this thesis introduced a novel face-iris multimodal recognition system that applied the fusion at the feature level, which used Deep CNN to extract the features from the face and both irises. Furthermore, it proposed an effective face-iris multimodal system based on Deep Learning that applied the fusion at the score level. This proposed system improved the state-of-the-art accuracy of two multimodal benchmark databases.

VI. In order to evaluate the proposed systems' performance and compare them with state-of-the-art methods, this doctoral research used more than twelve benchmark datasets: five face benchmark datasets, five iris benchmark datasets, and two multimodal datasets.

1.3 Thesis organization

The rest of the thesis has been structured in two parts: the first one (Chapters 2, 3, and 4) is devoted to the state-of-the-art in the biometrics field and deep learning for computer vision. The second one (Chapters 5, 6, and 7) presented our contributions to the biometrics community. Below, we briefly show the content of each Chapter:

- Chapter 2 "The Biometrics": This Chapter defines biometrics, discusses its history and applications, and provides an overview of the biometric system's structure and operating modes. Additionally, it highlights some famous modalities used for user recognition and discusses the choice of biometric traits and limitations of unimodal systems. Furthermore, Chapter 2 presents the multi-biometric system as a solution to these limitations, illustrating the motivations and characteristics of the multi-biometric system.

- Chapter 3 " Background Study and Literature Review": This Chapter covers the basics of face and iris recognition, including their various stages, and briefly overviews recent research on the unimodal face and iris biometric systems. It also explores the face-iris recognition system and its processes. Furthermore, the Chapter reviews recent studies that have combined face and iris modalities.

- Chapter 4 "Deep Learning for Computer Vision": In this Chapter, Artificial Neural Networks are briefly introduced, and the focus then shifts to Convolutional Neural Networks, the main topic of the thesis. The Chapter answers the question "What is CNN and why is it used?" and provides an overview of the layers of a CNN. Additionally, popular deep CNN architectures are presented in this Chapter, along with a discussion of techniques used to enhance the performance of deep CNN models.

- Chapter 5 "Face Recognition System": In this Chapter, three face recognition systems are proposed, and their architecture is presented. The systems are then evaluated on benchmark face datasets and compared with current state-of-the-art methods to demonstrate their efficiency. The first proposed face recognition system used handcraft approaches, and the second is based on Deep Learning. The third system is proposed for recognizing faces under illumination variation and poor lighting.

- Chapter 6 "Iris Recognition System": This Chapter presented a new iris recognition system based on Deep Learning, detailing its architecture and evaluating its performance using four benchmark datasets. The system was also compared to current state-of-the-art methods to highlight its superiority.

- Chapter 7 "Face-Iris Multimodal Recognition System": This Chapter proposed three multimodal face-iris systems based on Transfer Learning; the first used image-level fusion, the second applied the fusion at the feature level, and the third at the score level. In this Chapter, the detailed design of each system was presented, followed by multiple experiments on various benchmark datasets to demonstrate the robustness of the proposition. The systems were evaluated on unimodal databases to show their effectiveness even when one biometric trait is missing or unavailable. In addition, the evaluation was conducted on real and chimeric multimodal databases, and the results were compared to state-of-the-art methods to prove their superiority.

1.4 Publications list

This research has resulted in numerous publications in international journals and conference papers.

1.4.1 Journal papers:

- **A. HATTAB and A. BEHLOUL**, “Face-Iris multimodal biometric recognition system based on deep learning”, *Multimed Tools Appl*, pp. 1–28, 2023. doi: https://doi.org/10.1007/s11042-023-17337-y

- **A. HATTAB and A. BEHLOUL**, “A robust iris recognition approach based on transfer learning”, *International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 1065–1080, 2023. doi: http://dx.doi.org/10.12785/ijcds/130186.

- **A. HATTAB and A. BEHLOUL** , “New Approaches for Automatic Face Recognition Based on Deep Learning Models and Local Handcrafted ALTP”, *ICST Transactions on Scalable Information Systems*, vol. 9, no. 34, pp. e11–e11, 2022. doi: http://doi.org/10.4108/eai.20-10-2021.171547.

1.4.2 Conference papers:

- **A. HATTAB and A. BEHLOUL**, “An illumination-robust face recognition approach based on convolutional neural network”, *In: Modelling and Implementation of Complex Systems: Proceedings of the 7th International Symposium, MISC 2022, Mostaganem, Algeria, October* 30-31, 2022, pp. 135–149, Springer, 2022. doi: https://doi.org/10.1007/978-3-031-18516-8_10.

- **A. HATTAB and A. BEHLOUL**, "A Robust Face Recognition Method Based on ALTP and SIFT", *In: Advances in Communication Technology, Computing and Engineering. RGN Publications; 2021 (pp. 155-169).* doi: http://doi.org/10.26713/978-81-954166-0-8.

- **A. HATTAB and A. BEHLOUL**, "Automatic Face Recognition System based on Data Augmentation and Transfer Learning", *In: 2023 International Conference on Theoretical and Applicative Aspects of Computer Science (ICTAACS)*

- **A. HATTAB and A. BEHLOUL**, and A. BOUDJEMA, “Automatic Face Recognition System Based on Fine- tuned Xception Model”, *In: First National Conference in Computer Science Research and its Applications RIA’23, Relizane, May 10, 2023.*

- A. BOUDJEMA, F. TITOUNA, **A. HATTAB**, and KH. DJOUZI, “A Light CNN Architecture for Human Activity Recognition Based Wearable Sensors”, *In: First National Conference in Computer Science Research and its Applications RIA’23, Relizane, May 10, 2023.*

- **A. HATTAB and A. BEHLOUL**, “An Efficient Iris Recognition System based on Fine-tuned Xception Model”, *In: the National Conference on Artificial Intelligence: From Theory to Practice, Tebessa.*

Part I

State-of-the-art

CHAPTER 2

THE BIOMETRICS

"Biometrics is the key to securing our digital lives and ensuring privacy in the digital age."

David Pollington, Biometrics Expert.

2.1 Introduction

In daily life, a person needs to get to know the caller before starting to converse with them because each one has certain limits in dealing with, and some topics are not discussed except with a specific person. The same goes for many computer applications that require identifying a user before interacting with him. Several applications use traditional proofs such as passwords, personal identification numbers (PINs), or ID cards to recognize users. Still, these proofs are very easy to steal, forget, share, manipulate, or lose. Biometrics have been introduced to replace traditional proofs with user physiological traits and behavior characteristics, such as the face, iris, fingerprint, ear, signature, and voice. These biometric modalities are generally stable, universal, unique, and invariant over time [22, 23].

This introduction Chapter presents the definition, history, and applications of biometrics; it also introduces the structure of the biometric system and its operation modes. In addition, we show some famous modalities used for recognizing users. Also, the selection of biometric traits and the limitation of unimodal biometric systems are outlined. Moreover, we will introduce the multi-biometric system as a solution to these limitations, where we illustrate the motivation and some characteristics of the multi-biometric system.

2.2 Biometrics Definition, History, and Applications

The biometric word has been created by concatenating two words in Greek, bio (life) and metric (to measure). In our context, biometrics is the science that deals with recognizing (identifying or authenticating) people based on their physiological traits like face, iris, and fingerprint or their behavioral characteristics like voice or signature [23, 24]. The biometric has been chiefly used in systems security because it is based on robust measurements that could effectively utilize for automatic user recognition. The history of biometrics extended into prehistory when the Qin Dynasty used the fingerprint in China to sign documents [25] . According to the Portuguese historian Joao De Barros, the Chinese merchants used ink to stamp children's footprints and palmprints on paper to differentiate young children. In addition, the Babylon merchants, 3000 years before JC, used fingerprints as a signature for commercial exchanges [26–28] . In 1880 based on empirical observation, Henry Fauld proved the uniqueness and individuality of fingerprints and their effectiveness for person identification [29]. To identify convicted criminals, the police desk of Paris developed the Bertillonage system that recognized criminals based on physical characteristics [30] . Also, in 1924, the United States Congress started collecting criminals' fingerprint images [31].

With the development of technology in the last decades in signal processing and image processing, several automatic person recognition systems have been proposed by using several biometric modalities. In 1966, Li et al. [32] proposed a system for speaker verification. In addition, with the growing fingerprint databases and technology development. In 1977, law enforcement agencies applied Automated Fingerprint Identification Systems (AFIS) to recognize a person based on the fingerprint. To improve the process of fingerprint identification, twenty-two years later, a partnership between the Federal Bureau of Investigation (FBI) and some other law enforcement agencies established the Integrated Automated Fingerprint Identification System (IAFIS) [33]. In the 1980s, a biometric system for signature verification [34] and a face recognition system [35] were proposed. In 1993, Daugman proposed the first algorithm used for iris recognition [36]. Furthermore, many recognition systems based on several biometric traits have been suggested, such as body odor [37], heartbeat [38], eye movement [39], and thermal infrared face [40]. With the recent development in machine learning and deep learning techniques, researchers developed several robust biometric systems that achieved great accuracy rates. Many biometric applications have been developed in the last few years to facilitate person recognition in daily living. Some of them have been illustrated in Figure 2.1, in which the biometric applications can be classified into four categories [41]:

- Data access control: like controlling the logging into PCs, smartphones, and networks.
- Physical access control: like controlling the access to military, government, and companies agencies.

- Validate a claimed identity: like controlling the access in a border environment and confirming the identity of a bank consumer.
- Identify persons: like recognizing persons in a surveillance video using face traits and identifying alleged criminals using fingerprints.

Figure 2.1: Some applications of biometrics

2.3 Biometric systems' modules and operating modes

The biometric system exploits the uniqueness of the physical and behavioral traits to identify persons; it is an automatic system based on two main phases for recognizing people. In the enrollment phase, the system used a special sensor that captured information from the biometric traits, and then it used different algorithms to extract features. Finally, it stored the features (template) extracted in a database. In the testing phase, the system captures the biometric trait and extracts features from this trait, and then compares the features with the template stored in the database. Based on the matching result, the system makes the final decision.

2.3.1 Biometric systems' modules

The biometric system process consists of four sequential modules. The enrollment phase uses only two, and the testing phase uses all sequential modules. In the following, we briefly explain these modules:

1. **Sensor or acquisition module:** In this module, the biometric systems use different sensors and cameras (fingerprint sensor, visible light camera, infrared camera, etc.) to collect the data from the biometric traits, where the data is collected in a digital format to be processed by the different algorithms. In recent years, several sensors have been developed for scanning traits with high quality to achieve more accuracy.

2. **Pre-processing and feature extraction module:** The system pre-processes the data captured by the sensor to improve its quality. The pre-processing process has been divided into three principal steps. First, the system evaluates the goodness of the collected trait, then based on a pre-defined threshold, the system asks the user to re-capture the biometric characteristic if the quality is terrible, but if the quality is acceptable, the system starts further processing. Secondly, the system starts a segmentation step where it separates the region of interest from the captured data to reduce the noise introduced by the sensor and ignore the unimportant regions. Finally, the system used enhancement algorithms to improve the data quality, like histogram equalization, logarithmic transformation, etc. In the feature extraction module, the system applied different techniques to extract effective features from the data; these techniques can be divided into two categories: handcraft methods such as local binary pattern (LBP), local ternary pattern (LTP), Gabor wavelets (GW), etc. The second category is the deep learning methods such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long short-term memory (LSTM), etc. The feature extraction module is the most important step in the process of biometric recognition. Therefore, the feature extraction method must be chosen very carefully.

3. **Matching module:** In this stage, the system calculates the matching scores by comparing the feature extracted from the user in the testing phase and the features stored during the enrollment phase. Several techniques can be used in this stage, such as hamming distance, Euclidean distance, etc.

4. **Decision module:** The system identifies the user based on the matching scores if we use a biometric identification system. On the other hand, the system rejects or accepts the user based on the scores calculated in the previous module if we use a biometric verification system.

   Figure 2.2 presents the global structure of the biometric systems.

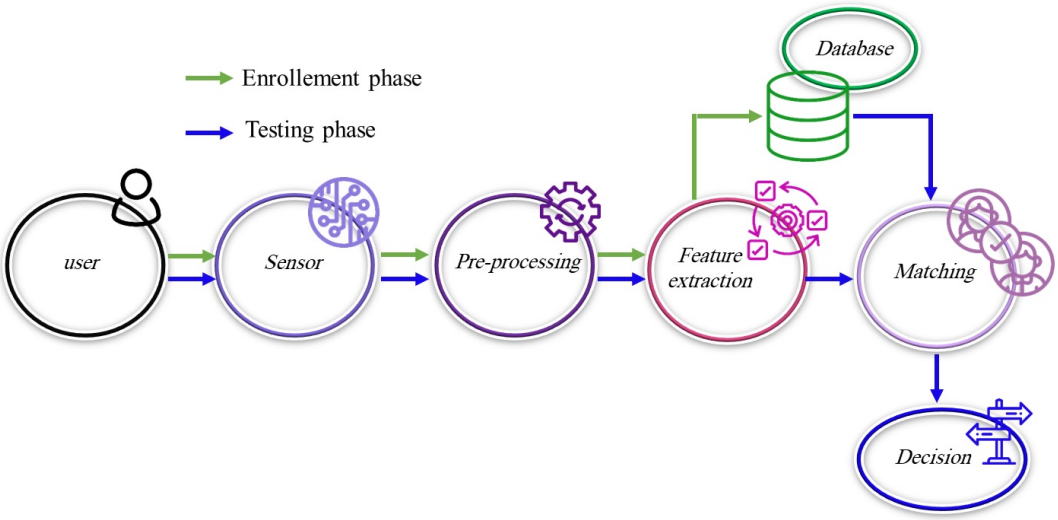


Figure 2.2: Generic structure of a biometric system

2.3.2 Biometric systems' operating modes

The biometric systems use three operating modes depending on the application context:

2.3.2.1 Enrollment

The enrollment operation is the primary step in all biometric systems, where it is an irreplaceable operation. During this operation, the biometric system stores a user's template (features extracted from the biometric traits) with his personal information (first name, last name, ID card) in a database. In both biometric recognition modes (identification and authentication), the systems used the biometric information stored in the database to recognize persons. Figure 2.3 illustrates the steps of the user's enrollment in the biometric system.

2.3.2.2 Authentication

In the authentication mode, the user inserts his identity (user name, ID card, PIN code, etc.) in the system; after that, the system captures the biometric trait and then confirms the claimed identity (authentic or imposter). Figure 2.4 illustrates the steps of authentication on a biometric system. Authentication is used to verify the identity of a person. For example, if a user with a banking customer ID (X) wants to withdraw money from an ATM, the customer inserts his credit card into the ATM. Then the user confirms his identity by presenting a biometric trait; for example, he places his fingerprint on the sensor. The system captures the biometric trait and extracts its features, then compares these features with the template of the banking customer ID (X) stored in the database during the enrollment mode; this comparison is called a one-to-one comparison. Finally, after the comparison, the system accesses the banking account of the

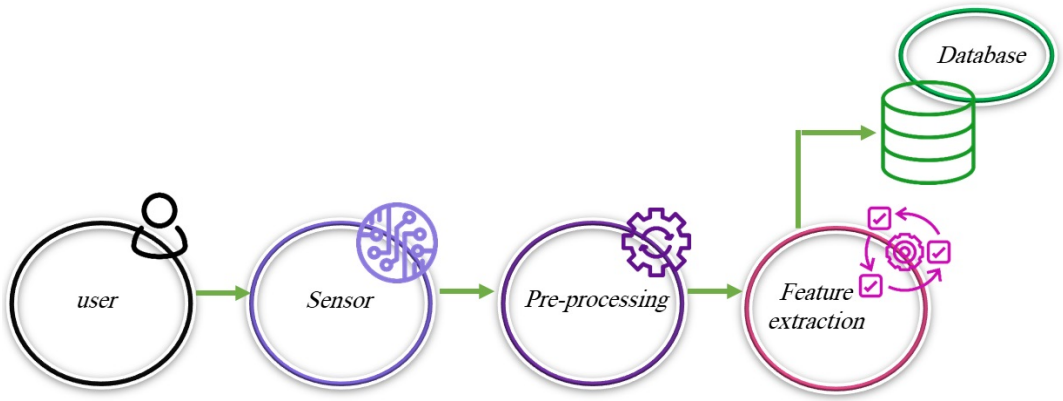


Figure 2.3: User's enrollment in the biometric system

customer ID (X) if the degree of similarity is high, or it refuses access if the degree of similarity is low.

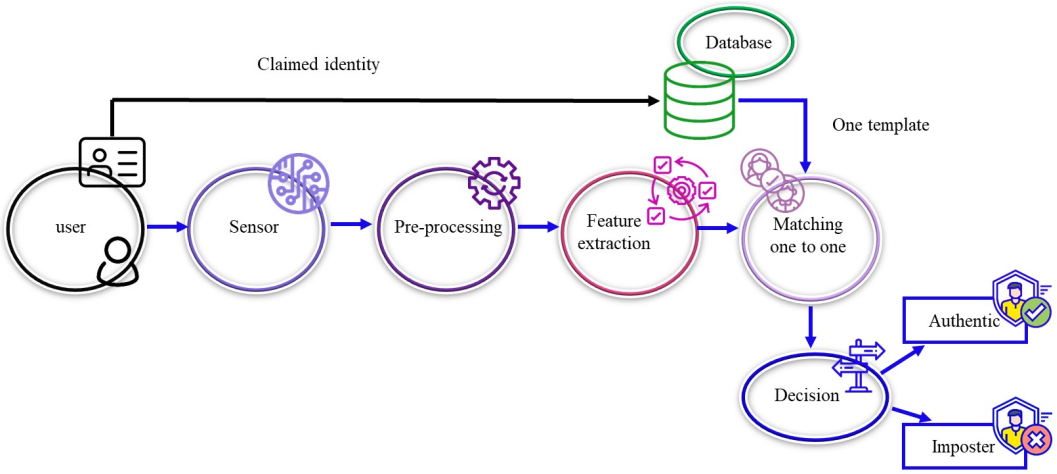


Figure 2.4: User's authentication in the biometric system

2.3.2.3 Identification

In the identification mode, the system captures the biometric traits of a person and extracts a vector of features from the data captured. Then, the system compares this vector with all the templates stored in the enrollment mode (one-to-N comparison). After the comparison, the system's output is the personal information (first name, last name, and ID card) stored with the template has the highest degree of similarity. Figure 2.5 illustrates how the biometric system identified a person. For example, a government company uses a biometric system to monitor

employees' entry and exit. The system captures the face of the employee and extracts features from the image captured. Then the system compared the features vector extracted with the templates stored in the database. Based on the template with the highest degree of similarity, the system defines the employee's identity. The identification mode is very complex, and its effectiveness is lower than the authentication mode [42]. This thesis discussed the problems of the biometric identification systems proposed in recent years and proposed effective biometric identification systems. In this thesis, when we use the term recognition, we mean identification because this thesis doesn't care about the authentication mode.

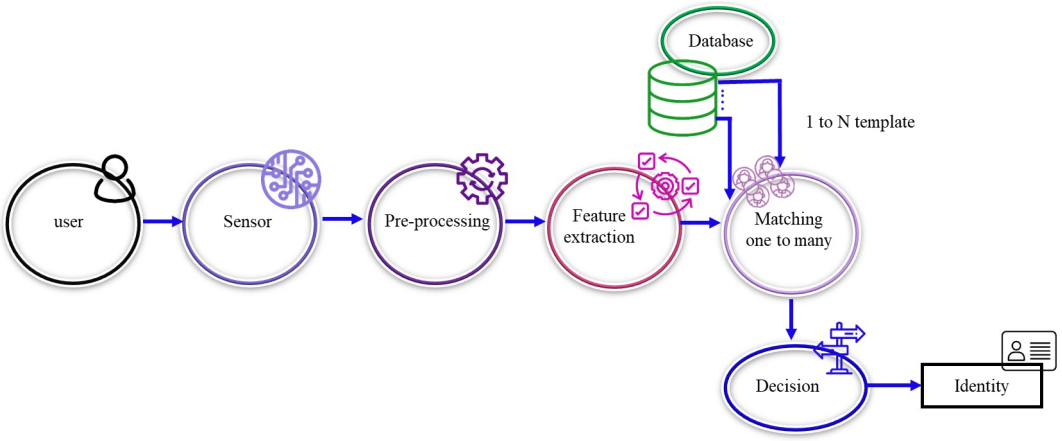


Figure 2.5: User's identification in the biometric system

2.4 Biometric Modalities

Several biometric systems have been proposed in the literature to recognize persons. This section briefly reviews the most important modalities used by these recognition systems. Before we start, we present in Figure 2.6 the modalities reviewed in this section, classified into physical, behavioral, and medico-chemical traits. In addition, we distinguished the traits used by the contactless recognition systems with the blue color because the users preferred these systems [43].

2.4.1 Face Recognition

Face recognition has become the most familiar modality used for personal identification in daily life [44]. The top two advantages that helped face biometrics to spread considerably are user acceptability and requiring no user cooperation. The user accepts the use of face images in the recognition process because the face is not a personal thing that should be hidden. In addition, the face image capture does not require the user's cooperation because the face is always

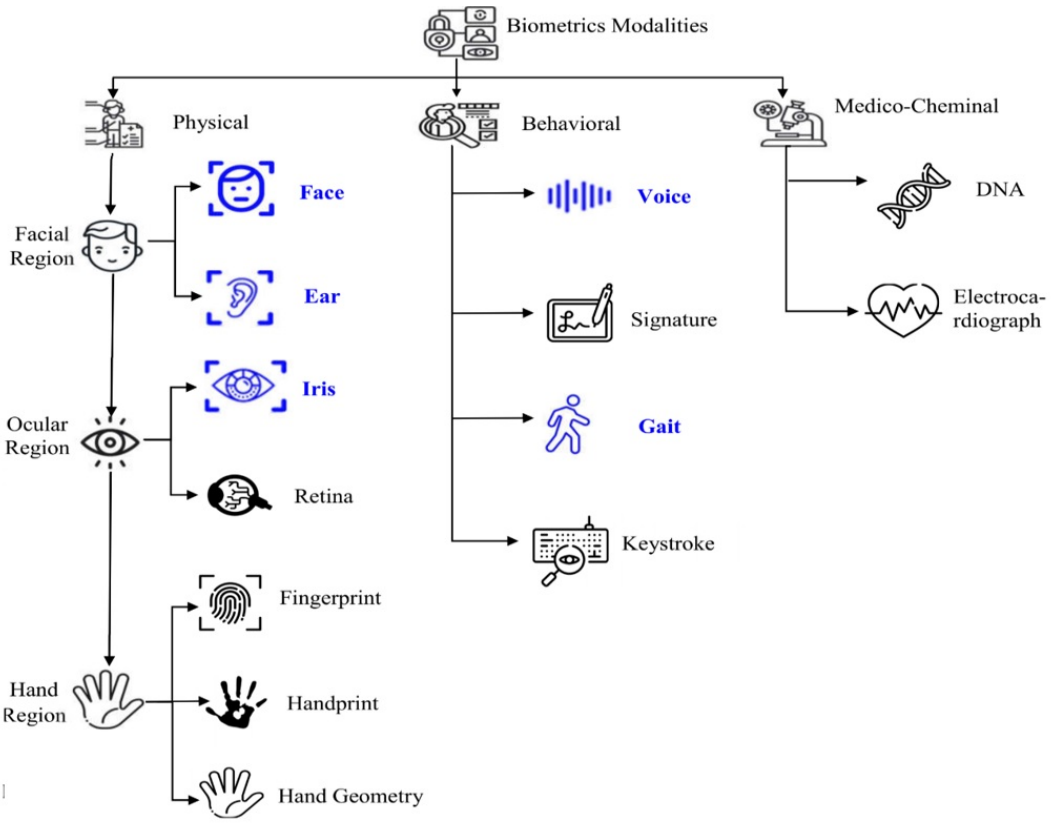


Figure 2.6: Biometric modalities classification, where the contactless biometric traits are shown in blue.

exposed. Face recognition achieved a high accuracy rate in the controlled environment. Still, this accuracy decreased in the constrained environment, where the variety of poses, illuminations, and expressions affects the face recognition performance.

2.4.2 Iris Recognition

Iris is one of the safest recommended modalities for recognizing persons [4]. The iris regions contain unique features that can be used even to differentiate between identical twins. Furthermore, with the development of camera sensors, we can capture the iris region at a distance with less user cooperation. Moreover, humans do not hide their iris regions in daily life, which facilitates the capture of this region. Still, the performance of iris recognition decreased if the images were captured under uncontrolled conditions such as illumination variation, blur, occlusions due to eyelid/ /glasses /eyelash, pupil dilation/constriction, and specular reflection.

2.4.3 Ear Recognition

Researchers also use the ear trait to recognize persons [45]. The uniqueness of the shape and appearance of the ear makes this technique more robust. In addition, the form of the ear is unchangeable with the growth and even aging of persons [46]. Also, the ear is resistant to the influences of the environment [47]. Ear recognition is based on matching the distance between the multiple small valleys and hills. Several research papers proved that the ear contains unique physiological features, which encouraged the researchers to be used it in many identification/ authentication biometric systems [48, 49]. Still, several limitations affect the ear recognition performance, like poor illumination conditions, hair or accessories covering the ears, and the variation in the viewing angles.

2.4.4 Retina Recognition

The retina is a hidden layer of the eye that receives the light from the cornea, then processes it and sends it as a neural signal into the brain for visual recognition. The retina contains blood vessels that form a unique pattern; this pattern is stable and unchangeable during the growth and even aging of persons. The retina is one of the most secure techniques used for person recognition, where the change or replication of this trait is very complicated [50]. Still, this technique requires the cooperation of users in the retina image acquisition stage, where it needs between 12 and 15 seconds of careful concentration, which inconveniences users. Generally, this expensive technology is applied to ensure a maximum level of security, where it is used in the military and government agencies [51, 52].

2.4.5 Fingerprint Recognition

The fingerprint trait has been used for hundreds of years to verify persons, in which the ancient Chinese used the fingerprint to sign documents [25]. With the development of technology, the fingerprint has become one of the most used modalities to authenticate persons automatically. The initial embryogenic development condition affects the formation of the fingerprints, making the fingerprint unique to each finger and person [53]. In our daily life, the fingerprint trait is applied to identify criminals. In addition, the fingerprint is also used for smartphone user authentication, where most smartphones contain a fingerprint sensor. The three main advantages of fingerprint recognition are: the fingerprint sensor is inexpensive, the fingerprint pattern is unchangeable with aging, and the ability to use more than one fingerprint to improve the system's accuracy rate [54]. On the other hand, the main disadvantage of fingerprint recognition is the users' cooperation, where the sensor can not scan the fingerprint without the user's collaboration. Also, some users fear the spread of germs due to touching the sensor.

2.4.6 Palmprint Recognition

Palmprint recognition is an ancient technique Chinese merchants used to differentiate young children [26]. The main cause of palmprint recognition success is the large variety of palm features. Furthermore, this technique achieved a high accuracy rate even when using a low-resolution scanner [55]. However, this technique requires the user's cooperation, and hand injuries create obstacles to recognizing persons. Besides, users do not feel safe when using this technique, especially with the spread of some viruses transmitted between people by touching surfaces.

2.4.7 Hand Geometry Recognition

Hand geometry is one of the most widely used techniques to recognize persons, where the first biometric system based on hand geometry was commercially available in the 80s. To effectively scan the hand geometry, the technology used two cameras; the first was used to capture the top of the hand, and the second was based on a mirror to capture the side of the hand. The hand geometry system recognizes persons based on the measurements extracted from the images captured, such as the hand's length, width, surface, and thickness [53]. The main disadvantage of this modality is that its performance decrease if it is used to identify persons from a large population because it is not very unique. For this reason, this type of system was mostly used to authenticate persons [46].

2.4.8 Signature Recognition

The signature is one of the most traditional ways to authenticate in commercial, legal, and administrative transactions. The signature is a behavioral biometric trait that emotional, psychological, and medical conditions can influence. In addition, this biometric trait can change over time. Furthermore, many people forget their signature if they do not use it for a long time.

2.4.9 Voice Recognition

The voice is a behavioral biometric trait that can be applied to recognize persons. The easy implementation and the low price characterizes voice recognition systems. Plus, most people accept using their voice in the recognition task. Still, several limitations affect voice recognition technology, like Background noise and throat disease. Moreover, the voice trait is not sufficiently distinctive to identify a person on an extensive database; for this reason, it is often used in the authentication mode.

2.4.10 Keystrokes Recognition

When it comes to the way of writing, people seem to draw the lines differently. Similarly, people type on the keyboard in their unique way. The features extracted from the keystroke trait are

not unique but offer sufficient discrimination to use in the person authentication field. We can extract several features from the keystroke trait, such as the sequence of the letter, the time of striking, the pauses, etc. Implementing keystroke recognition is very easy and does not require additional hardware such as scanners or cameras. Still, this trait can be applied only with an application that uses a keyboard. Moreover, emotional, psychological, and medical conditions and keyboard formats (QWERTY, AZERTY, etc.) can affect this type of recognition [56].

2.4.11 Gait Recognition

The technique that recognizes persons by their manner of moving and walking is called gait recognition. Most gait biometric recognition applications are used for person authentication because the gait trait is not very distinctive and cannot achieve a high accuracy identification rate, especially in a large database [46]. In addition, the gait characteristics can be affected by footwear, walking surface, and medical conditions (pregnancy, fatigue, and drunkenness) [57].

2.4.12 DNA Recognition

DeoxyriboNucleic Acid (DNA) is one of the most reliable biometric recognition traits. We extract the DNA from the body's cells, where we can extract from a fragment of hair, blood, urine, nails, saliva, or skin. The DNA trait is used chiefly in the applications of forensics to recognize persons. DNA achieved a high identification accuracy rate even when applied to large databases. Moreover, the DNA trait is unchangeable during growth, aging, and even after the person's death. Still, we cannot use DNA recognition in real-time applications because DNA analysis requires a long time. In addition, this person recognition technology is very costly. Also, the DNA biometric trait cannot distinguish between identical twins. Furthermore, the DNA biometric identity is straightforward to steal, which may cause problems, especially if thrown at a crime scene.

2.4.13 Electrocardiograph Recognition

Electrocardiograph (ECG) is one of the most promising traits used for recognizing humans; it is complicated to deceive this type of trait compared to other traditional traits [58]. In addition, it is possible to capture ECG through wearable devices like smartwatches or by placing the finger on the sensor [59, 60]. Still, besides the fact that ECG signals can change over time, heart rate variability and privacy issues may also pose challenges to this promising biometric trait [61, 62].

2.5 Biometric Modalities Selection

Many biometric traits have been proposed to recognize persons like face, iris, fingerprint, DNA, etc. Each trait has advantages and limitations and is suitable for a specific biometric application. For example, the DNA trait achieved an excellent accuracy rate. Still, most persons refuse to

use their DNA in the recognition, and we cannot use this biometric trait in real-time automatic person recognition. For these reasons, the DNA trait is only used in particular domains such as forensics applications. Jain, Bolle, and Pankanti [46] defined seven factors to evaluate the performance of biometric traits; these factors have been used successfully in the literature to select effective modalities; we summarize these seven factors as follows:

2.5.1 Universality or availability

This factor means that each person among the community members should own this biometric modality. For instance, a voice authentication system cannot be used in an application made for deaf people.

2.5.2 Distinctiveness or uniqueness

It is the most important factor in building a robust recognition system. This factor means that the biometric modality should be able to differentiate between all community members. For example, the iris biometric modality has a high distinctiveness factor where it can distinguish even between identical twins.

2.5.3 Permanence or robustness

This factor means that the trait resists under different conditions where it stays stable over time. For instance, the DNA trait is unchangeable during growth, aging, and even after the person's death.

2.5.4 Collectability or accessibility

This factor means that the biometric modality should be captured without disturbing the user or even without the user's cooperation. Plus, most important, the modality must be measured quantitatively. For example, the scanner of the retina trait is not comfortable for people, requiring more than 12 seconds of careful concentration. This factor is among the essential factors in selecting a biometric modality. Because users increasingly use systems that capture the biometric trait more comfortably.

2.5.5 Performance

This factor means the robustness and effectiveness of the biometric system under different conditions. The recognition accuracy rate, plus the speed and the cost of recognition, generally measure the performance of a biometric system.

2.5.6 Acceptability

The community members approve of collecting their biometric traits, where users present their biometric traits without fearing anything, especially their private life. This factor is critical to the success of any biometric system because persons' objection to giving their biometric trait inevitably leads to system failure.

2.5.7 Circumvention

This factor determines the degree of resistance against fraudulent methods. The imposter applies different fraudulent methods to cheat the biometric system to gain illegitimate access. In addition, the system uses different algorithms to discover spoofing. For instance, the fingerprint biometric system applies different algorithms to discover the fake fingerprint before starting the recognition process. Table 2.1 shows a comparison between biometric traits, which proves that each modality has some limitations. No ideal modality does not contain any disadvantages.

Biometric Modalities	Universality	Distinctiveness	Permanence	Collectability	Performance	Acceptability	Circumvention
Face [23]	H	L	M	H	L	H	H
Iris [23]	H	H	H	M	H	L	L
Ear [23]	M	M	H	M	M	H	M
Retina [23]	H	H	M	L	H	L	L
Fingerprint [23]	M	H	H	M	H	M	M
Palmprint [23]	M	H	H	M	H	M	M
Hand Geometry [23]	M	M	M	H	M	M	M
Signature [23]	L	L	L	H	L	H	H
Voice [23]	M	L	L	M	L	H	H
Keystrokes [23]	L	L	L	M	L	M	M
Gait [23]	M	L	L	H	L	H	M
DNA[23]	H	H	H	L	H	L	L
Electrocardiograph [63]	H	H	H	M	H	M	L

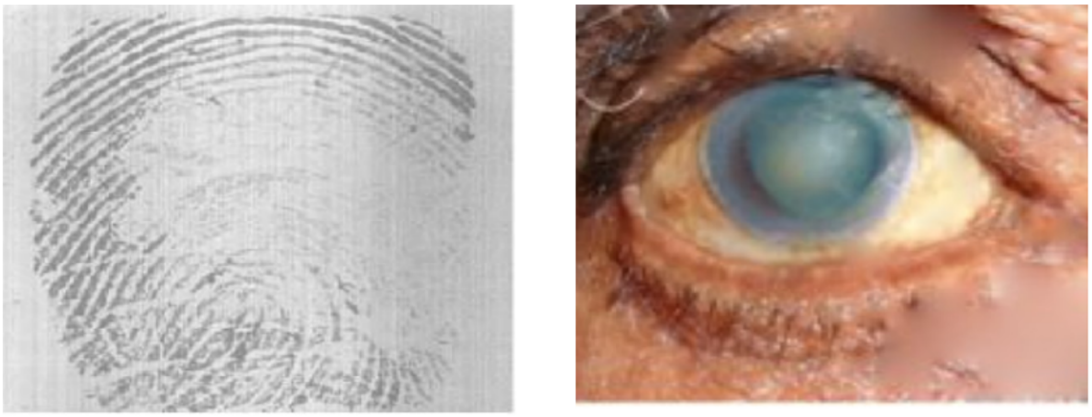
Table 2.1: Comparison between biometric modalities High, Low, and Medium are denoted by H, L, and M, respectively

2.6 Unimodal Biometric Systems Limitations

In recent years several unimodal biometric systems have achieved great success in many human recognition applications. Although that, the unimodal systems failed to meet the requirements of many sensitive governmental, military, and civilian applications because these types of systems have many disadvantages and limitations that affect the accuracy, performance, and reliability of these sensitive applications. Below, we summarized the most critical unimodal biometric systems limitations:

2.6.1 Non-universality

Certain categories of the population do not possess or cannot present a specific biometric trait. For example, the National Institute of Standards and Technology (NIST) reported that about 2% of the population could not use their fingerprint as a biometric modality, in which some people can't present their traits to the fingerprint scanner due to disabilities related to the hand and the absence of fingers. Plus, some persons have very oily or dry fingers [64]. Furthermore, eye abnormalities and long eyelashes can hinder the iris recognition system. Also, certain diseases can affect the performance of the retina recognition system. Figure 2.7 (a) shows a non-universal fingerprint due to finger dryness, and figure 2.7 (b) shows a non-universal iris image due to abnormalities eye.



(a) (b)

Figure 2.7: Two examples illustrate the non-universality: (a) a fingerprint of a dry finger and (b) an iris trait of an abnormalities eye.

2.6.2 Noisy sensor data

Noise in the biometric data reduces the accuracy rate of the biometric recognition systems [65]. The two main reasons for noises in scanned biometric data are uncontrolled environmental conditions and inadequate sensor maintenance. For example, poor illumination conditions affect

the accuracy rate of the face recognition systems; also, the noisy environment reduces the performance of the voice recognition systems. In addition, dirt accumulation on the handprint sensor or fingerprint scanner affects the goodness of the data scanned [66]. Figure 2.8 shows a noisy fingerprint due to the dirt of the sensor.

Figure 2.8: A noisy fingerprint image [67]

2.6.3 Intra-class variation

The samples collected from the user in the recognition stage are very different compared to those acquired from the same user in the enrollment stage. Generally, the incorrect user's interaction with the sensor is the main problem that produces intra-class variation. In addition, modifying the sensor's setting and using different sensors can cause intra-class variation. Figure 2.9 illustrates an example of face images with an intra-class variation.

Figure 2.9: Three face images illustrate an example of an Intra-class variation [68]

2.6.4 Inter-class similarities

This issue means the lack of uniqueness, where the traits extracted from different persons are very similar, which makes a large difficult to distinguish between these persons. For example, the face recognition system mostly fails to distinguish between twins. Figure 2.10 shows the similarities between two images of two twins.

Figure 2.10: Face images of two twins [69]

2.6.5 Spoof attacks

This issue concerns the possibility of fooling the biometric system using specific spoofing techniques. Behavioral traits are more vulnerable to attacks than physical traits [70]; for instance, the imposters fool the behavioral recognition systems by imitating the voices or reproducing users' signatures. In addition, many techniques have been used to spoof the physical biometrics recognition systems, like constructing an artificial fingerprint to fool the fingerprint recognition system or using a contact lens to fool an iris recognition system. Figure 2.11 shows some examples of biometric spoofing.

The abovementioned limitations degrade the unimodal biometric system's performance and recognition accuracy. To overcome these limitations, the researchers proposed multi-biometric systems [71, 72], where these systems use more than one source of biometric characteristics to improve biometrics performance.

(a) (b) (c)

Figure 2.11: Examples of biometric spoofing : (a) face spoofing by using a face photo, (b) fingerprint spoofing by using an artificial fingerprint, and (c) iris spoofing by using a contact lens.

2.7 Motivation for Multi-Biometric

The multi-biometric systems use multiple sources of biometric characteristics from the same user to overcome the limitation of the unimodal system, where the multi-biometric systems achieve butter recognition efficiency and a high level of security. Below, some advantages of the multi-biometric systems are motioned [73]:

- **Universality:** The multi-biometric system saves the system's universality. If one modality is unavailable or unacceptable by the system, the system uses the remaining modalities. For example, suppose a user has abnormalities in the eyes. If the iris recognition system can not enroll this user successfully, the availability of other modalities, such as the face, allows enrolling the user successfully.

- **More robust to noise:** Multi-biometric systems are robust against the problem of noisy data. The multi-biometrics systems use other biometric sources, such as multiple biometric traits and multiple sensors, to overcome the limitation of noisy data. For instance, in some environmental conditions, the system cannot use the voice trait to recognize persons; it uses another trait like face or iris to recognize persons effectively.

- **Solve the problem of intra-class variance and inter-class similarity:** Combining multiple traits can solve the problem of intra-class variation and inter-class similarity. For example, suppose the system cannot differentiate between two identical twins by using face images due to their significant similarity (inter-class similarity). In this case, the multi-biometric system uses another trait, such as the iris, that can easily distinguish between the two twins. In another case, if a high variation between the images captured in the enrollment stage and the images collected in the recognition stage due to beard, glasses, hat, or scarves (intra-class variation), the multi-biometric system can easily recognize the user based on another trait such as fingerprint or palmprint.

- **More secure (hard to spoof):** The multi-biometric systems have more excellent resistance to spoofing than the unimodal systems, where it is complicated to spoof multiple biometric traits at once [70].

2.8 Muli-biometric Systems

The multi-biometric systems achieved high success in the last years in the person recognition field, where the characteristics of these systems encouraged the researchers to use them in several critical applications. To build a multi-biometric system, we need to know the following three main factors: the fusion scenarios, the architectures, and the fusion levels of the multi-biometric systems. We will explain in detail the three factors in the following subsections.

2.8.1 Multibiometric Fusion Scenarios

In the last decades, researchers have used many sources of information to design multi-biometric systems. Ross and Jain [74] classified these sources into five scenarios; the first four use only one trait, and the fifth use more than one trait. Figure 2.12 illustrates the five scenarios. In addition, Ross and Jain [74] added a sixth scenario that combines two or more of the five scenarios mentioned below.

2.8.1.1 Multiple Sensors

The multiple sensors system uses two or more sensors to collect the data from a single trait. This scenario makes the system more effective because the various sensors capture diverse features. For instance, some systems use ultrasonic sensors and optical to collect more features from the fingerprint trait. In addition, in face recognition, some systems use a visible-light camera in conjunction with an infrared camera to make the system more robust.

2.8.1.2 Multiple algorithms

The multiple algorithms systems collect the data from a single trait using a single sensor, and then the system applies different feature extraction methods and/or numerous matching algorithms. For example, in the literature, some face recognition systems used PCA [75] (Principal Component Analysis) and LDA [76] (Linear Discriminant Analysis) techniques to extract features from the same face image.

2.8.1.3 Multiple Samples

This category is used to overcome the intraclass variation, where the system uses only a single sensor to collect multiple samples from the same biometric trait. Using more than one sample makes the representation of the modality more complete and makes the system more effective

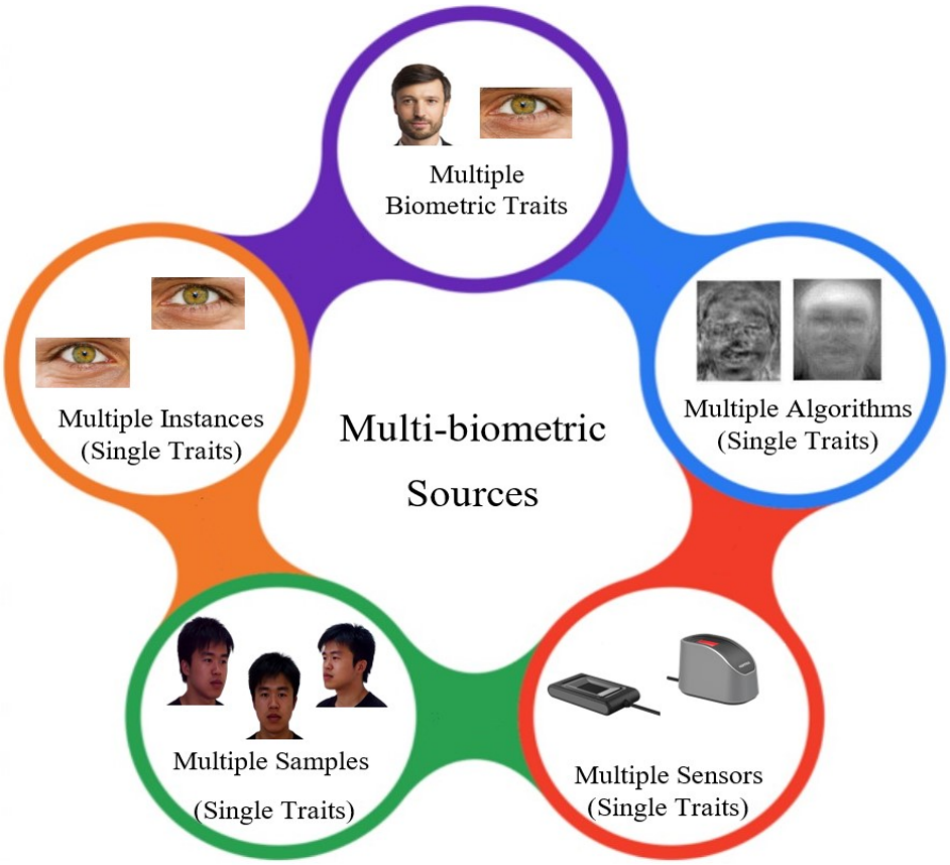


Figure 2.12: The five sources of information that the multi-biometric systems can use

[66]. For example, to overcome the challenge of face recognition under varying poses, some systems capture multiple samples from many angles, like the frontal, right, and left profiles.

2.8.1.4 Multiple instances

The same sensor captures multiple instances from the same biometric trait [77]. The most important advantage of this category is its low cost, where it does not require more than one sensor, a feature extraction technique, and a classification method. For example, using both irises to recognize persons or multiple fingers to verify a person's identity.

2.8.1.5 Multiple modalities (multiple biometric traits)

The systems of this category use multiple biometric traits for establishing identity. This category is also called multimodal biometric systems; it combines the evidence presented by more than one biometric trait to recognize persons. For example, some systems used the face and fingerprint traits to improve the recognition system's performance [78, 79]. In addition, others used the iris

and periocular modalities to prove the effectiveness of multimodal systems [80, 81]. This type of system has been receiving increasing attention for researchers because it increases the number of features by combining different biometric traits, which improves the system's accuracy rate [53, 82, 83].

2.8.2 Multi-biometric Architectures

After choosing the biometric traits used to build the multi-biometric system, the most critical challenge is determining which type of architecture we apply to design the system. Indeed, the architectures of multi-biometric systems can be classified into two fundamental modes: serial operation mode and parallel operation mode.

2.8.2.1 Serial Mode

In the serial mode, the multiple biometric traits do not capture simultaneously. The system captures and processes the data collected from the first trait. In this case, if the features extracted from this trait are not enough to recognize the user, the system captures another biometric trait and fuses them with the previous biometric traits (see Figure 2.13). The systems based on this mode sometimes make the final decision before capturing all traits used by these systems [84]. For example, in the fingerprint, palmprint, and iris authentication system, the user uses fingerprint verification; if the authenticate fails, the system captures the palmprint trait and then fuses the fingerprint and the palmprint; if this authentication also fails, the system fuse the iris with the previous traits [84].

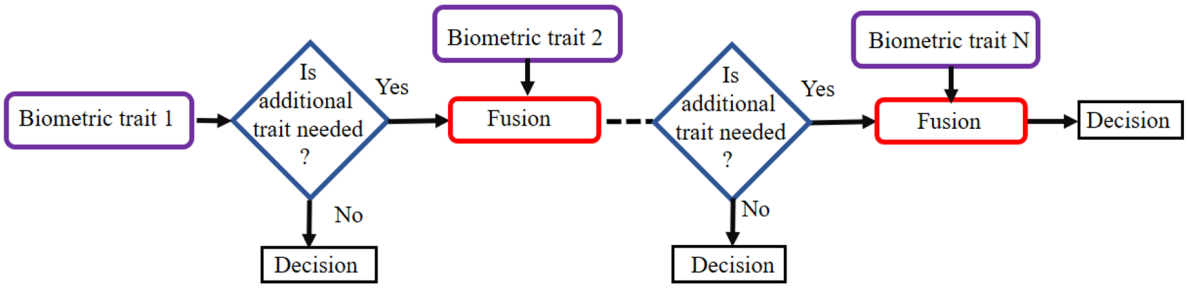


Figure 2.13: Architecture of fusion in series

2.8.2.2 Parallel Mode

This operation mode is the most used because all the available data are considered to make the final decision. So the system based on this operation mode collects all the required biometric traits and then processes them independently and concurrently simultaneously (see Figure 2.14). This operation mode is highly recommended for sensitive applications such as government and

military applications because it achieves high recognition accuracy and has a significant level of security compared with the serial mode [66]. The critical limitation of the parallel mode is the recognition time which is very long compared with the serial mode. Still, this limitation became less worrisome with the recent availability of powerful computing resources. For example, in the face-iris recognition systems relying on this mode, the systems capture the face and the iris, simultaneously using both biometric traits to recognize the person.

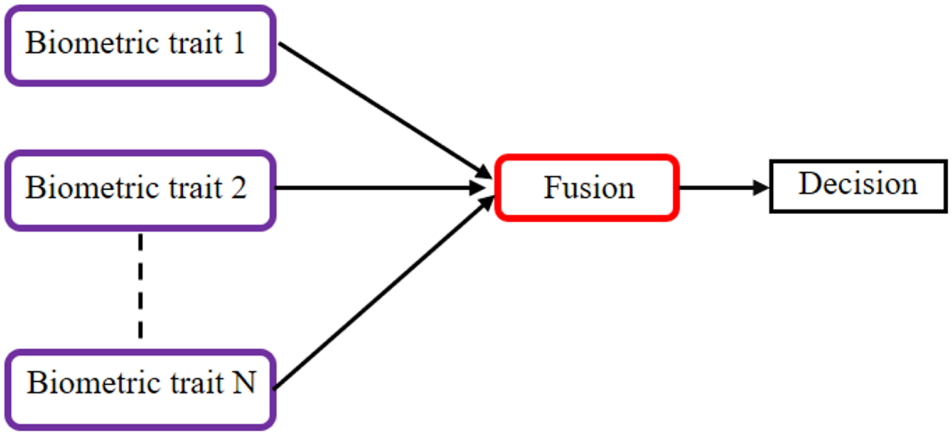


Figure 2.14: Architecture of fusion in parallel

In addition, another architecture called hierarchical fusion combines previous architectures, but the research community does not pay much attention to this type of fusion architecture [66].

2.8.3 Multi-biometric Fusion Levels

The biometric systems have been based on four main modules. In the data acquisition module, the sensors collect the data from the persons. Then, the second module extracts the features from the data collected. In the matching module, the matching algorithm compares the features extracted in the previous module with the features stored in the database and generates matching scores. The decision module uses these scores to recognize the persons. The multi-biometric systems fusion can rely on the information built by each module (see Figure 2.15). Also, based on the scores, we can generate other information 'Ranks' that the multi-biometric systems can fuse to identify the persons. In the following, we will explain in detail the different fusion levels.

2.8.3.1 Sensor Level Fusion

In this fusion level, the system combines the data captured by the sensors and then uses the fused data to extract features. For instance, we can fuse the face images collected by a visible-light camera and an infrared camera in the face recognition system. In addition, some studies proposed the fusion of data collected from different biometric traits. For example, to combine the face and

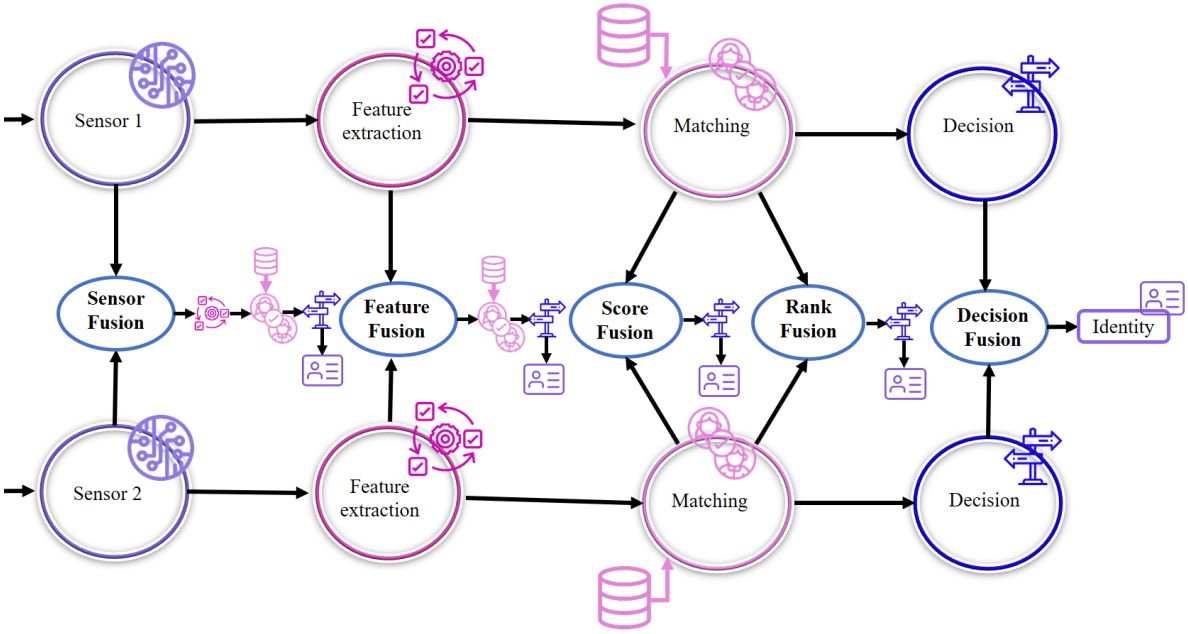


Figure 2.15: Five levels of fusion that the multi-biometric systems can use.

the palmprint images, a sensor-level fusion scheme has been proposed by Kisku et al. [85]. Also, Noushath et al. [86] used sensor-level fusion to combine the face and palmprint traits.

For image-based biometrics, sensor-level fusion is also known as data-level fusion or image-level fusion [87].

2.8.3.2 Feature Extraction Level Fusion

In feature-level fusion, the system creates a new feature vector representing the person by fusing the feature vectors extracted from multiple sources. The feature vectors can extract from multiple biometric traits, multiple sensors, multiple samples, multiple instances, or from the same biometric trait using different feature extraction approaches (e.g., handcraft and Deep learning approaches). The feature level has richer information about the raw data collected from the biometric trait. Thus, it is highly probable that fusion at this level achieves high success compared with different fusion levels [73]. However, some difficulties may affect the success of this type of fusion, such as the high dimensionality of the newly created feature vector; also, the incompatibility between the feature vectors poses some problems. Nevertheless, many normalization schemes are suggested in the literature to make the features used in the fusion compatible [88]. In addition, dimensionality reduction and feature selection techniques have been applied to reduce the dimensionality of the new feature vector [66].

2.8.3.3 Score Level Fusion

The score level fusion, also called the confidence level or the measurement level, is the most common level of fusion applied to build multimodal biometric systems [89]. It ranked third in information availability, where first place was gained by the sensor level and second by the feature level. However, it gained the best compromise between the availability of information and the ease of implementation [53]. In score-level fusion, each unimodal biometric system measures the similarity between the features extracted from the presented trait and the features stored in the enrollment stage based on different matching classifiers. Then, we normalize the obtained matching scores to achieve a good recognition accuracy. The most common normalization techniques are min-max and z-score. In addition, to calculate a single matching score, the score level fusion used an arithmetic operation like summation and product or a comparison operation like greater than and less than. Finally, based on the single matching score calculated, the multimodal biometric system takes the final decision. For example, after extracting the features in a face-iris multimodal authentication system, we compute the scores based on two different matcher scores. Then we use the min-max technique to normalize these scores and apply the product operation to obtain a single match score that can authenticate the user.

2.8.3.4 Rank Level Fusion

The rank level fusion can be used in the biometric identification systems where the output of each matcher in the unimodal system is a list of possibilities (list of enrolled users) sorted in descending order of confidence. Ho and al.[90] proposed three approaches to fuse the rank lists produced by the different unimodal systems: logistic regression, highest rank, and Borda count approach. The output of these approaches is a list of users; the first user in this list is the output of the multimodal system.

2.8.3.5 Decision Level Fusion

In this fusion level, each biometric trait is processed by a unimodal biometric system, and then we fuse the decision of the different unimodal systems to make the final decision. The decision-level fusion is also called abstract-level fusion because the information available at this level is very limited. The decision of the authentication system is only one information (accept or reject), and the output of the decision level in the recognition system is only the label of the class. Therefore, the fusion power in this level is lower than the previous levels [74]. Several methods have been applied in the literature to make the final decision, like OR, AND, majority voting [91], and Bayesian decision [92].

2.9 Conclusion

In this Chapter, we introduced the concept of biometrics and its application. Then we briefly presented the structure of the biometric system and its operation mode. Further, we give a brief view of the different biometric modalities that can be used to recognize persons. Moreover, we answered the question, how to select an effective biometric modality? In addition, we presented some limitations that affect the performance of biometric systems that use only one biometric modality. Most of the limitations of the unimodal biometric system can be reduced by combining multiple biometric modalities. Furthermore, we introduced the concept of a multimodal system, the source of information that this system can use. Also, the architectures and the levels of fusion that can be applied in this type of system have been presented in this Chapter.

As we mentioned in this Chapter, the most important step to building a robust biometric system is the selection of the best modalities. In this thesis, we have combined the face and both irises to design a robust multimodal biometric system. Besides, we chose the two traits because they complement each other and can be captured by using only a single high-resolution camera. In the next Chapter, we will present face recognition, iris recognition, and face-iris multimodal recognition; also, we give a brief view of previous works in these three fields.

CHAPTER 3

BACKGROUND STUDY AND LITERATURE REVIEW

"The secret of success in research is to see an old problem in a new way."

Harold Edgerton, Electrical Engineer.

3.1 Introduction

Researchers in recent years focused on multimodal biometric systems to overcome the limitation of the unimodal approach and build reliable biometric systems. The combination of face and iris modalities achieved an encouraging result in the literature because the face trait overcomes the limitation of the iris trait and vice versa [66]. In addition, with the potential spread of the COVID-19 pandemic and some other diseases by touch, users have become apprehensive over much about using touch-enabled biometric systems. Therefore necessitates proposing reliable biometric systems that use biometric modalities that can be collected without physically touching the users' body, like face and iris traits.

This Chapter introduces face recognition, iris recognition, and their stages. In addition, it briefly summarizes the previous most recent works related to the face and the iris unimodal biometric systems. Furthermore, it includes the face-iris recognition system and its process. Moreover, the Chapter discusses the relevant recent works that used face and iris modalities in the different fusion levels.

3.2 Face Recognition

Face trait is the most popular biometric modality used to identify humans automatically. The top two characteristics that contributed to the success of this biometric trait are: Firstly, the users accept the use of their face images in the recognition process. Secondly, the collection of face images does not require the user's cooperation. A considerable number of research works have been published in the last three decades to develop effective systems for face recognition [93] [94]. Despite that, the performance of most of these systems is dramatically degraded when the face images are collected in an unconstrained environment [95] . Moreover, facial expressions, ageing, and occlusions affect the face recognition system's performance.

3.2.1 Face recognition process

Three main steps are applied in the literature to build a face recognition system. The system's input is an image or a video, and the output is the label of the face. Figure 3.1 illustrates these fundamental stages required by each face recognition system. In the following, we describe these three steps.

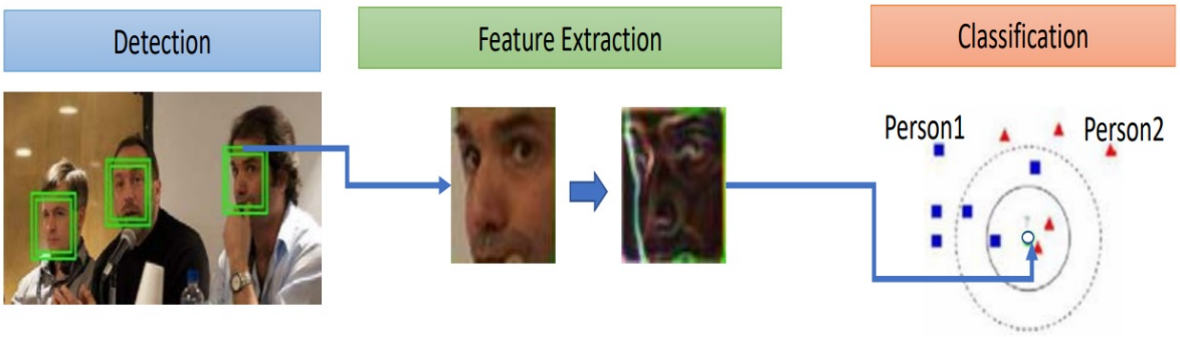


Figure 3.1: face recognition process [96]

3.2.1.1 Face detection

The facial recognition system's performance principally depends on the effectiveness of the face detection algorithm. In the face detection step, we locate the faces by creating bounding boxes in the image where each box contains a face. Several approaches have been published in the literature to detect faces, such as Viola-Jones [97] and Yolo-face [98]. Generally, the face recognition proposed systems do not focus on this step because the state-of-the-art face databases mostly contain cropped face images [96].

3.2.1.2 Feature extraction

The extraction of features is the most critical stage in face recognition, in which discriminant descriptors' extraction helps the system to achieve a high recognition accuracy. Many feature extraction approaches have been published in the literature that can categorize into two classes: Deep Learning approaches, and handcraft approaches. For more details about the classification of the facial feature extraction approaches, see Figure 3.2.

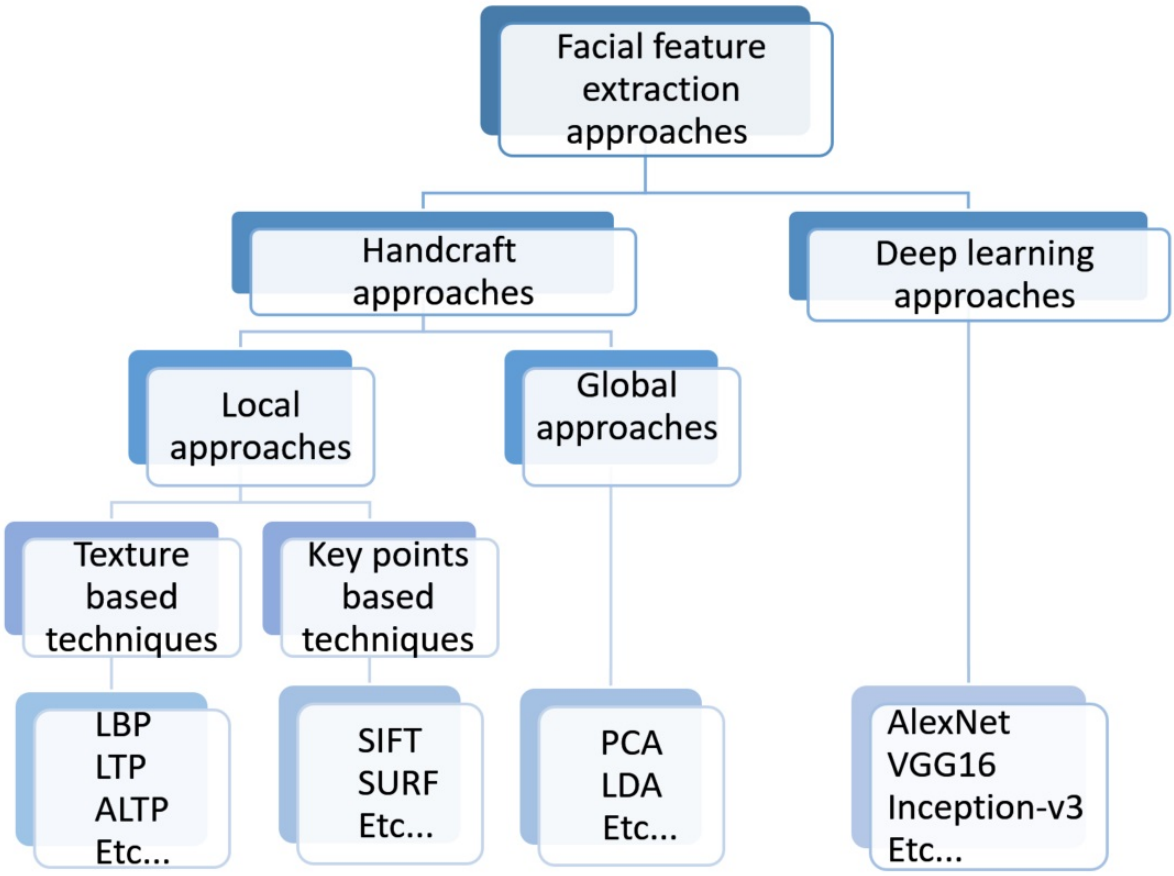


Figure 3.2: Taxonomy of facial feature extraction approaches

I. Handcraft approaches: Most face recognition systems proposed in the previous works have been based on the handcraft approaches to extract discriminant and effective feature descriptors. The handcraft approaches can be classified into two classes: local handcrafted-descriptor approaches and global handcrafted-descriptor approaches.

A. Local approaches: Several approaches have been applied in the literature to extract local features from face images. This sub-section focuses on five approaches (LBP [99], LTP [100],

ALTP [19], SIFT [18], and SURF [101]) that achieved high success in many face recognition systems.

1. Local Binary Patterns (LBP): is a robust feature extraction method based on local texture descriptors. The LBP was published in 1995 by Harwood et al. [102] to measure the local contrast. A year later, Ojala et al. [99] used the LBP technique as a texture descriptor. In the following years, many works have been published in several domains based on this technique, such as object recognition [103] and object detection [104].

The idea of the LBP is to calculate the relationship between the intensity of a pixel and his neighbouring. The resulting LBP of a centre pixel p_c can be represented in decimal form as:

$$LBP_{R,N} = \sum_{k=0}^{N-1} s(p_k - p_c)2^k, s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \tag{3.1}$$

Where p_k represents the intensity value of the neighbour pixels ofp_c, N sampling points of the circle whose centre is p_c and whose radius is R. Figure 3.3 shows an example illustrating how to calculate an LBP code where N=8 and R=1 (the direct neighbours). Moreover, Figure 3.4 shows an example of an LBP image and its histogram.

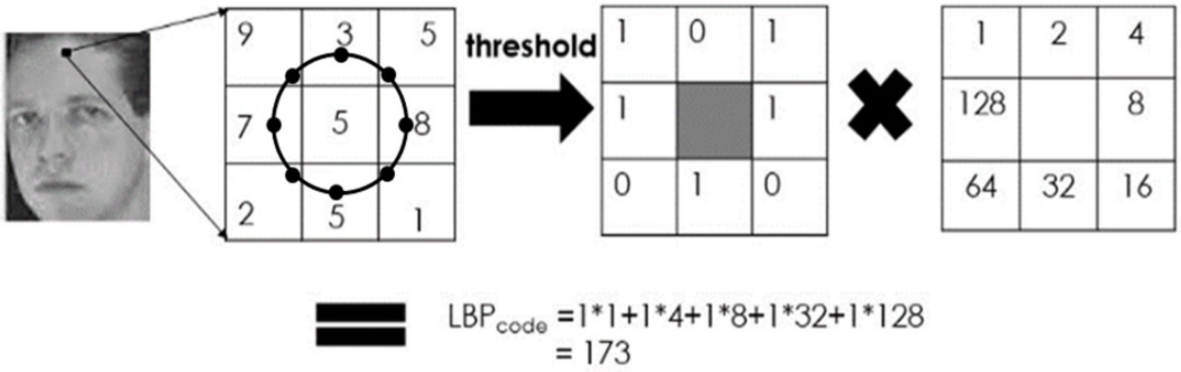


Figure 3.3: Example of LBP operator

2. Local ternary patterns (LTP): To overcome the noise sensitivity limitation that affected the LBP's performance, Tan and Triggs [100] proposed an LBP extension called LTP. The LTP used three values to calculate the Ternary code: 0 is assigned to the pixel if its grey-level value (u) is in the interval [$p_c - t$, $p_c + t$] where p_c is the central pixel's value, 1 is assigned if the grey-level value (u) of the pixel is upper than $P + 1$, and -1 is assigned in the rest case.

$$S(u, p_c, t) = \begin{cases} 1, u \geq p_c + t \\ 0, |u - p_c| < t \\ -1, u \leq p_c - t \end{cases} \tag{3.2}$$

The ternary pattern matrix is divided into two matrices (negative matrix and positive matrix). Based on both matrices, we calculate two LTP_{codes} (see figure 3.5). After scanning all the pixels

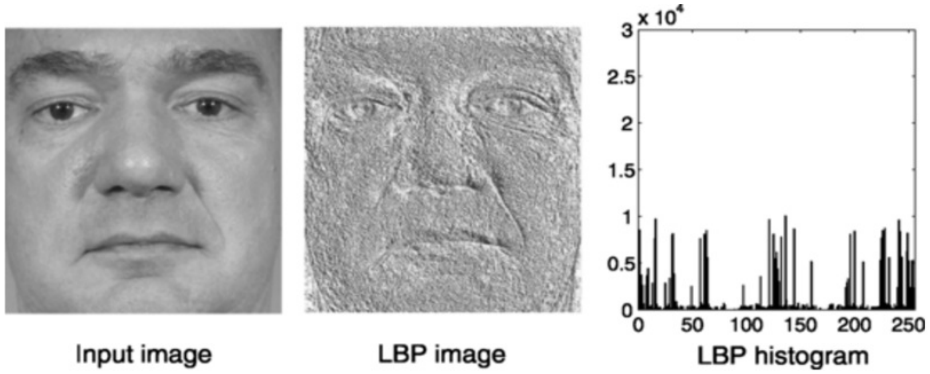


Figure 3.4: Example of LBP image and its histogram

of the image, we got two images. Finally, we calculate the histograms of the two images and concatenate them to obtain the final LTP descriptor for the image.

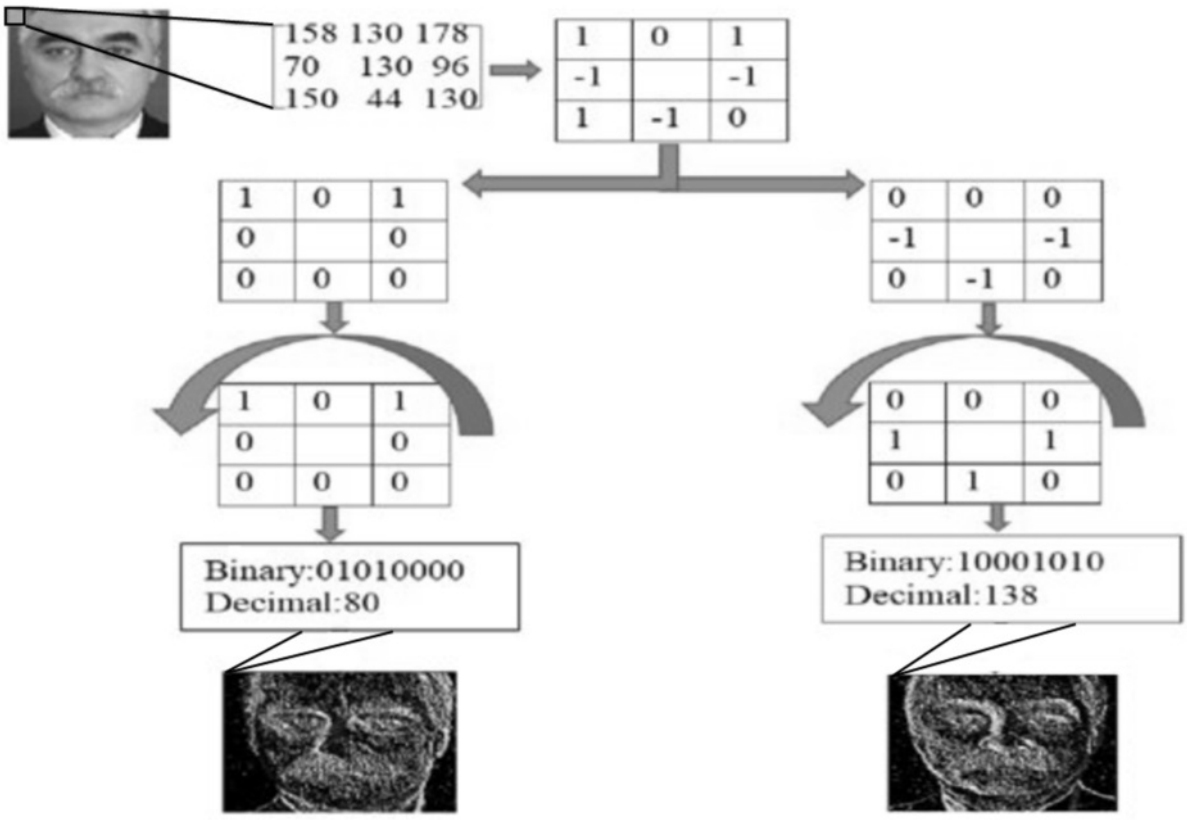


Figure 3.5: Local Ternary Pattern competition, adapted from [105]

3. Adaptive Local Ternary Patterns (ALTP): To overcome the noise's influence, the LTP uses a threshold (t). Still, this parameter's main problem is: the threshold is set manually, which

causes problems in applications that contain different data sets. To solve this problem, ALTP [19] proposed an automatic threshold inspired by Waber's law. The following equation illustrates how ALTP calculates the new threshold :

$$t = p_c * k \tag{3.3}$$

Where p_c is the central pixel's value, and k is Weber's law parameter. Like LTP, the ALTP count the histograms of the two images produced by the negative and the positive matrices. The obtained histograms are concatenated to obtain the final ALTP descriptor.

4. Scale Invariant Feature Transform (SIFT): is a practical approach proposed by Lowe [18] to detect and extract the local features (key points) and compute their descriptors. The SIFT is robust against illumination conditions, scaling, translation, and rotation of images. The SIFT feature extraction approach has four main stages:

a. Constructing the Difference of Gaussian Pyramid: At first, in this step, the image is represented in scale-space by convolving the image $I(x,y)$ with a variable-scale Gaussian $G(x,y,\sigma)$ as follows:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \tag{3.4}$$

Whereas $G(x,y,\sigma)$ is computed as follows:

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.5}$$

Secondly, we compute the difference between two nearby Gaussian scales (Gaussian images at scales: $k\sigma$ and σ):

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \tag{3.6}$$

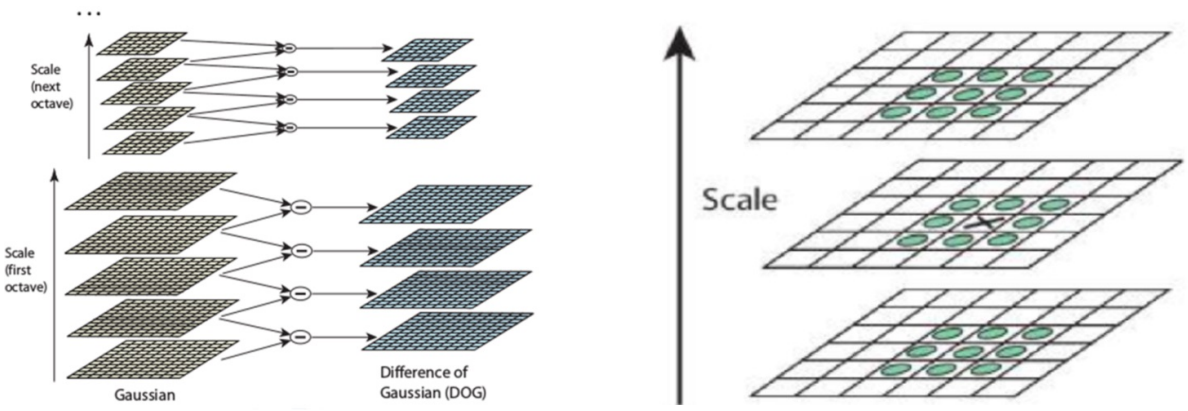
These differences are called Differences of Gaussians (DOG). Figure 3.6 (a) illustrates the construction of a DOG Pyramid.

b. Key point localization: In this stage, SIFT detects the Interest points (key points) by determining the local minima and maxima of the DOG pyramid, in which each pixel in the DoG images is compared to its direct neighbours in the current image and (3×3) neighbourhood in adjacent scales (see Figure 3.6 (b)). We call an extremum to point if its value is the minimum or maximum of the 26 neighbourhoods.

c. Orientation assignment: The SIFT assigns a magnitude and orientation to each interest point based on a neighbourhood around this key point (see Figure 3.7 (a)).

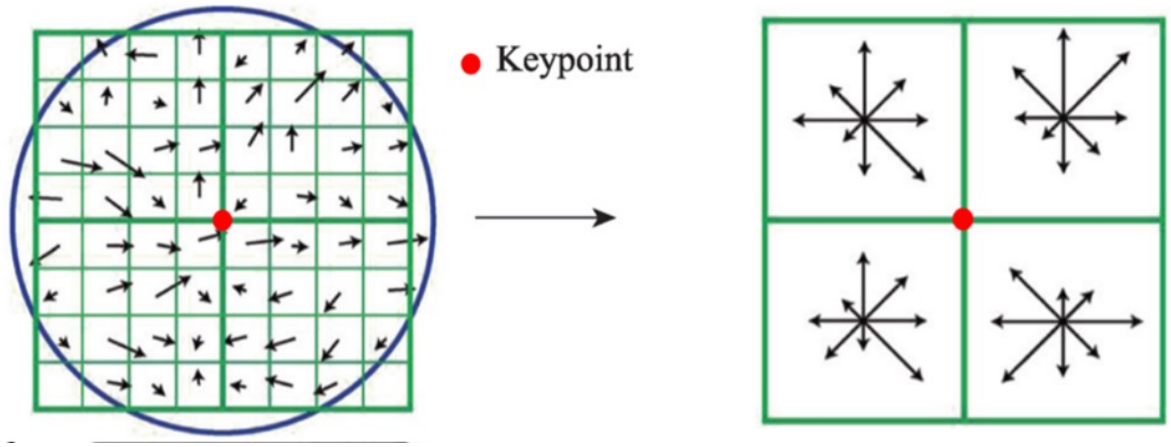
d. Key point descriptor generation In each point around the interest point areas, we calculate the global gradient magnitude and orientation (see Figure 3.7 (b)).

4. Speeded-up robust features (SURF): Relying on the SIFT descriptor, Herbert Bay et al.[101] constructed SURF. The SURF's performance is the same as the SIFT, but its complexity is significantly reduced [106].



(a) (b)

Figure 3.6: SIFT extremums detection: (a) Creation of DoG pyramid (b) Detect extremum in DoG pyramid.



(a) (b)

Figure 3.7: SIFT feature descriptor: (a) Neighborhood gradient direction. (b) key point descriptor.

The SURF used the determinant of the Hessian blob detector to obtain faster feature extraction. In addition, SURF used a second-order Gaussian Filter to approximate the image's partial derivatives because the mentioned filter allows scale and space analysis. Moreover, the Haar wavelets responses' sum is used to describe the feature of a key point.

B. Global approaches : Many global handcraft feature extraction approaches have been applied in the last three decades to create automatic face recognition systems, in which the Principal Component Analysis (PCA) and the Linear Discriminant Analysis (LDA) achieved great performance in this task.

1. Principal Component Analysis (PCA): is a famous statistical unsupervised technique to extract features and reduce the data's dimensionality. PCA was proposed by Turk and Pentland [75] in 1991.

The original paper of the PCA [75] introduced a new term called "eigenface" used in the face recognition field. The PCA has been applied on a large face image set to project every face image in a lower dimensional space, where each image is transformed into a small vector of weights called eigenface.

The main idea of the PCA is to build a matrix by converting each image into a vector and fusing these vectors to create a matrix. Then, PCA calculates the covariance matrix. In addition, PCA computes the eigenvalues of the covariance matrix and computes the eigenvectors of the eigenvalues. The PCA used these principal components (eigenvalues) to reconstruct the face image in a lower dimensional space.

2. Linear Discriminant Analysis (LDA) is a supervised statistical technique applied to extract features and reduce the dimensionality of data proposed by Belhumeur et al. [76] in 1997. The LDA or the Fisherface is similar to the PCA, but it is commonly applied to solve supervised classification problems. The LDA needs a labelled dataset (supervised learning) to maximize the variation interclass and minimizes the variation intraclass, unlike the PCA, which does not require class labels (unsupervised learning) to maximize the variance in the dataset. Compared to the Eigneface, the Fisherface is robust against facial expressions and illumination variation but needs more processing time and significant storage resources.

II. Deep Learning approaches: The computer vision field, especially the face recognition task, has improved dramatically in the last decade thanks to the significant development in the Deep Learning (DL) field.

DL, specifically the deep Convolutional Neural Network (CNN), is adopted to receive the images and extract their in-depth features.

CNN is an end-to-end architecture applied to solve the classification and the regression task, in which it uses convolutional layers to extract features and some fully connected layers to predict the label or the variable. We can use CNN as a feature extraction technique by keeping the convolutional layers and removing the fully connected layers.

Several deep CNN models have been proposed in the last decade, such as AlexNet [107], VGG [17], GoogleNet [108], ResNet [109], and Xception [110]. Deep CNN models achieved a high accuracy rate in many computer version tasks, in which some models outperformed humans' capacity in the face recognition field [111] [112].

Due to the importance of Deep CNN in developing many computer vision systems, we will dedicate the next Chapter (Chapter 4) to detailing the component of the CNN architecture and reviewing some famous Deep CNN models.

3.2.1.3 Classification

Several classifiers have been used in the literature to augment the pattern recognition systems' performance. In the following, we presented three robust classifiers that achieved extraordinary

success in the face recognition field.

I. Nearest Neighbor (NN): is one of the first algorithms proposed to solve supervised classification problems. It characteristics by its simplicity compared with other classifiers, which did not apply any training. The Nearest Neighbor (NN) calculates the average distance between the features extracted from the testing image and the features of each training image. Finally, the Nearest Neighbor assigns the label of the training image that has the minimum distance to the testing image.

The k Nearest Neighbor (k-NN) is a famous technique based on the Nearest Neighbor classifier widely used in literature. The k-NN takes 'k' training images that have the minimum distance from the testing image and selects the majority label among these images. Figure 3.8 represents the k-NN algorithm with k=3 and k =5.

In the literature, several distances are used by the k-NN, like Hamming distance, Cosine distance, Euclidean distance, and Chi-square distance.

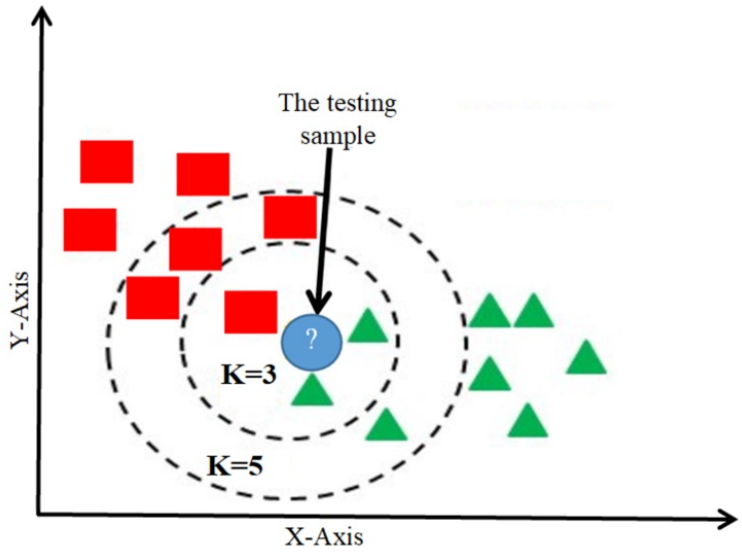


Figure 3.8: Representation of the KNN algorithm with K=3 and K=5

II. Support Vector Machines (SVM): is a robust supervised technique based on building separating hyperplanes, which is applied in the regression and classification analysis.

The SVM was originally proposed in 1963 by Vapnik [113] to create a linear classifier. After that, SVM used the kernel method to create non-linear models with a higher dimension. The SVM classifier seeks to build the best line or decision boundary that helps us to segregate a space with n-dimension into classes relying on the training data. In the future, based on this segregation, we

can put the testing data in the correct class. These best lines or decision boundaries are generally called hyperplanes.

The Support Vector Machine term comes from the extreme points used to create the hyperplane, which these points are also called support vectors. The SVM seeks to maximize the margin between classes by maximizing the margin that separates the hyperplane and the support vectors. Figure 3.9 illustrates the components of the SVM.

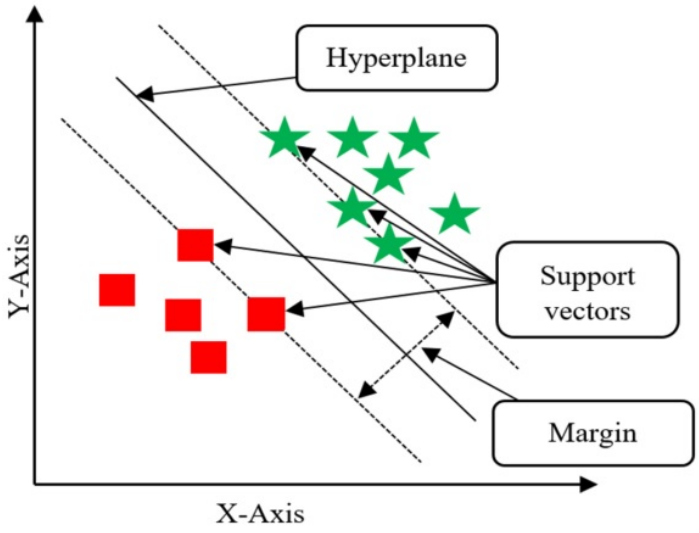


Figure 3.9: Components of SVM

III. Artificial Neural Networks (ANN): is a supervised computational system applied to solve regression and classification; it mimics the biological neural networks that build animals' brains. The ANN consists of three layers: input, output, and a layer between them called the hidden layer. Each layer contained a set of neurons connected to the next layer's neurons by weighted links. Figure 3.10 illustrates the components of the ANN. The input layer receives the data, process them by applying an activation function, and passes the results to the hidden layer. Generally, the ANN uses more than one Hidden layer to solve complex problems. The hidden layers applied a weighting function to the inputs followed by an activation function, then passed the results to the next layer's nodes. The output layer processes the last hidden layer's outputs based on a set of weights and its own activation function before delivering the final decision. The next Chapter (Chapter 4) presents More detail about ANN.

3.2.2 Face recognition state-of-the-art

Recently, face recognition has received extensive attention, and many papers have been published in this research field. In this section, we reviewed relevant and interesting works, starting with

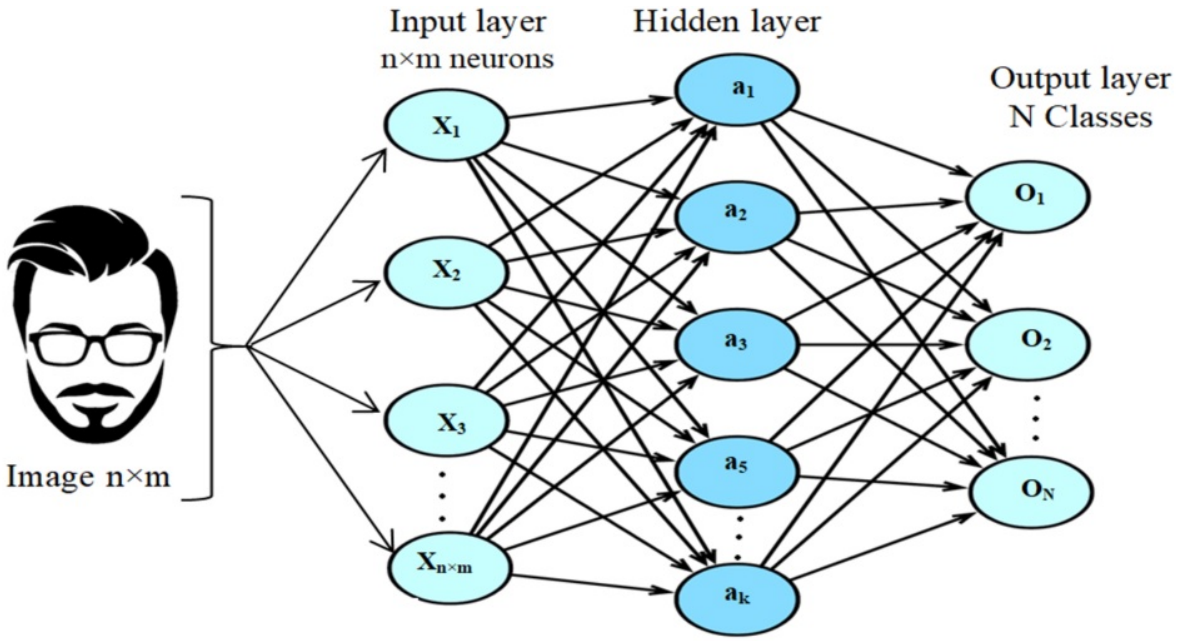


Figure 3.10: ANN architecture

the methods that used handcraft approaches and then that applied deep learning.

Ouyang et al. [114] proposed a hybrid method for automatic face recognition; the authors combined a new proposed algorithm called Improved Kernel Linear Discriminant Analysis (IKLDA) and Probabilistic Neural Networks (PNNs). Initially, the authors used the IKLDA algorithm to reduce the dimensionality and retain only the relevant information. Then, the IKLDA-PNN hybrid algorithm is used to recognize faces. The proposed algorithm improved the computational efficiency and achieved encouraging results with the small datasets, achieving an accuracy rate of 97.22% on the ORL database. However, its accuracy rate dramatically degraded when the number of training samples was very low. Ousliman et al. [115] proposed an effective descriptor for texture classification called Rotation-Invariant features. The new descriptor is mainly based on Directional Coding (DC). The Rotation-Invariant features integrated into the face recognition field and achieved acceptable results. Sapijaszko et al. [116] proposed a new face recognition system that extracts features from enhanced face images. The system is based on the two-dimensional DWT (Discrete Wavelet Transform) and the two-dimensional DCT (Discrete Cosine Transform). In the classification stage, the authors applied a multilayer sigmoid neural network. Qin et al. [117] applied a non-linear transformation that converted the pixel's intensity value into an alternative representation. The newly proposed method reduced the image's dimensionality and enriched the similarity intra-class. The accuracy rate of this idea reached 95% on the ORL face database. In addition, it reached 76.50% on the Feret database.

Yadav et al. [118] introduce a novel face recognition system robust to lighting variation. Firstly,

the system reduced the illumination effect by applying an effective filter called Homomorphic Filter (HF). Furthermore, the system adapts the intensity value of each pixel with its neighbour by applying the Reflectance Ratio (RR). Secondly, the system extracted features based on the Discrete Wavelet Transform (DWT). Finally, the system reduces the dimensionality of the extracted features using PCA before applying the k-Nearest-Neighbor (KNN) in the classification task. In another proposed system, Yadav et al. [119] applied the Reflectance ratio (RR) to cancel out the illumination deviation, and then Contrast Stretching (CS) was used to normalize the lighting variation. The system used the Integer Wavelet Transform (IWT) method to extract features and Fisher Linear Discriminant Analysis (FLDA) to determine the relevant features. In the classification stage, the authors used k-NN based on cosine angle and Euclidean distances.

Moreover, Yadav et al. [120] proposed a robust system for face recognition under lighting variation. Firstly, the system applied a proposed framework to annul the lighting variation. Secondly, a robust feature extraction technique relying on the Discrete Wavelet Packet Transform (DWPT) domain is applied to the Reflectance Ratio and Histogram Equalization (RRHE) images. Finally, the system used PCA for dimensionality reduction and the Nearest Neighbour in the classification stage. The LDA is one of the robust techniques proposed in the face recognition field. Still, its performance dramatically degraded under uncontrolled environments. Wen et al. [121] proposed a new version of the LDA called Robust Sparse Linear Discriminant Analysis (RSLDA) used to extract face descriptors. The new method proved its robustness in the experiments carried out by the authors on five databases captured under an unconstrained environment.

Yuan et al. [122] developed a novel system to extract features from face images. The system combines Collaborative Representation (CR) and discriminant analysis. Specifically, the system is based on two discriminant analysis techniques, Fisher Discriminant Analysis (FDA) and LDA. The presented idea achieved a good recognition accuracy on the small dataset. Hu et al. [123] introduced a novel framework to extract latent low-dimensional descriptors from high-dimensional data. The Joint Sparse Locality-Aware Regression (JSLAR) framework is mainly based on the non-squared L2 norm to improve local intra-class compactness. The proposed framework proves its robustness in several classification tasks like object and face recognition. In another work, Ayyavoo and Suseela [124] developed a face recognition system based on new illumination pre-processing method. Firstly, the author used Gamma Intensity Correction (GIC) to enhance the face image. Secondly, the 2D DWT divided the enhancing image into low-frequency and high-frequency coefficients. Finally, the system applied Contrast Limited Adaptive Histogram Equalization (CLAHE) to the low-frequency coefficients before using the Gabor Magnitude (GM) to extract features.

Dahmouni et al. [125] used multiple methods to build a novel face recognition system for education applications. First, the authors extracted face features based on the Local Gradient Probabilistic Pattern (LGPP), then used 2D DWT, 2D PCA, and 2D LDA for feature dimensionally reduction. In addition, the system used SVM and ANN in the classification stage. The system has

been evaluated on small faces databases and achieved an acceptable result raching 98.45% on the AR dataset. A new method for face recognition based on the region of interest has been introduced by Aldhahab et al. [126]. The authors localized and cropped both eyes, mouth, and nose from the face images, then applied two algorithms to extract features from this important facial region (Vector Quantization and DWT). Finally, In the classification stage, the Euclidian Distance is applied to check the similarity between the localized region and the training part. Complete Gabor Filter with Random Forest (CG-RF) is a hybrid face recognition technique introduced by See and Noor [127] for face recognition. CG-RF used Magnitude Responses (MR) of a Gabor filter and Oriented Gabor Phase Congruency Image (OGPCI) [128] in the feature extraction stage. In addition, Random Forest was applied in the classification stage. The proposed hybrid technique achieved acceptable results on the FERET and Georgia tech databases. Elaggoune et al. [129] combine many handcraft methods to extract features from the face image. In addition, they applied the cosine distance metric in the classification stage. As a result, the authors achieved a good accuracy rate on the ORL database, reaching 98.5% by combining the Gabor filter, Histogram of Oriented Gradients (HOG), and Particle Swarm Optimization (PSO).

Multiple face recognition systems based on Deep learning have been developed in recent years. Min et al. [130] introduced a system for face recognition that used only a single sample per person in the training stage. The system used the Transfer Learning technique to recognize faces, in which a Deep CNN model was trained on a comment multi-sample face dataset, and then the pre-trained weights were used to a target data set. Furthermore, the system proposed a sample expansion method to enrich intra-class variation information. The proposed method is called K Class Feature Transfer (KCFT). The system achieved an encouraged accuracy rate on the ORL database, reaching 97.77%. Hosgurmath et al. [131] developed a face recognition system based on CNN. The authors proposed a CNN architecture used to extract face features, in which the proposed CNN contains three convolution layers and three max pooling layers. In addition, it applied a batch normalization layer after each convolutional layer. After extracting 2048 face features, the classification is carried out using Linear collaborative discriminant regression classification (LCSRC) [132]. Tamilselvi et al. [3] propose a system for face recognition under uncontrolled environmental conditions based on Hybrid Robust Point Set Matching Convolutional Neural Network (HRPSM_CNN). Firstly, the authors overcome the limitation of noise sensitivities by applying the Laplacian of Gaussian (LoG) filter. Secondly, the Viola-Jones algorithm is used for face detection. Then a hybrid approach containing CNN and Robust Point Set Matching is used to extract features from the detected face images. Finally, the proposed system outputted the label prediction as audio because the system is proposed for a visually challenged person.

Elaggoune et al. [129] introduced a new face recognition system based on transfer learning. At first, the system used the ImageNet pre-trained AlexNet model to extract features from the faces. Then, it used the Particle Swarm Optimization (PSO) algorithm to keep only the relevant face characteristics. Finally, in the classification step, the system used Softmax. A face recognition

system that combines handcraft and Deep learning approaches was proposed by Yallamandaiah et al. [133] in 2022. The authors proposed a CNN architecture containing three convolution layers, each layer followed by a max-pooling layer. In addition, they used two fully connected layers. The system extracts end combines the features extracted by the Histogram of Oriented Gradients (HOG), the Local Binary Patterns (LBP), and the proposed CNN. Finally, the system used SVM to predict the label's face image.

Bendjillali et al. [2] developed a new system that used the Viola-Jones face detection algorithm to localize the face region. Then, before extracting features, the system applied the novel Modified Contrast Limited Adaptive Histogram Equalization (M-CLAHE) algorithm to enhance the face region. In the feature extraction and classification stages, the authors used three famous deep CNN models VGG16, ResNet50, and Inception-v3. The proposed system achieved an average accuracy rate of 98% on the Extended Yale B database. Yang Et al. [134] proposed a new Deep Reconstruction CNN architecture to recover and recognize the face images captured under poor lighting and illumination variation. The proposed architecture proved that it could efficiently remove unwanted lighting information. In the feature extraction step, the authors evaluated multiple methods in which the Relative Gradient Histogram Features (RGHF) achieved the best results compared to LBP and Patterns of Oriented Edge Magnitudes (POEM) [135]. The authors applied the Nearest Neighbor in the classification stage. In addition, the Generative Adversarial Network (GAN) was used to build a face recognition system by Yang et al. [1]. The authors proposed a GAN model to remove unwanted illumination from the image by building a new image based on the original face image captured under lighting variation. The GAN contains two networks built based on CNN, the generative network used to generate a new image and the discriminative network applied to evaluate the built image, in which the two network works in an adversarial manner. In the feature extraction step, the RGHF achieved good results compared to LBP and POEM. Also, in this work, the authors used the k-Nearest-Neighbor in the classification stage.

Zeng et al. [136] proposed a new scheme combining the handcrafted and deep CNN features. In addition, they presented an expanding sample method to augment the training set. The authors achieved remarkable accuracies compared with several face recognition systems, especially when they used only one sample in the training sets. Morever, Zeghina et al. [137] applied the Harris Detector to detect the important regions from the facial images. Then the authors applied a proposed CNN on these extracted parts to recognize faces. The system's peformance was evaluated on small datasets and achieved an acceptable result on the occluded face images. Almabdy et al. [138] used pre-trained deep CNN models to creat a new face recognition system. The system applied AlexNet and ResNet-50 models for the feature extraction task, then concatenated the two extracted vectors to generate a longer vector containing 6144-dimensional features. Finally, the system used SVM in the classification stage.

Table 3.1 compares the accuracy of some state-of-the-art face recognition techniques. The

results illustrated the superiority of the methods based on the Deep learning approach compared to the methods based on the handcraft approach.

Approach	Database	method	Accuracy
Handcraft	ORL	Ouyang et al. [114]	97.22%
	FERET	Qin et al. [117]	76.50%
	Extended Yale B	HFRIN-SFDWT FR [118]	95.56%
Deep	ORL	Min et al. [130]	97.77 %
Learning	ORL	Yallamandaiah and Purnachand [133]	98.48%
	FERET	Zeng el al. [136]	93.90%
	FERET	Min et al. [130]	93.04%
	Extended Yale B	Bendjillali et al. [2]	98%
	Extended Yale B	Yang et al. [1]	99.90%

Table 3.1: The performance of some state-of-the-art face recognition methods

3.3 Iris Recognition

Iris recognition is one of the most authentic and secure techniques used for personal identification. The iris texture has unique features that can be applied even to distinguish between identical twins.

In this sub-section, we presented the process that uses by most of the iris identification systems, and then we reviewed the state-of-the-art method that applied the iris modality to recognize persons.

3.3.1 Iris recognition process

Generally, the iris recognition systems apply four steps to identify the persons, as illustrated in Figure 3.11. At first, the system segmented the iris region. Then, it applies the normalization process to improve the iris region's quality. Finally, the system extracted features and applied the classification stage (matching and decision). This sub-section focuses on the iris segmentation and normalization steps. In addition, it quickly passes on the feature extraction and classification steps, because they are very similar to the steps used in face recognition that were explained in the previous subsections.

3.3.1.1 Iris Segmentation

Perfect iris region segmentation plays a vital role in the improvement of the reliability and accuracy rate of the iris recognition system. In this stage, the system aims to detect two iris boundaries:

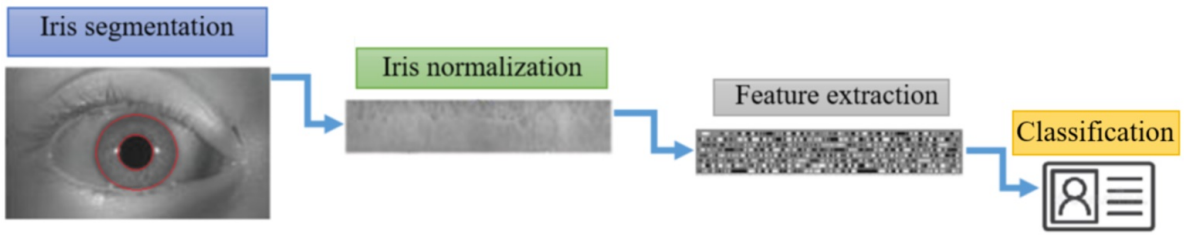


Figure 3.11: Iris recognition process

the iris-sclera boundary and the pupil-iris boundary. However, this sub-section presents two famous algorithms applied to segment the iris region: Daugman's Integro-Differential Operator [36] and the Circular Hough Transform [139].

I. Daugman's Integro-Differential Operator: Daugman proposed an iris segmentation algorithm based on a new integro-differential operator. The algorithm detects the iris and pupil region. In addition, it detects the upper and lower eyelids (see Figure 3.12). The proposed operator is defined as follows:

$$\max_{(r,x_0,y_0)} \left| G_\sigma(r) * \frac{\partial}{\partial r} \oint_{r,x_0,y_0} \frac{I(x,y)}{2\pi r} ds \right| \tag{3.7}$$

Where $I(x,y)$ is the intensity value of the input eye image in the pixel (x,y), r is the radius, $G(r)$ is the Gaussian smoothing filter controlled by σ, s is the contour of the circle that generated by (x,y) as coordinates centre and r as radius.

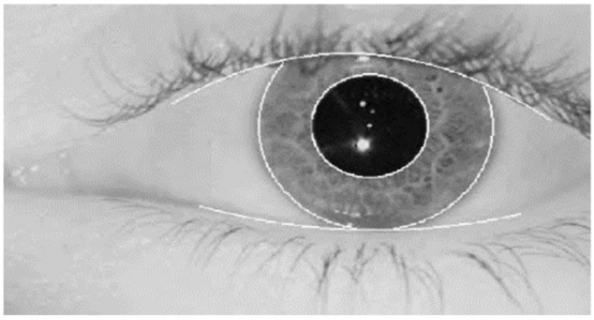
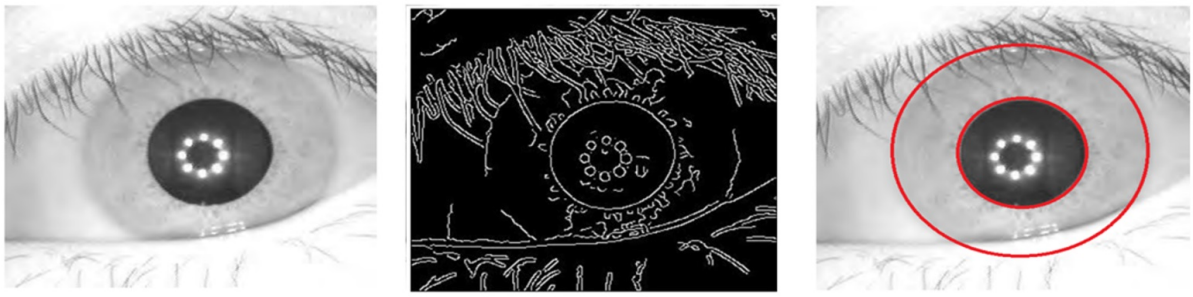


Figure 3.12: Iris segmentation computed by Daugman's Integro-Differential Operator, adapted from [140]

This algorithm works only on the local scale; for this reason, it mostly fails to gate good results due to the noise, such as reflections, and strong boundaries of lower and upper eyelids.

II. The Circular Hough Transform: is a computer vision algorithm applied to detect simple geometric elements such as circles and lines. The Hough transform technique is tolerant of gaps in the boundaries of objects and has an acceptable tolerance to noise.

The circular Hough transform specifically analysis the images to detect the aligned edges that create circles (See Figure 3.13). It is most commonly used to detect pupil and iris boundaries' radius and centre coordinates. First of all, the Circular Hough transform computes the first derivatives of the iris image intensity and then generates an edge map image based on a threshold value (See Figure 3.13 (b)). In addition, for each edge point, it draws a circle with a given radius r. Finally, the centre coordinates of the detected circle are the intersection point created by the largest number of circles.



(a) (b) (c)

Figure 3.13: Iris segmentation process using Circle Hough transform : (a) original image, (b) Edge map image, and (c) segmented iris image.

Generally, the algorithm is applied for a range of radii because we don't know the radius of the iris and the pupil. Then we select a radius r that can build an intersection point (x, y) based on the maximum number of circles. Finally, based on the selected radius r and the coordinates of the intersection point (x, y), we can define the detected circle by the following equation:

$$x^2 + y^2 = r^2 \tag{3.8}$$

Hough transform is based on the "brute-force" approach, which makes this algorithm very slow and computationally intensive.

3.3.1.2 Iris normalization

After detecting the iris region, the iris systems applied the normalization stage to build a fixed dimension feature rectangular that allows matching between two different iris images. Most iris recognition systems use a famous iris normalization algorithm called Daugman's rubber sheet model.

Daugman's rubber sheet model [140] is one of the most effective iris models used for iris normalization. The idea behind the model is to remap each cartesian coordinate (x, y) of the iris region into a polar coordinate (r, σ) system as shown in Figure 3.14.

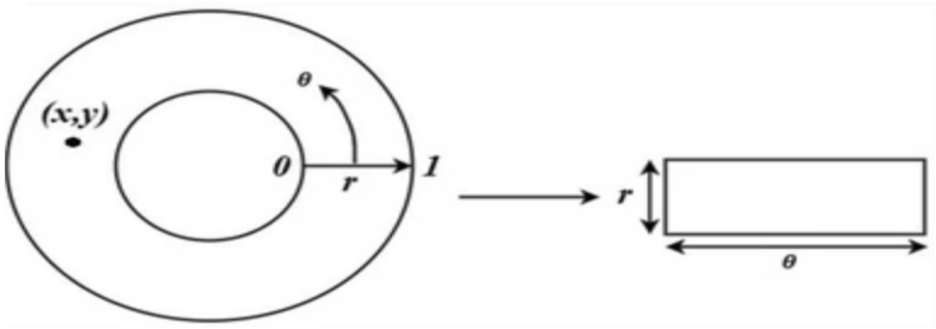


Figure 3.14: Daugman's rubber sheet model

The proposed model used the following equation to remap the points from the Cartesian to Polar coordinates:

$$I(x(r,\theta), y(r,\theta)) \rightarrow I(r,\theta) \tag{3.9}$$

With:

$$\begin{cases} x(r,\theta) = (1-r)x_p(\theta) + rx_l(\theta) \\ y(r,\theta) = (1-r)y_p(\theta) + ry_l(\theta) \end{cases} \tag{3.10}$$

Where $I(x, y)$ is the iris region, (x, y) is the original coordinates on the Cartesian space, and the (r, σ) is its correspondent coordinates on the polar space. In addition, (x_p, y_p) and (x_l, y_p) are the boundary coordinates along the direction of the pupil and iris.

3.3.1.3 Feature extraction

The features extraction module is a delicate step in the iris recognition process, where the extraction of discriminating information allows us to create a robust iris recognition system.

The researchers used many feature extraction methods that achieved great attention in the face recognition task to propose iris recognition systems, such as LBP [99], SIFT [18], and CNN [141]. However, some researchers encoded all the information included in the normalized iris region. The idea is to build a biometric template based on the feature encodings algorithms like Wavelet encoding, Gabor Filters, and Log-Gabor Filter.

3.3.1.4 Classification

The classification stage is applied to recognize the features extracted from a test iris image based on the features extracted from a training set. This task is used to match the features extracted from the testing image against the training features and decide the identity of the testing iris

image. In the literature, several classifiers have been used to build an iris recognition system, in which the most common classifiers used for this task are the classifiers explained in the face recognition process (SVM, k-NN, and NN).

3.3.2 Iris recognition state-of-the-art

Researchers have recently focused on the iris recognition field because it achieved extraordinary reliability compared with other biometric modalities.

Khorimah and Juniati [142] used the Hough transformation to detect the iris region and applied the rubber sheet model in the normalization task. The box-counting method is used by the authors to extract features. In addition, the k-Nearest-Neighbor (k-NN) is applied in the classification stage. Dua et al. [143] create rectangular iris images by applying the rubber sheet model after detecting the iris region by applying the circular Hough transform. The authors used the Log-Gabor filter to code the rectangular iris image and applied a neural network in the classification stage.

Abdo et al. [7] applied the Circular Hough Transform algorithm and Daugman's rubber sheet model to detect and normalize the iris region. In addition, the authors applied the Histogram Equalization method to build a well-distributed texture. Then they applied the Discrete Cosine Transform (DCT) to extract features from the texture obtained. Finally, the multi-class SVM is used in the classification task. Abdalla et al. [144] combined the DCT with the Discrete Wavelet Transform (DWT) and applied them to extract features from the normalized iris images. The proposed feature extraction technique obtained good results with the multi-class SVM classifier. Winston and Hemanth [145] introduced two modified Self Organizing Maps (SOM) used for iris recognition. The accuracy rate of the proposed maps reached 98.4% on the IITD database. Taha and Ahmed [146] applied the Canny edge detector, and the Circular Hough Transform to detect the iris region. Then, the authors used the Daugman rubber-sheet model in the normalization stage. Finally, they applied the Local Binary Pattern (LBP) to extract features and Euclidien Distance in the matching and recognition stage. The system achieved an iris recognition rate of 86% on the Ubiris-v1 database. Abdo et al. [147] introduced a new iris recognition approach based on the Fuzzy Local Binary Pattern (FLBP) [148]. Multiple classifiers evaluated the FLBP method on multiple iris image datasets, and it achieved good results with the SVM classifier.

Taha and Ahmed [149] used the Speeded Up Robust Features (SURF) Descriptor to extract features from the normalized iris region. Then the author applied Euclidien Distance in the classification task. Gad et al. [150] developed an iris recognition system that applied a new algorithm to localize the iris region based on the masking technique. In addition, the author applied two new techniques to extract features (Delta-Mean and Multi-Algorithm-Mean). The proposed system was evaluated on five iris datasets and achieved acceptable accuracies. Ak and Steluta [151] modified the method of pupillary border detection proposed by Tahir and Bindian [152]. The modified method achieved more accuracy on various databases. In addition,

the authors proposed a new algorithm for limbic border detection. After detecting and extracting the iris region, the authors detected the eyelids and masked them. Finally, they normalized the result of the previous stage by the Rubber-sheet method and applied the Gabor filter to extract features. Ak and Steluta [151] built a new iris recognition approach. The authors detected the iris region by applying a modified method based on the pupillary border detection technique [133] and a new proposed algorithm for limbic border detection. Then, they used the Rubber-sheet model to normalize the detected iris and the Gabor filter to extract features. In addition, the authors used the Hamming distance in the matching task.

Alaslani [153] introduced an iris recognition system relying on deep learning. The proposed system used the canny edge detector and the Circular Hough in the iris segmentation task. In addition, the system normalized the iris image based on Daugman's rubber sheet model. In the feature extraction task, the author applied the pre-trained Alex-Net model. Moreover, the multi-class SVM is used in the classification stage. The experiments with the pre-trained Alex-Net model and the SVM classifier proved that the segmented irises are more effective in iris recognition compared with normalized irises. Alaslni et al. [154] used Transfer learning to build a robust iris recognition system. The pre-trained VGG16 model has been fine-tuned on an iris dataset before being used in the feature extraction stage. The system achieved a high iris recognition rate on the small datasets, reaching 100% on the IITD database. Arora and Bhatia [155] proposed a new deep CNN model to extract features from the normalized iris. The system used the Circular Hough transform to detect the iris region. Then, it normalized the iris image before entering them in the proposed CNN to extract features. Finally, in the classification stage, the system used a Softmax classifier.

Chakraborty et al. [156] developed an iris recognition system relying on a new texture-aware lightweight Deep CNN framework. The proposed framework achieved encouraging results on the CASIA-Iris-thousand and CASIA-Iris-Distance databases. The new framework applied robust techniques to overcome the lack of labelled iris data. Yifeng Chen et al. [5] proposed a new loss function called Tight Center to train Deep CNN models. The authors evaluated the loss function using three deep CNN models (Tiny VGG, MobileNet, and ShuffleNet) on three iris image datasets. The experiments proved the robustness of the proposed function on large datasets. However, its accuracy rate degrades on the small iris datasets. The authors achieved an accuracy rate of 99.30% on the IITD database. Sujana and Reddy [6] proposed an iris recognition approach relying on deep learning. The authors introduced a new deep CNN model to extract features from the iris after normalizing it. The suggested model was trained from scratch on two iris image datasets and achieved remarkable results. Kranthi Kumar et al. [157] used a CNN model inspired by the mini VGG architecture to build a novel iris recognition system. The proposed system used a modified Hough Transform to detect the iris region and applied the rubber sheet model in the normalization stage. Then the authors train the CNN model before using it in the recognition. The new iris system trained on the CASIA-Iris-V1 dataset and achieved an excellent

result. Shanbagavalli et al. [158] developed a hybrid method for iris recognition. The authors applied data augmentation, cropping, rotation, flapping, and Color Space Transformation to augment the iris dataset. In the features extraction and classification tasks, the system applied a modified Gabor filter followed by a newly proposed method called EMiCoAReNet (Emerging Mixed Convolutional and Adaptive Residual Network). The proposed method achieved an accuracy rate of 95.2% on the CASIA-Iris-Interval database.

Jayanthi et al. [159] presented a robust framework for iris recognition based on Deep Learning features. In the pre-processing step, the framework applied Median filtering, Gamma correction, and Bottom Hat filtering to enhance the quality of iris images. Then, it used Hough Circle Transform to localize the iris region. In addition, the authors used a proposed Mask R-CNN with an inception-v2 model in the segmentation and recognition tasks. Shanto et al. [160] proposed a new iris recognition system based on CNN. The system used the Canny Edge Detection and the Circular Hough Transform (CHT) in the segmentation stage. Then, it applied a proposed CNN model to the segmented iris images. The proposed CNN used three convolutional layers to extract features, in which a max-polling followed each layer. In addition, it used dense, dropout, and softmax layers in the classification task.

Hassan et al. [161] used the Circular Hough Transform (CHT) and the Rubber-Sheet Model to segment and normalized the iris region. Then, they applied a new CNN model to extract deep features from the normalized irises. Finally, the system used SVM to classify the extracted features. The system achieved an iris recognition rate of 99.07% on the CASIA-Iris-V1 database. Zambrano et al. [162] used a pre-trained CNN model to build a novel iris recognition system. The authors applied many pre-processing stages to enhance the iris image, then used a pre-trained CNN model without fine-tuning it on the iris datasets. Jia et al. [163] proposed a new ConvNet deep neural network to recognise the iris. The authors used a multi-level interaction method to correlate the deep features extracted by several convolutional layers. Furthermore, they applied a mask network to eliminate the noisy factors during the matching and enhance the accuracy rate. The iris recognition system proposed by Patil et al. [164] detects the iris region based on the U-Net [165] deep learning technique. Then, it applied Gabor transform to extract features and the Hamming distance for the similarity check. The proposed system achieved an encouraged recognition rate on the Ubiris-v1 database, reaching 95.03%.

Table 3.2 compares the accuracy rate of some state-of-the-art iris recognition techniques. The results illustrated the superiority of the methods based on the Deep learning approach compared to the methods based on the handcraft approach.

Though most of the proposed iris recognition systems achieved an acceptable accuracy rate, several issues remain to be further addressed. Most existing systems applied many pre-processing steps to build normalized iris images. They used them in the feature extraction task, affecting these systems' computational efficiency. In addition, several proposed systems applied complex methods to extract features like the systems that used deep CNN models containing millions of

Approach	Database	method	Accuracy
Handcraft	IITD	Winston and Hemanth [145]	98.4%
	CASIA-Iris-V1	Dua et al. [143]	97%
	UBIRIS-V1 B	Taha and Ahmed [146]	86.00%
	UBIRIS-V1 B	Taha and Ahmed [149]	83.00%
Deep	IITD	Alaslni et al. [154]	100%
Learning	IITD	Yifeng Chen et al. [5]	99.30%
	CASIA-Iris-V1	Alaslni et al [154]	98.3%
	CASIA-Iris-V1	Hassan et al. [161]	99.07%
	UBIRIS-V1 B	Patil and Vasanth [164]	95.03%

Table 3.2: The performance of some state-of-the-art iris recognition methods

parameters. Furthermore, many proposed approaches should be evaluated on more benchmarks and compared their performances with more state-of-the-art methods.

3.4 Face-Iris multimodal biometric recognition

To overcome the limitation of the face and iris unimodal system, many multimodal biometric systems relying on the face and iris modalities have been published in the last few years. The combination of these traits achieved great success, especially since the face and the iris can be captured by using only one camera. In addition, the two modalities can be collected at a distance and without the cooperation of the users. Moreover, the face and the iris modalities complement each other, and each minimises the disadvantage of the other.

3.4.1 Face-Iris multimodal biometric recognition process

The face-iris multimodal biometric recognition process is based on four main stages. In the pre-processing stage, the system detects the face region from the captured face images; also, it segments and normalizes the iris region. In addition, in the features extraction stage, the system extracts the discriminate feature from the pre-processed images. Furthermore, the matching stage is applied to compare the extracted features with the training features. Moreover, the decision stage used the matching scores generated by the previous stage to identify the users.

Five fusion levels have been used in the literature to build face-iris multimodal systems (see Figure 3.15).

The face-iris multimodal systems based on the image-level fusion segmented and normalized the iris region, then combined the generated iris region with the detected face region. In addition,

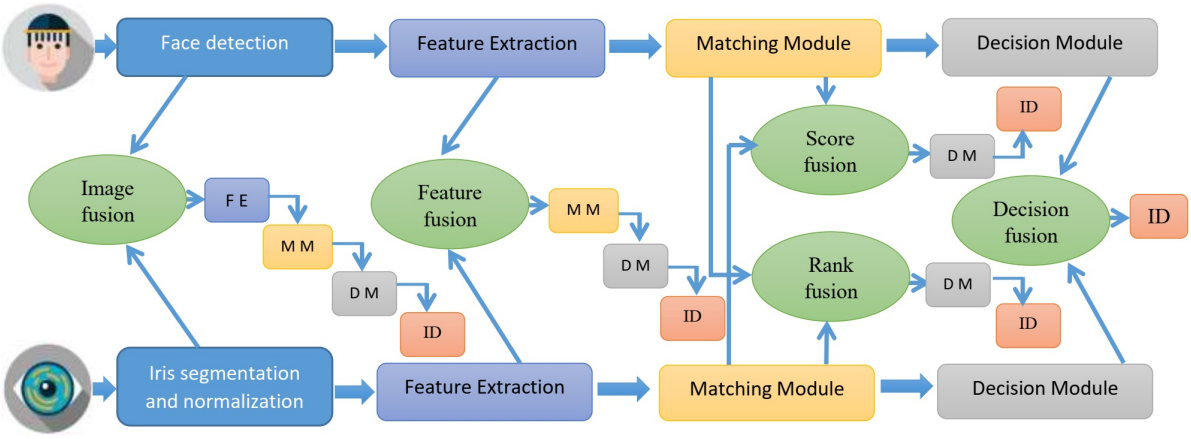


Figure 3.15: Face-Iris multimodal biometric recognition process

they extract the feature from the fused data and apply the classification process (matching and decision modules) to recognize the person.

With feature-level fusion, the systems extract the features from the detected face region and fuse the extracted features with the features extracted from the normalized iris image. Then, they applied the matching and the decision stages to identify the person.

Most iris-face multimodal recognition systems proposed in the literature are based on score-level fusion. This type of fusion measures the similarity between the features extracted from the face trait and the face descriptors stored in the training task. The same process is applied to the iris trait. Then, the system normalized and fused the obtained scores to use them in the decision module.

The rank-level fusion used the same process applied by the score-level fusion, but it used the matching module to calculate the ranked lists. Then, fused the two generated lists and applied the decision module to recognize the person.

The face-iris multimodal biometric systems based on the decision fusion applied the four modules to the iris image and the face images. Then, fuse the obtained decisions to obtain the final decision.

In addition, some researchers applied hybrid-level fusion to build a robust face-iris recognition system.

3.4.2 Face-Iris multimodal biometric recognition state-of-the-art

The multimodal biometric systems overcome several limitations of the unimodal systems, where they achieved high accuracy and reliability compared with the biometric systems that use only one modality in the recognition process. This sub-section will review recent papers combining face and iris images to recognize users. Amour et al. [12] proposed a face-iris multimodal biometric

system for person recognition. The system used the singular spectrum analysis and normal inverse Gaussian to extract features faces. In addition, the authors used a multiresolution 2D Log-Gabor to extract features from the normalized iris images. Finally, the system applied a hybrid level fusion (score fusion and decision fusion) to identify the user. Alsubari and Ramteke [166] proposed a novel multimodal biometric system based on face and iris feature fusion. The authors applied the Circle Hough Transform and the Daugman rubber sheet to normalize the iris region. Moreover, the system extracted the features from the normalized irises and the face images based on LBP and Gabor-Zernike Moments. Then, it fused the two feature vectors and applied different classifiers to evaluate the proposed system. Mansoura et al. [167] used the singular value decomposition (SVD) and applied Fast Fourier Transform (FFT) to extract features from the face and the iris modalities. Then, they calculate the Euclidean distance in the matching stage before fusing the scores.Rasool [13] applied four feature extraction approaches to extract features from the face and iris modalities. Then, the author used the fusion in two levels (features and score level). The best recognition rate has been achieved by applying the score-level fusion based on the texture descriptor and the statistical feature extraction methods.

A parallel architecture has been used by Al-Waisy et al. [11] to build a face-iris multimodal biometric recognition system. The authors used a Deep Belief Network (DBN) that contained two layers to extract the features and a softmax layer used as a discriminative model. In addition, the authors used an iris recognition network called IrisConvNet [168] to recognize the normalized iris images. The IrisConvNet contains five convolutional layers, each followed by a max pooling except the fifth. Moreover, the IrisConvNet applied two FC layers followed by a Softmax layer. The proposed multimodal fused the matching score obtained by the two networks (DBN and IrisConvNet) to achieve a high accuracy rate. Alay and Al-Baity [169] used the VGG16 to build a multimodal face recognition system based on face, iris, and finger veins modalities. The authors built two multimodal systems; in the first, they extracted the features from the three modalities, fused them, and applied the softmax classifier to recognize the persons. In the second, they used the VGG16 model to extract the features and the Softmax classifier to obtain the score of each trait. Then, the system applied a score-level fusion method based on the Sum or Product rule to identify the persons. In addition, Alay and Al-Baity [170] used the VGG16 to propose a new face-iris multimodal system. The authors combined the face and the iris trait based on two-level fusion (features and score). Moreover, Soleymani et al. [171] proposed a CNN model containing five convolution blocks, each followed by a max-polling layer. The authors used the proposed CNN to extract features from the face, iris, and fingerprint. Then, they Fused the learning features extracted by the different levels of the proposed CNN.

In another work [172], the authors proposed a CNN model containing four convolutional blocks, each followed by a max-polling layer, except the fifth block was followed by a FC layer. The proposed system applied the proposed CNN on the face, iris, and fingerprint images, and then it fused the outputs of the FC layers based on a Compact Bilinear Feature Fusion Algorithm. Xiao

et al. [173] introduced a new face-iris multimodal recognition architecture. The authors extracted learning features from the face images and applied the Chi-square distance to calculate the similarity. They used the Gabor filter on the normalized iris images to extract features and the hamming distance for the matching task. The authors introduced a new algorithm to select the best score-level fusion weights based on the quality of the face and the iris images. The algorithm applied various image quality metrics to choose the trait that has the best quality for increasing its weight. Harakannanavar et al. [174] proposed a multimodal system that used the Stationary Wavelet Transform and the LBP to extract features from the face and the iris modalities. The proposed system used PCA to reduce the dimensionality of the two feature vectors extracted from the face and iris images. In addition, it normalized and concatenated the reduced vectors before applying the Euclidian Distance classifier.

The proposed face-iris multimodal biometric systems achieved a high accuracy rate compared with the unimodal system, but they have some limitation that needs to be addressed. Most of the previous work used the pre-processing stage before extracting the features, which is a complex and delicate task. In addition, many face-iris multimodal systems have been evaluated on chimeric databases that combine two unimodal datasets. These systems do not use real databases to evaluate the proposed system performance [12, 167]. Furthermore, several proposed systems used handcraft approaches to extract features, which pose problems in unconstrained environmental conditions. Also, some methods applied Deep CNN models containing tens to hundreds of millions of parameters. Moreover, if one of the face or iris traits is missed or unavailable, the accuracy rate of the many proposed systems is dramatically degraded.

3.5 Conclusion

This Chapter introduces the face recognition system and discusses its steps. Also, it reviews the most recent works that proposed face recognition systems. In addition, the Chapter presents the iris recognition system and its process. Furthermore, it outlines the face-iris multimodal system and its different levels of fusion. Also, it reviews the relevant recent works that proposed multimodal biometric systems relying on the face and iris traits.

In this Chapter, we remark that the Deep Learning approaches applied by the unimodal and multimodal biometric systems achieved a high success compared with the handcraft approaches. For this reason, we focused on Deep Learning to propose robust biometric recognition systems. In the next Chapter, we outline the deep learning techniques used in the computer vision field, in which we introduce CNN and review some famous deep CNN models.

CHAPTER 4

DEEP LEARNING FOR COMPUTER VISION

"Deep Learning is a superpower. With it, you can make a computer see, synthesize novel art, translate languages, render a medical diagnosis, or build pieces of a car that can drive itself. If that isn't a superpower, I don't know what is."

Andrew Ng, AI Expert.

4.1 Introduction

Deep Learning (DL) is one of the most important fields of Machine Learning, which, in turn, is one of the most critical fields of Artificial intelligence (AI).

Deep Learning, also called Deep Neural Network (DNN), is a machine learning technique that mimics the human brain's function and structure in processing data, such as pattern detection and recognition. The Deep Learning technique is a way to automatically extract useful patterns from raw data, where it is an extension of the classical Artificial Neural Network (ANN), requiring huge datasets and High-Performance Computing (HPC) resources to learn its parameters.

Deep Convolutional Neural Network (DCNN) is a primary class of Deep Neural most commonly applied to identify and classify patterns in images and videos. Since 2012, and after the tremendous success achieved by DCNN in the ILSVRC (ImageNet Large Scale Visual Recognition Competition), which outperformed traditional image processing techniques, DCNN is still state-of-the-art in the computer vision field [175].

The CNN automatically extracts features from data without any preprocessing steps. Therefore, CNN is more independent of human experts than the traditional technique that uses handcraft methods in the feature extraction task.

In this Chapter, we briefly introduce Artificial Neural Networks (ANNs), and then we focus more specifically on Convolutional Neural Networks, which are the basis of this thesis. We respond to the question: what is CNN, and why is it used? Then we present the layers of CNN. In addition, we present some popular deep CNN architectures. We also discuss some techniques used to improve the performance of the Deep CNN model, and then we conclude the Chapter.

4.2 From human neural networks to artificial neural networks

No tool is more effective and powerful in pattern recognition and classification than the human brain. Therefore, encouraging the researchers to develop a machine-learning system that mimics the human brain's work. This section briefly presents human neural networks and introduces Artificial Neural Networks.

4.2.1 Human neural Networks

The human brain contains about 100 billion interconnected neurons that design sophisticated neural networks. The neuron receives inputs (electrical impulses) through its dendrites. The cell body processes the input signals and then transmits the output signal by the axon to other neurons through the synapse. Therefore, the synaptic allows making a connection between two neurons, connecting the axon of a neural to the dendrite of another neural [176]. Figure 4.1 shows the architecture of a human brain neuron.

Generally, the neurons in the human brain are organized into layers that create a network of neurons. The signal is sent chiefly between layers in one direction, where the neurons of a layer are connected with neurons in an adjacent layer [177]. After the above brief presentation of the human neural network, it is possible to design an Artificial Neural Network (ANN).

4.2.2 Artificial Neural Networks

An artificial Neural network (ANN) is a system inspired by the human brain neural network that contains a set of layers of neurons used for computational processing, which is specialized in pattern recognition [178]. In the case of a fully connected network, each neural is connected to all nodes of the next layer (See Figure 4.2).

The input of an ANN is a set of variables (features or attributes) for which we need to predict a discrete value in the classification or a continuous value in the regression tasks.

An Artificial neural network composes of a set of artificial neurons inspired by biological neurons. Each neuron in a layer is connected to the nodes of the next layer through weights, which refers to synaptic connections in the human brain neuron. Each neuron calculates the

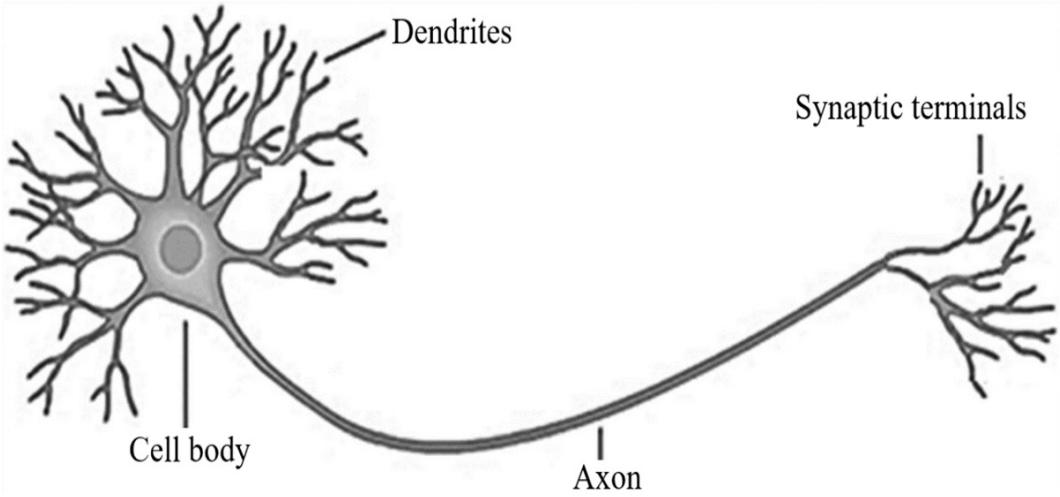


Figure 4.1: A neuron of the human brain, adapted from [177]

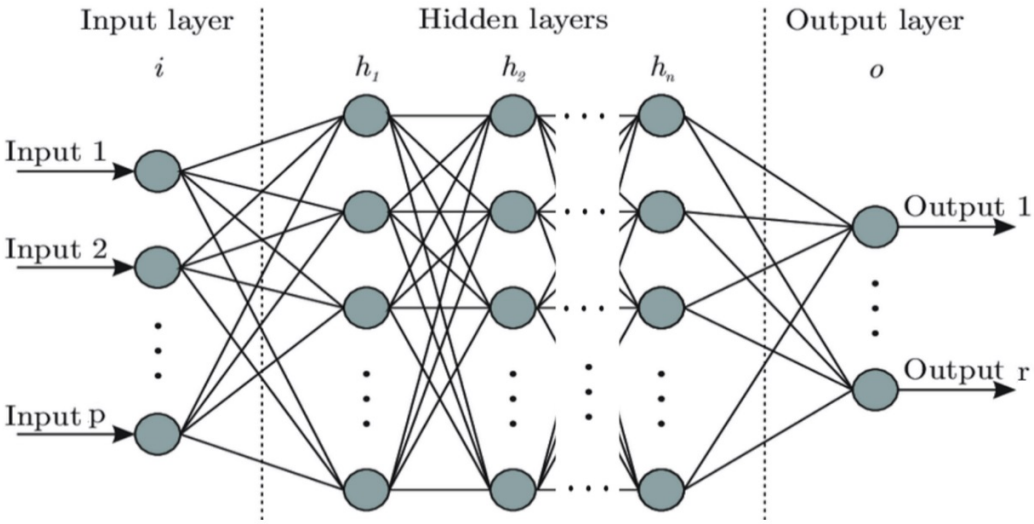


Figure 4.2: Artificial neural network contained n hidden layer, adapted from [178]

weights sum of its inputs plus the bias of the neuron (a corrective factor), and then this sum (equation (4.1)) is transformed by applying an activation function. The result of the activation function is the output of the neuron. Figure 4.3 shows the representation of an artificial neuron.

$$y = F\left(\sum_{i=1}^{d}(X_i w_i) + B\right) \tag{4.1}$$

An ANN propagates the computing through the network from the input to the output neuron(s), using weights as intermediate parameters. The output of this forward propagation is the

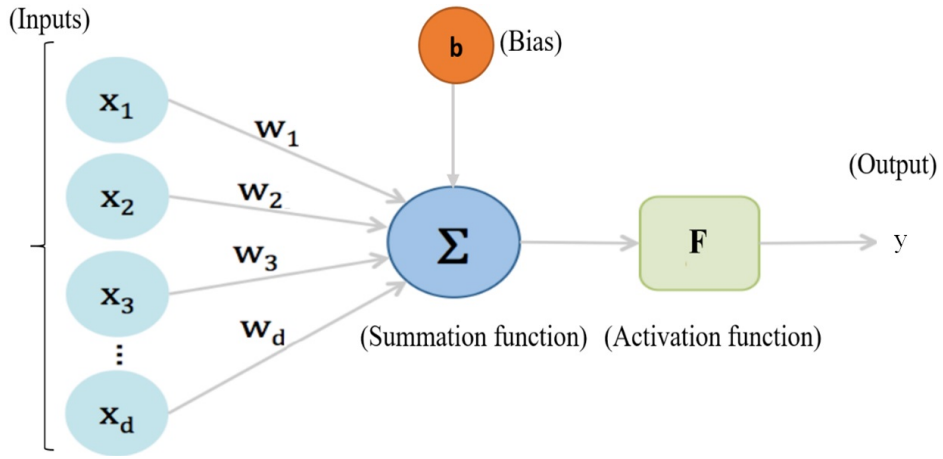


Figure 4.3: Representation of an artificial neuron

predicted value.

To make the ANN more accurate in future iterations and more correct in production, we fine-tune its trained parameters (weight and bias) using a backpropagation algorithm.

After each iteration, the backpropagation algorithm calculates an error metric by the loss function based on the predicted value and the desired output value. Then the ANN applies an optimization algorithm to minimize the loss function. Generally, the ANN uses an optimization algorithm called Gradient descent that fine-tunes the trainable parameters of the ANN until the convergence, where the value of the loss function does not minimize or a very little optimization [179].

4.3 Convolutional Neural Networks (CNN)

The success of CNN started in 2012 with the famous model AlexNet [107], which gained the first award in the ILSVRC [180] . The Alexnet model outperforms state-of-the-art handcraft methods in this competition. The CNN models have grown exponentially in the last decade, where their layers have progressed from 8 layers in AlexNet [107] to 152 layers in ResNet [109] just in three years (2012-2015). In this section, we answer why we need CNN and what CNN is. Then we briefly review the primary types of CNN layers.

4.3.1 Why do we need CNN, and what is CNN?

In image classification, the ANN takes the features extracted by handcraft methods as inputs and then fine-tunes its parameters to classify the input features.But why ANN uses auxiliary methods (handcraft methods) in the classification tasks? Why do we not receive the raw data as inputs?

The ANN can not take raw data as input because the image is usually built by tens of thousands to millions of pixels, which makes the network very complex, where the input layer requires thousands to millions of neurons to receive all the pixels. In addition, the ANN does not consider the local spatial information and the relationship between the neighbouring pixels, because it converts the image into a vector to input it into the network.

Since the development of the AlexNet CNN model in 2012, the Convolutional Neural Network has become state-of-the-art in the computer vision field [175]. CNN is an extension of the ANN that extracts features from the raw images directly, without using handcraft methods. Then it applies a classification process to classify the input images based on the features extracted in the previous step. The CNN is based on some layers used to extract features and reduce the dimensionality of data to save computational resources. Figure 4.4 illustrates the architecture of the basic CNN.

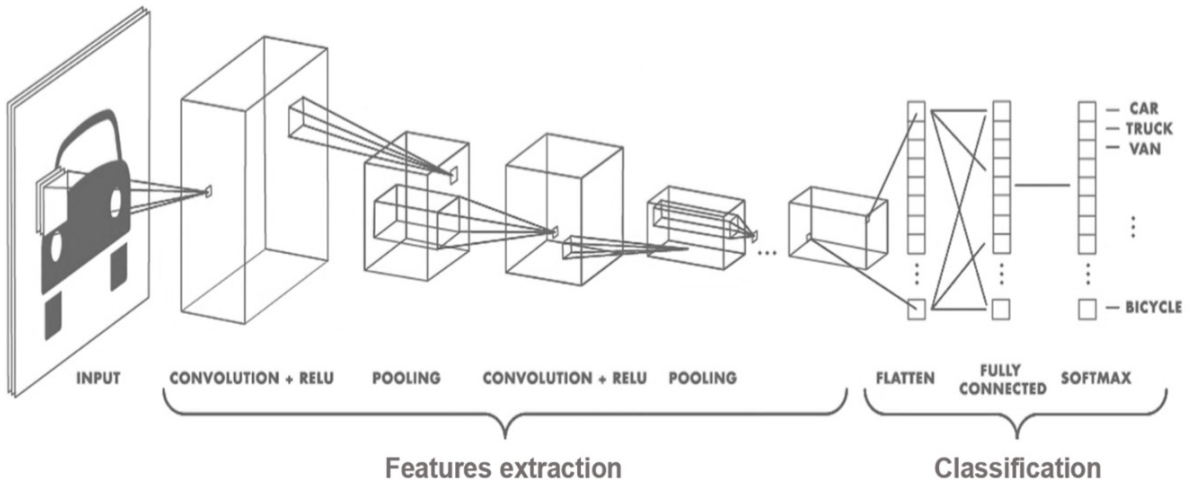


Figure 4.4: Architecture of Convolution Neural Network, adapted from [181]

4.3.2 The layers of CNN architecture

The CNN architecture contains different types of layers that are used to extract features and classify them. In this section, we review the primary layers that are used in most CNN models.

4.3.2.1 Convolution layer

The convolution layer is the most critical layer to building a CNN, where the first letter in the word "CNN" is an abbreviation that refers to this layer. This layer produces a feature map based on a mathematical operation called convolution. The following equation represents the convolution operation:

$$O_{x,y} = \sum_{j=0}^{k-1} \sum_{i=0}^{k-1} p_{x+i,y+i} w_{ij} \tag{4.2}$$

Where $O_{x,y}$ is the value of the output matrix at the point of (x, y), $p_{x+i,y+i}$ is the pixel's intensity value on the input matrix at the point of $(x+i, y+j)$, and k is the size of the filter. Figure 4.5 illustrates an example of the convolution operation.

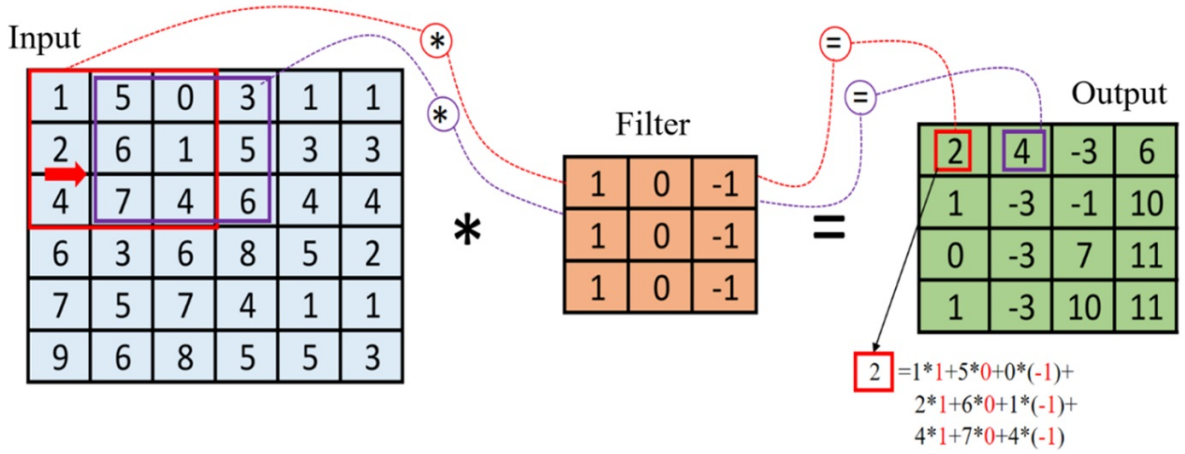


Figure 4.5: The convolution operation

The convolution layer uses a learnable filter, sliding on the input data (image/ tensor) to extract local features without human supervision. These filters, also called kernels, generally have small spatial dimensionality, but they have a depth equal to the depth of the input data. In practice, we applied numerous convolutions operations on the input, where each used a different kernel, in which each kernel produced a 2D feature map. Finally, the convolution layer fuses the 2D feature map to produce a 3D feature map. (See Figure 4.6)

After the convolutional operation, an activation function is applied to increase the feature map expressiveness. The activation functions can be classified into linear or nonlinear activation functions. Generally, we apply a nonlinear activation function because the convolutional operation output is normally linear. Figure 4.7 illustrates a famous nonlinear activation function called ReLU.

The activation function is not considered a layer from the CNN layers because it does not contain any parameters or trainable weights; for this reason, most CNN diagrams omitted it and considered it part of the convolutional layer.

The ReLU (Rectified Linear Units) has recently become the standard and the most common activation function on CNN design. However, the ReLU activation function is defined by The following formula:

$$\text{ReLU}(z) = \max(0, z) \tag{4.3}$$

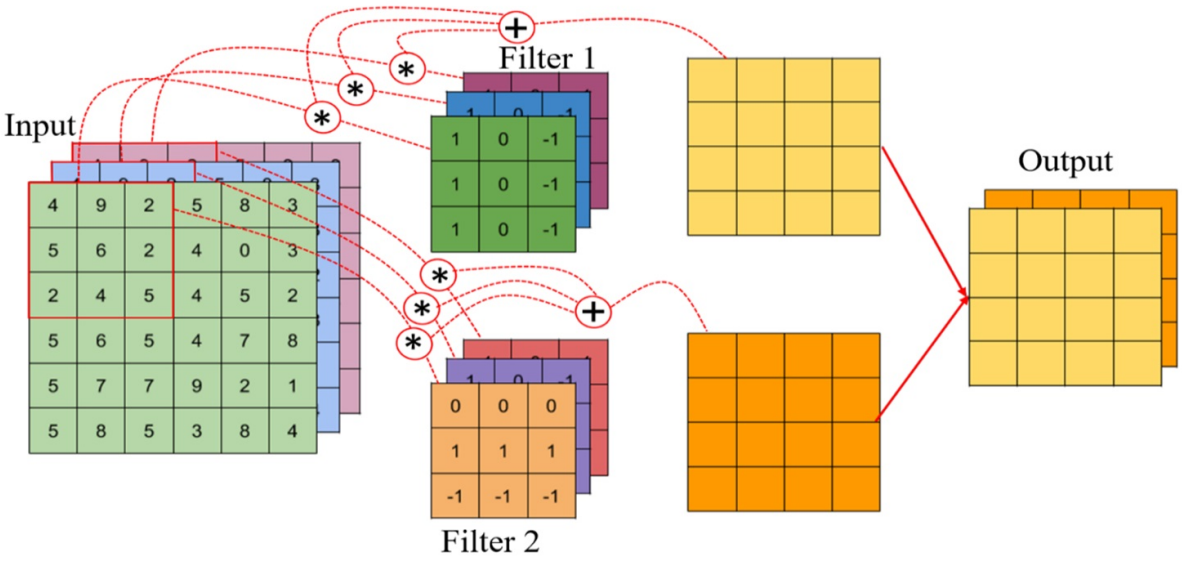


Figure 4.6: Convolution of RGB image using two different filters, adapted from [182]

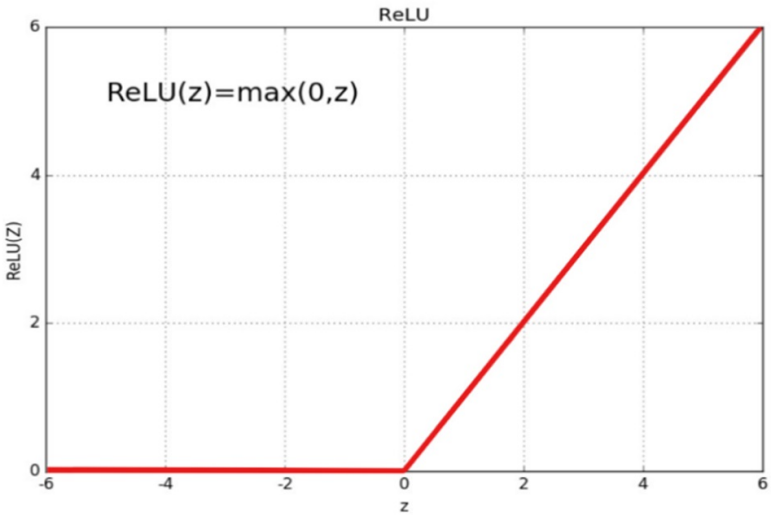


Figure 4.7: ReLU activation function.

Where Z is the output of the convolutional operation, the Relu function inverses the negative values to 0 and saves the positive values.

4.3.2.2 Pooling layer

The pooling operation is applied to reduce the feature maps' dimensionality, which reduces the number of parameters and the model's computational complexity. This operation reduces the spatial dimensionality height and width, and preserves the depth. The pooling layer applies a

kernel that divides the feature maps into non-overlapping rectangular regions. Then a specific function is used to reduce each region into only one value.

Max-pooling and Average-pooling are the most popular pooling function. The first one takes the max value from the region currently covered by the kernel (see Figure 4.8), and the second calculates the average of all values in the covered region (see Figure 4.9).

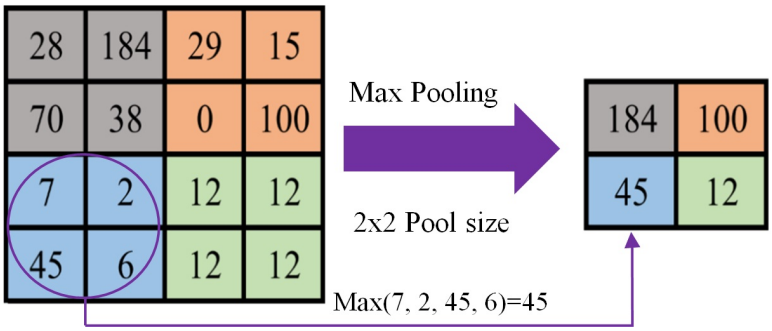


Figure 4.8: Max-pooling layer.

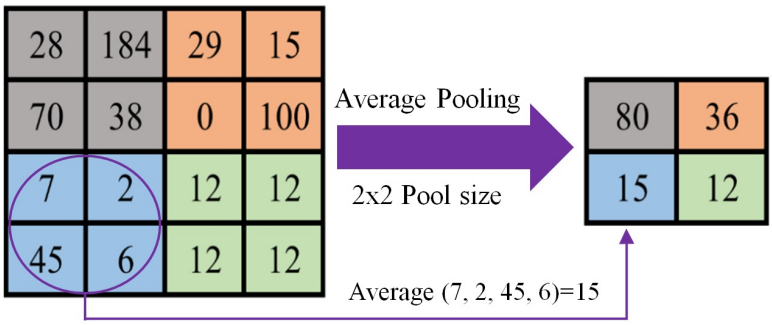


Figure 4.9: Average-pooling layer.

Generally, Max-pooling and average pooling are used in the middle of the CNN model to reduce the spatial dimensionality. In the last years, famous models (e.x., ResNet, Inception, Xception) applied a specific pooling function called Global Average-Pooling in the final layer of the network to minimize the size of the Fully connected layer. An example of a Global Average-pooling layer can be seen in Figure 4.10.

4.3.2.3 Fully connected layer

Fully Connected (FC) layers, also known as the Dense layers, are placed at the end of the CNN architecture to classify the original image. They are used after many feature extraction steps by alternating between convolution and pooling layers. The input of the first FC layer is the output of the last feature extraction layer. The FC layers entail a lot of parameters because each

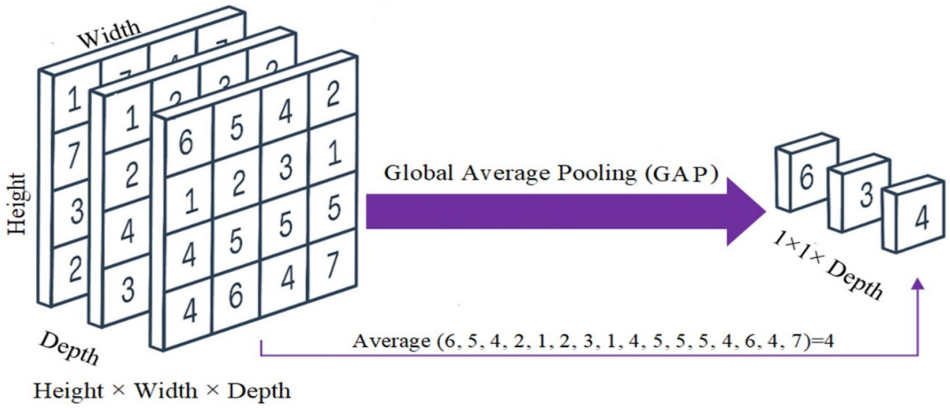


Figure 4.10: Global Average-pooling layer.

neuron of these layers is connected to all neurons of the previous layer, in which we need a weight for each connection. The last FC layer has N neurons, where N is the number of image classes. Generally, a softmax or a sigmoid activation function is applied to calculate the class probabilities of the input image.

4.4 Deep CNN architecture

Several Deep CNN architectures have been proposed in the last decade; in this section, we present some effective and robust models.

4.4.1 AlexNet architecture

In 2012, Krizhevsky et al. [107] proposed a deep CNN model Called AlexNet; in the same year, the Alexnet model gained first place in the ILSVRC (ImageNet Large Scale Visual Recognition Competition). The development of deep Learning has been accelerated dramatically after the superiority of AlexNet over the classical methods, where several effective models based on Deep CNN were proposed in the following years, and they dominated the ILSVRC competition.

The AlexNet architecture contains five convolutional layers, three max polling layers, and three Fully connected layers, where it has 62,378,344 parameters. Each of the first two convolutional layers is followed by a max pooling layer, and then the last three convolutional layers are placed before a final max polling layer. The default input for the AlexNet model is an RGB image with a size of 224×224. For the classification task, the model flattens the output of the final max-pooling layer and inputs them into the first Fully Connected layer, in which the AlexNet uses three subsequent fully connected layers. The first two layers contain 4096 neurons, and the third Fully Connected layer contains 1000 neurons, where 1000 is the number of classes in the ImageNet database. Figure 4.11 shows more details about the AlexNet architecture.

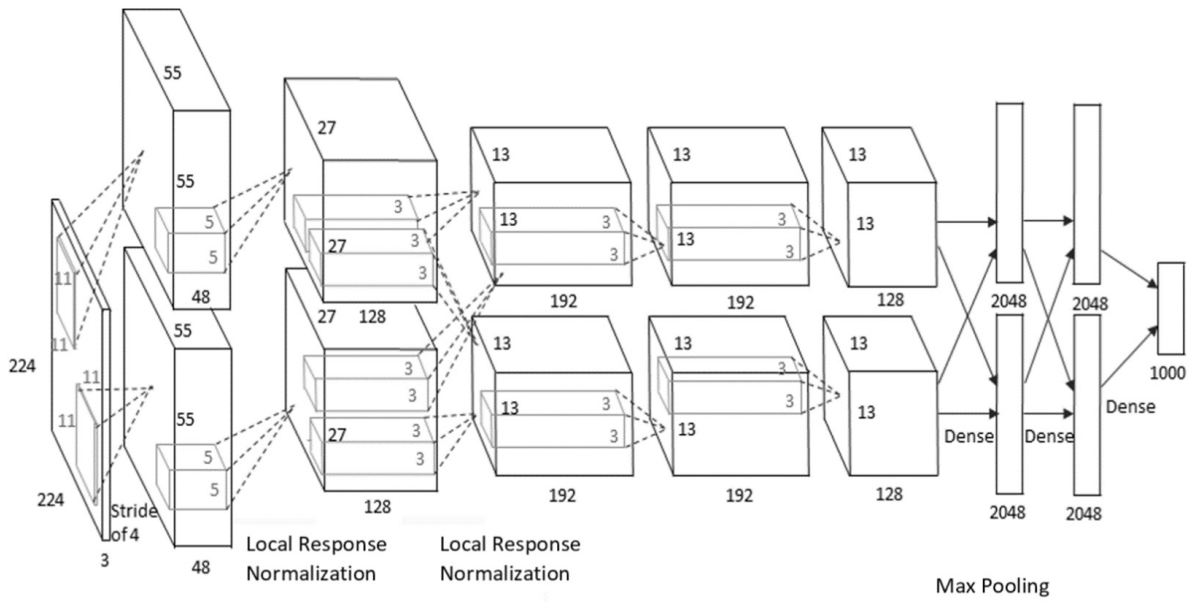


Figure 4.11: Two-branch AlexNet architecture [107]

The following python code is used to build an AlexNet model based on the Keras framework, where Figure 4.12 shows a summary of the model created.

```
import tensorflow as tf
from tensorflow.keras import layers
from tensorflow.keras.models import Sequential
alexnet =Sequential()
alexnet.add(layers.Conv2D(96, 11, strides=4, input_shape=(224, 224, 3),padding='same'))
alexnet.add(layers.Lambda(tf.nn.local_response_normalization))
alexnet.add(layers.Activation('relu'))
alexnet.add(layers.MaxPooling2D(3, strides=2))
alexnet.add(layers.Conv2D(256, 5, strides=1, padding='same'))
alexnet.add(layers.Lambda(tf.nn.local_response_normalization))
alexnet.add(layers.Activation('relu'))
alexnet.add(layers.MaxPooling2D(3, strides=2))
alexnet.add(layers.Conv2D(384, 3, strides=1, padding='same'))
alexnet.add(layers.Activation('relu'))
alexnet.add(layers.Conv2D(384, 3, strides=1, padding='same'))
alexnet.add(layers.Activation('relu'))
alexnet.add(layers.Conv2D(256, 3, strides=1, padding='same'))
alexnet.add(layers.Activation('relu'))
alexnet.add(layers.MaxPooling2D(3, strides=2))
alexnet.add(layers.Flatten())
alexnet.add(layers.Dense(4096, activation='relu'))
alexnet.add(layers.Dense(4096, activation='relu'))
alexnet.add(layers.Dense(1000, activation='softmax'))
```

```
_________________________________________________________________
 Layer (type)                 Output Shape              Param #
=================================================================
 conv2d (Conv2D)              (None, 56, 56, 96)        34944
 lambda (Lambda)              (None, 56, 56, 96)        0
 activation (Activation)      (None, 56, 56, 96)        0
 max_pooling2d (MaxPooling2D )(None, 27, 27, 96)        0
 conv2d_1 (Conv2D)            (None, 27, 27, 256)       614656
 lambda_1 (Lambda)            (None, 27, 27, 256)       0
 activation_1 (Activation)    (None, 27, 27, 256)       0
 max_pooling2d_1 (MaxPooling   (None, 13, 13, 256)      0
 2D)
 conv2d_2 (Conv2D)            (None, 13, 13, 384)       885120
 activation_2 (Activation)    (None, 13, 13, 384)       0
 conv2d_3 (Conv2D)            (None, 13, 13, 384)       1327488
 activation_3 (Activation)    (None, 13, 13, 384)       0
 conv2d_4 (Conv2D)            (None, 13, 13, 256)       884992
 activation_4 (Activation)    (None, 13, 13, 256)       0
 max_pooling2d_2 (MaxPooling   (None, 6, 6, 256)        0
 2D)
 flatten (Flatten)            (None, 9216)              0
 dense (Dense)                (None, 4096)              37752832
 dense_1 (Dense)              (None, 4096)              16781312
 dense_2 (Dense)              (None, 1000)              4097000
=================================================================
Total params: 62,378,344
Trainable params: 62,378,344
Non-trainable params: 0
_________________________________________________________________
```

Figure 4.12: Summary of the AlexNet model.

The AlexNet model contained more than 62 million trainable parameters. The CNN models learn in the convolutional layer and the fully connected layer; for this reason, it needs trainable parameters in these two layers. The following equation calculates the number of parameters used by a convolutional layer:

$$n = ((w \times h \times d) + 1) \times k \tag{4.4}$$

In which: w is the width of the filer used by the convolutional layer, h is the high of the filter, d is the number of filters in the previous layer, and k is the number of filters in the courant layer. We add 1 because each layer has a bias.

For example, the number of parameters in the second convolutional layer in the Alexnet model is ((5×5×96)+1)×256)= 614656. Because this layer used 256 filters with a size of (5×5), and 96 is the number of filters used by the previous convolutional layer.

In addition, the number of parameters used by a fully connected layer is calculated by the following equation:

$$n = c \times p + 1 \times c \tag{4.5}$$

In which: c is the number of neurons in the current layer, and p is the number of neurons in the

previous layer. We added $1 \times c$ because each layer uses a bias. For example, the number of neurons in the first and the second fully connected layer in the AlexNet model is 4096; therefore, the number of training parameters in the second fully connected layer is(4096×4096+1×4096)=16781312.

In 2014, Krizhevsky et al. [20] proposed a new version of the AlexNet model. The AlexNet-v2 achieved a good performance, although it removed the Local Response Normalization layers. Figure 4.13 shows the two versions of AlexNet.

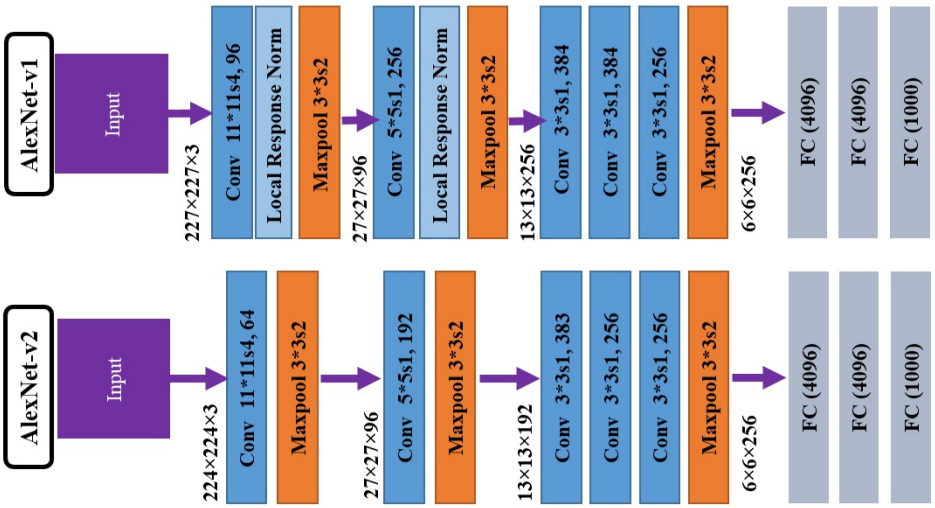


Figure 4.13: One-branch AlexNet-v1 and AlexNet-v2 architectures.

4.4.2 VGGnet architecture

In 2014, Two researchers named Simonyan and Zisserman from the Visual Geometry Group (VGG) of Oxford University proposed two CNN architectures called VGG16 [17] and VGG19 [17]. The VGG16 architecture was the runner-up of the ILSVRC 2014 classification, achieving a top-five accuracy of 91.90%.

The VGG16 model has five convolution blocks and three Fully Connected layers, where it has 138,355,752 parameters. Each block uses a max polling layer after some convolutional layer to remove the noise and reduce the output size. Each of the two first blocks contains two convolution layers. In addition, three convolution layers have been used in each of the three last blocks. Throughout the entire network, the VGG16 used a very small convolutional filter (3 x 3) with a stride and a padding of 1. As for the max pooling layer, the VGG16 used a 2×2 pixel window with a stride of 1 and no padding.

After the convolutional blocks, the VGG16 adds a flattened layer that converts the 3D feature map to a 1D vector to add the Fully Connected layers. The two first Fully Connected layers contain 4096 neurons, and the last contains only 1000 neurons. To ensure that output probability

summation is one, the VGGnet inserts a softmax activation function layer after the last Fully Connected layer. In addition, all the hidden layers applied the ReLU activation function to ensure non-linearity.

The VGG19 is a deep CNN network very similar to VGG16. The main difference is that the VGG19 uses three additional convolutional layers, VGG19 instead of 16 convolutional layers. Therefore, the number of parameters in the VGG19 increases to 143,667,240. The main drawback of the VGG16 architecture is the vast number of weights used in the classification task, in which the model uses 3 Fully Connected layers using 4069, 4096, and 100 neurons, respectively. Therefore, the number of trainable parameters in the classification layer reached 123,642,856. Figure 4.14 shows more details about the VGGnet architecture.

4.4.3 ResNet architecture

In 2015, the researchers of Microsoft -He et al. [109]- proposed a very deep CNN model that won ILSVRC 2015 with a top-five accuracy rate of 94.29%. The model is called ResNet, which stands for Residual Network. The new Residual Learning concept introduced in the ResNet model helps to go very deep for extracting complex high-level features. In just three years (2012-2015), and with this new concept, the CNN architectures progressed from 8 layers (AlexNet) to 152 layers (ResNet). Several version of the ResNet architecture has been proposed by He et al. [109], where the least deep version is the ResNet-18[109], which contain only 18 layers, and the deepest is the ResNet-152[109], which include 152 layers. In addition, other ResNet versions containing 34, 50, and 101 layers have been proposed in [109]. The number of parameters in the ResNet model varied between 11.4 and 58.5 million according to the depth of the version. The residual Learning concept introduced in this model overcomes the problem of vanishing the weights learned from the previous layers during the backpropagation on the very Deep CNN [183].

All the versions of the ResNet started with a convolutional layer that has a filter of (7x7), then they applied a batch normalization before the activation function and then used a Max Polling layer (3x3) to reduce the feature maps' dimensionality. Generally, batch normalization is a very effective quite used to improve and accelerate the training of Deep CNN models [184].

After the starting block, the ResNet-50 used 16 residual blocks, including convolutional and identity blocks. Figure 4.15 shows the structure of the two different blocks, in which each block applies three convolutional layers and uses a shortcut connection. The ResNet50 uses four identity blocks, where two convolutional blocks follow the first and the fourth identity blocks, and three convolutional blocks follow the second and the third.

The difference between the different ResNet versions is the number of residual blocks used. For instance, the ResNet-18 used only eight residual blocks, and the ResNet-152 used 50 residual blocks. The paper [109] that introduced the ResNet model illustrated the detailed architectures of the different ResNet versions.

In the end, the ResNet model used a Glolab Average Polling layer to reduce the dimension-

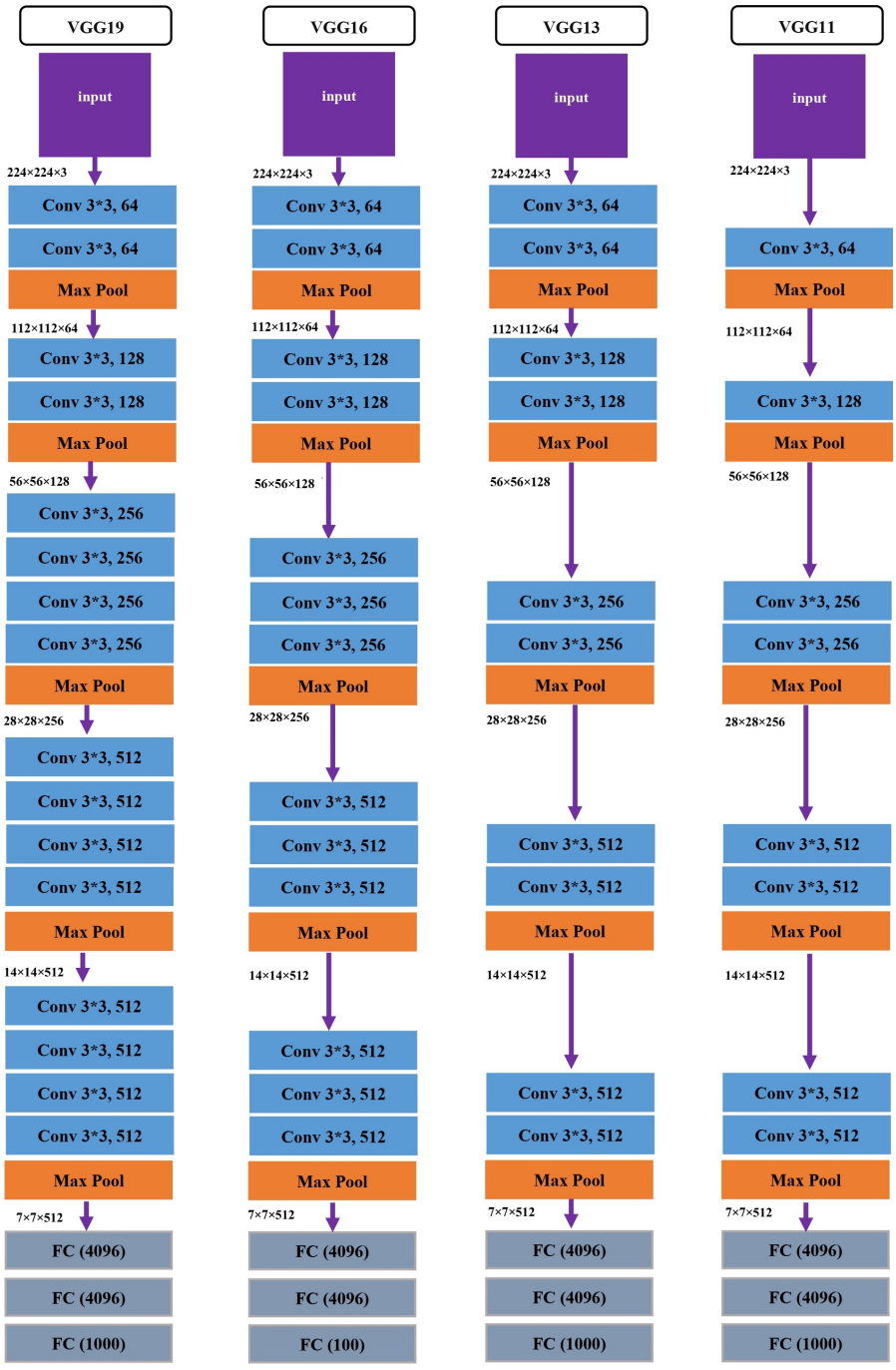


Figure 4.14: VGGnet architecture.

ality and used a Fully Connected layer of 1000 neurons in the classification task. The global architecture of the ResNet-50 model has been illustrated in Figure 4.15. The main drawback is that the ResNet model is very hungry for data, in which it needs very large databases to train its

huge number of parameters properly [109].

4.4.3.1 The residual block

The residual block is an idea introduced to enhance feature mapping by building higher-level features. It used a skip connection between the input of the residual block and its output to elevate the vanishing gradient phenomenon, accelerating the convergence process [109]. Since some information will be lost after applying a set of convolution layers, using the skip connection is a good idea to exploit all the information in the features building [185]. The ResNet50 presented in Figure 4.15 used two residual blocks:

- The identity block: is a standard residual block that contains three convolutional layers, each followed by batch normalization. It uses a skip connection connecting the input to the output layer see Figure 4.15 (b).
- The conv block: is very similar to the identity block but adds a 1x1 convolution to the skip connection to reduce the output depth. Figure 4.15 (c) shows a convolutional block.

4.4.4 GoogLeNet architecture

In 2014, Google researchers proposed a deep CNN model called GoogLeNet [108] that won ILSVRC 2014 with a top-five accuracy rate of 92.2%. After the great success achieved by the GoogLeNet model, which is also known as Inception-V1, the Google researchers proposed other robust versions such as Inception-V2 [186] and Inception-V3 [187].

The inception models applied a special convolution kernel (1×1) that was not used to extract local spatial features. However, it is used to extract cross-channel features. In addition, it reduces the tensor depth, reducing the number of weights and computational times.

GoogLeNet introduced a new concept called the inception module; the inception module contains many small convolution kernels (i.e., 1x1, 3x3, 5x5) distributed in parallel branches to extract the maximum of multi-scale features. The new concept optimized the number of parameters, increasing the width of the network and contributing to going deeper into the CNN architecture, in which the GoogLeNet contains 22 convolutional layers but with only 7 million parameters. Despite optimizing parameter numbers, GoogleNet outperformed AlexNet and VGGnet models that have more than 62 and 138 million parameters, respectively.

In 2015, the Inception-V3 model gained 1st runner-up in the ILSVRC 2015 competition, achieving a top-five accuracy of 94.4%. The inception-V3 factorized the filter 5×5 into two small filters (3×3) to reduce the number of parameters without decreasing the performance accuracy of the model. Also, to improve the computational speed, a convolution operation with a filter of 5×5 is 2.78 times more computationally expensive than a convolution operation with a filter of size (3×3) [187]. In addition, Inception-V3 factorized the convolution operation of filter size n×n

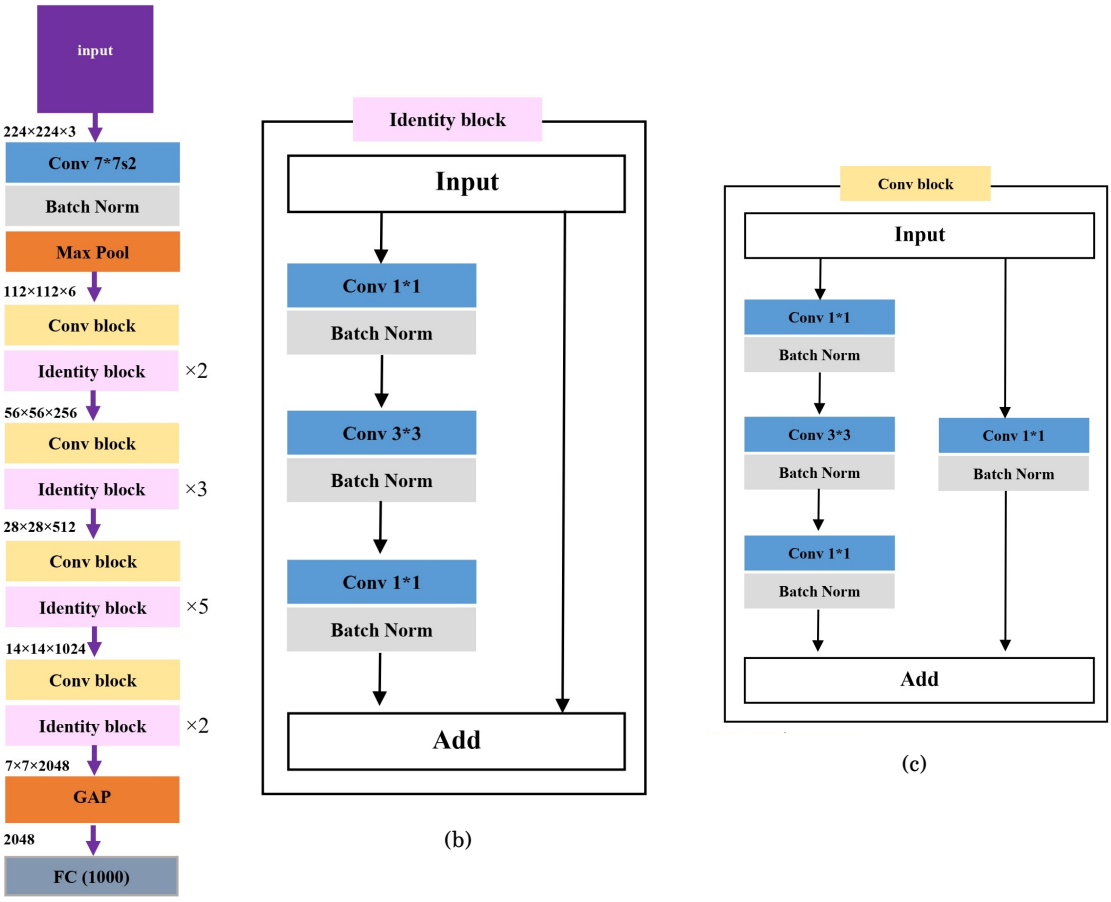


Figure 4.15: ResNet50 architecture: (a) The global model, (b) Identity block, (c) Conv block.

into two convolution operations with filter sizes 1×n and n×1. Also, this factorization reduces the number of parameters and saves computational resources. The Inception-V3 applied an n×1 filter followed by a 1×n filter (serial mode); also, it applied the two previous convolution operations in a parallel mode to make the model wider instead of deeper. Figure 4.16 shows the Incaption-V3 architecture.

4.4.5 Xception architecture

In 2017, the researcher of google François Chollet proposed an effective deep learning model called the Xception model [110]. The Xception deep learning model was inspired by the Google model 'Inception-v3', in which Xcepiton stands for eXtreme inception. The Xception model contains 14 modules built using 36 convolutional layers and has less than 23 million parameters . However, the model achieved significant success on the ILSVCR challenge with top-5 accuracy up to 94.50%; this result outperformed the results achieved by the Inception-v3 model, although it used fewer parameter numbers.

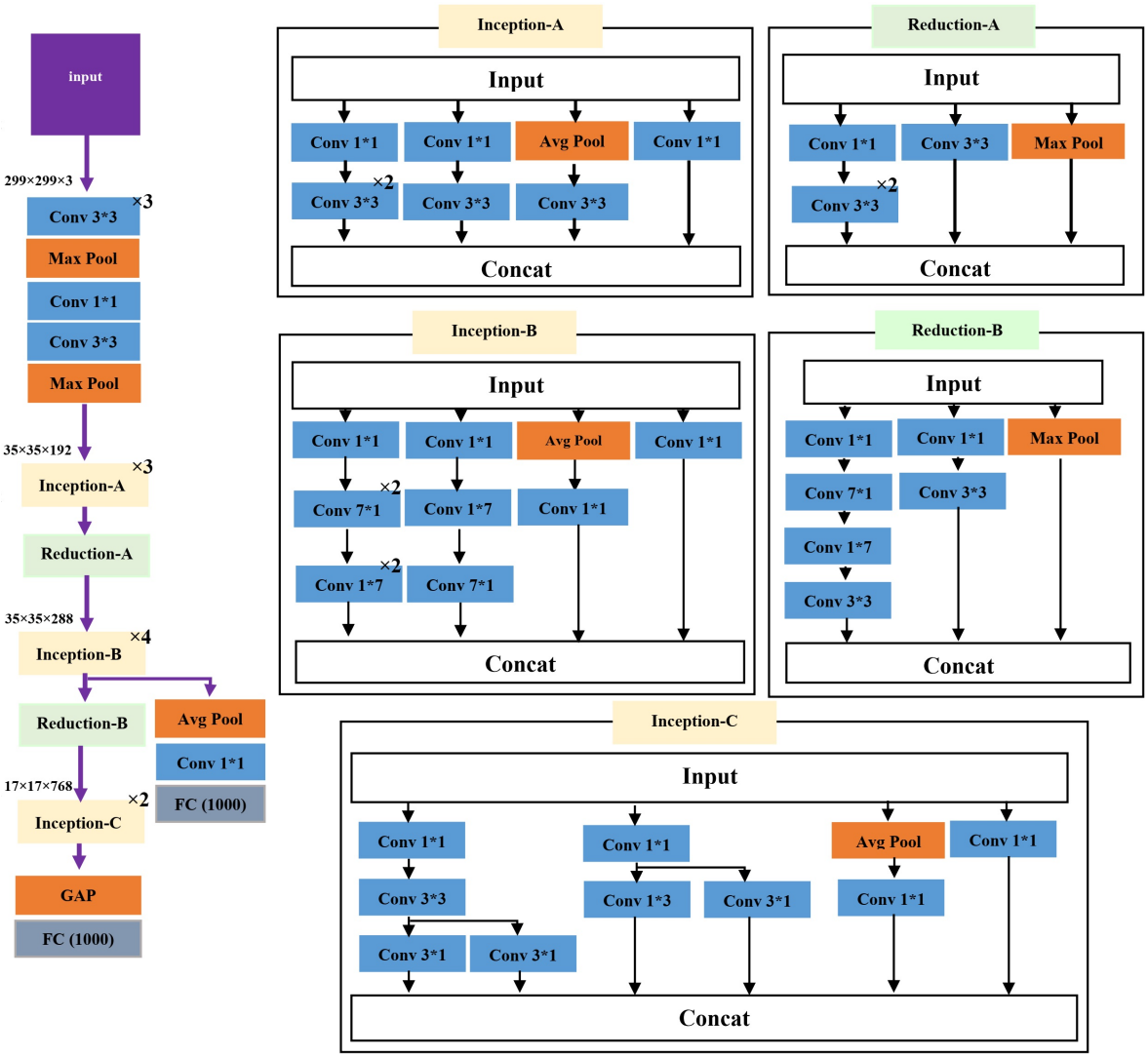


Figure 4.16: Inception-v3 architecture.

The main idea of the Xcepiton module is to apply a new depthwise separable convolution operation instead of the inception module, in which this operation is very identical to the traditional depthwise separable convolution that was presented in 2014 [188]. This type of operation used two primary convolutions: the depthwise convolution and the pointwise convolution. In the first convolution operation, each kernel independently convolutes in only one channel, and in the second operation, a 1x1 dimensional kernel is applied.

Two main differences between the two versions of the depthwise separable convolution [110]:

- The order of operation: the old version used a depthwise convolution first and then applied pointwise convolution, whereas, as shown in Figure 4.17, the module in the Xception architecture used the pointwise convolution.

- The presence of the non-linearity: each module of the Xception architecture applied a ReLu non-linearity function after both operations, whereas the original version used the non-linearity function only after the second operation (no intermediate non-linearity function).

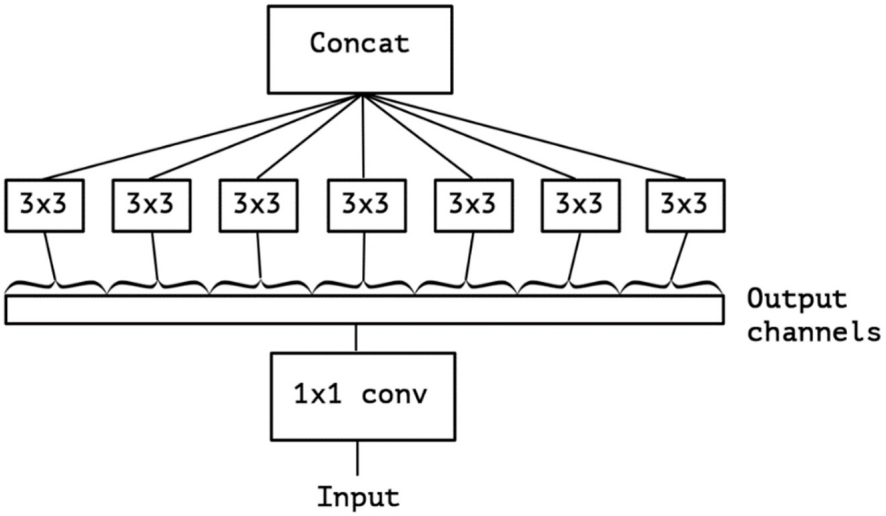


Figure 4.17: An "extreme" version of an Inception module .

Figure 4.18 shows the Xception architecture.
A brief comparison between the presented models has been shown in Table 4.1.

Year	Architecture	Top-1 accuracy	Top-5 accuracy	parameters
2012	AlexNet	84.60%	63.30%	62,378,344
2014	VGG16	91.90%	74.40%	138,357,544
2014	GoogleNet	92.2%	74.80%	23,000,000
2015	ResNet-152	94.29%	78.57%	25,000,000
2016	Inception-v3	94.4%	78.8%	23,851,784
2017	Xception	94.50%	79.00%	22,910,480

Table 4.1: Comparison between popular Deep CNN architectures, where top-1 and top-5 accuracy were achieved in the ILSVRC challenge.

4.4.6 YOLO architecture

The deep CNN was also applied in pattern detection, in which the YOLO model achieved high success in this task.

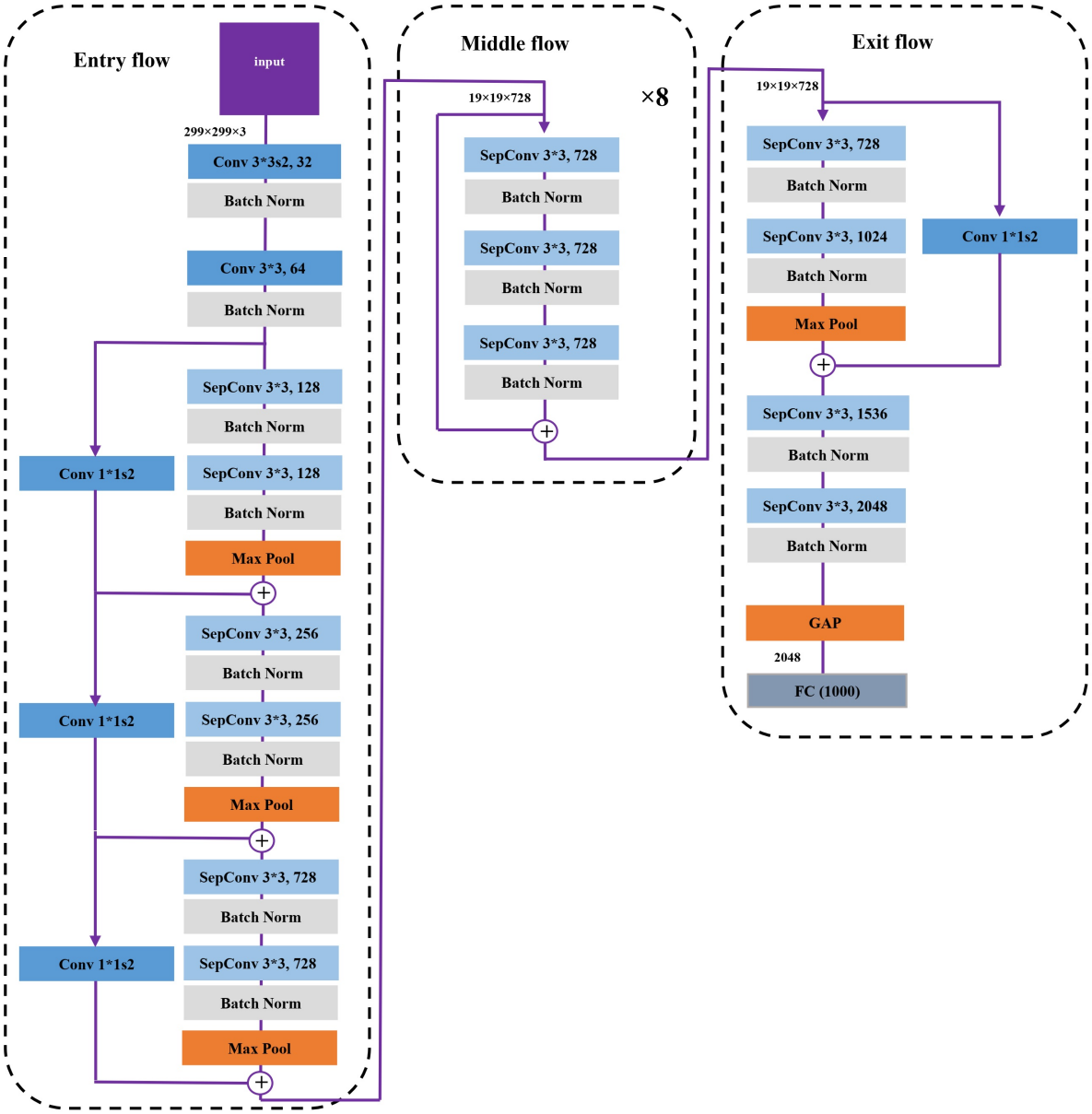


Figure 4.18: Xception architecture.

In 2016, Joseph Redmon et al. [16] proposed a CNN architecture called You Only Look Once (YOLO) for real-time object detection. The YOLO splits the images into small regions where each region is included in a bounding box and predicts the probability of classes of each bounding box [189].

Several versions of YOLO have been proposed after building the first version (YOLOv1 [16]), in which YOLOv2 [190] and YOLOv3 [191] were created in 2017 and 2018, respectively. In April 2020, the fifth version of YOLO [192] was published; also, there exists a fifth version of YOLO, but

it is not supported by any scientific paper [193, 194]. The latest scientifically supported version (YOLOv4 [192]) is the most accurate and fastest version in terms of both speed and accuracy [192]

To make the YOLOv4 more efficient, the authors add Cross Stage Partial Network (CSPNet) to the Darknet53 structure. The CSPNet is applied to improve the CNN learning ability, enhance object detection accuracy, lower memory cost, and save the computational bottleneck [195]. The YOLOv4 used the ImageNet dataset to train its feature extraction model separately. In addition, the Microsoft COCO dataset is used to evaluate the object detection performance [193].

The YOLOv4-tiny [21] is a faster and simple version of the YOLOv4 model. The YOLOv4-tiny network removes some convolutional layers of the original YOLOv4 to save memory and computing resources [195]. Therefore, the training of this version does not require big datasets compared with the original version because this version has fewer parameters. However, the tiny version achieved satisfactory performance in object detection. Moreover, it achieved a high speed of object detection, reaching 371 FPS (Frames per second) using a 1080Ti GPU [196]. Figure 4.19 illustrates the YOLOv4-tiny architecture.

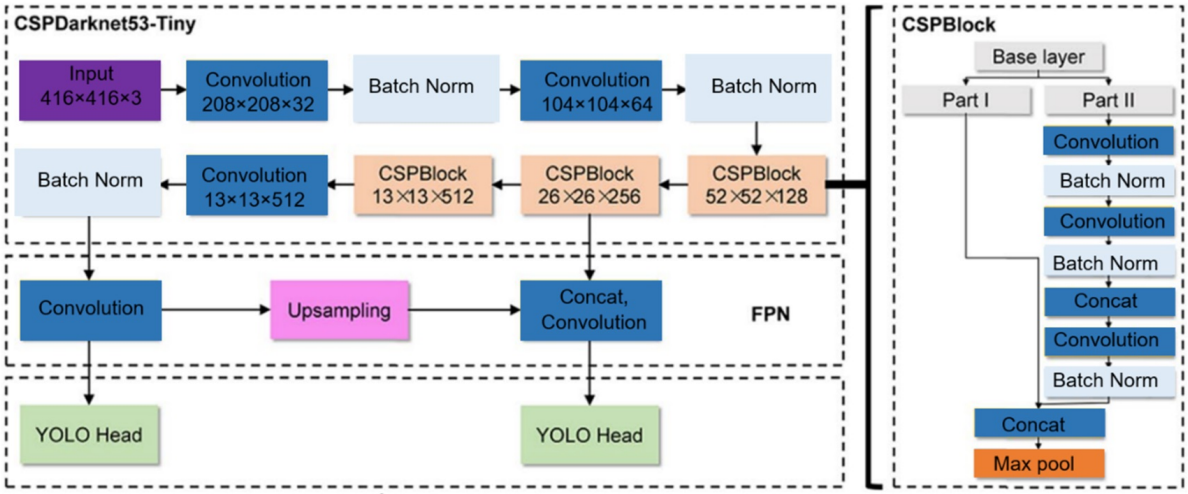


Figure 4.19: Yolov4-tiny architecture, adapted from [189]

4.5 Performance-improving techniques for deep CNN models

The most important challenge that has limited the success of the Deep CNN models in some tasks is the lack of large datasets. Due to the overfitting curse, it is impossible to train a Deep CNN from scratch to solve a complex problem based on a database containing a few samples. The Deep CNN models have tens of thousands to tens of millions of parameters that need huge amounts of data for a successful training process [175].

This section presents two techniques successfully used in the literature to overcome the challenge of lacking samples for achieving high performance with the Deep CNN.

4.5.1 Data Augmentation

The training of a deep CNN model that contains tens of hundreds to millions of trainable parameters is prone to the curse of overfitting [175].

Data Augmentation (DA) is one of the most practical techniques applied in the literature to deal with the curse of overfitting. It is a technique that applies different transformations to genuine training images to increase the amount of the training set. These transformations include not limited rotating, cropping, flipping, zooming, vertical or horizontal mirroring, and colour space transformation (see Figure 4.20).

Figure 4.20: Data Augmentation for face recognition by using some geometric transformations [197]

Generally, Data augmentation combines these transformations randomly to generate different artificial images. Practically, the data augmentation can decrease the top-1 error in the CNN by more than 1% [161].

4.5.2 Transfer Learning

The deep CNN models have achieved high success in the past decade due to the availability of huge datasets used to train this type of network hungry for data. The ImageNet and the COCO are the most famous datasets in Deep Learning.

The ImageNet[198] is a large benchmark image dataset used to train and evaluate new object recognition models, containing over 1.2 million images categorized into 1000 classes. The COCO [199] is a large image dataset developed by Microsoft, used to train and test the different

models proposed for object detection and segmentation. It contains more than 328K images and includes more than 2.5 million object instances, including 91 object classes, in which each object is localized and labelled precisely by humans.

Most of the state-of-the-art CNNs architectures are explicitly built for object detection and recognition on these two datasets. Reusing the state-of-the-art deep CNN architectures for different domains mostly fails due to insufficient annotated data [197]. The researchers introduce a new term called Transfer Learning, inspired by humans that can quickly exploit knowledge learned from a previous solving problem to solve a new problem. The Transfer Learning term in the machine learning field means using the knowledge picked up from one or more solving tasks in an alternate but related task [200]. Yosinski et al.[201] conducted many experiments demonstrating that we can use the pre-trained models on a new similar target dataset and achieve a high accuracy rate. Several pre-trained models trained on the ImageNet dataset, such as VGG, ResNet, Inception, and Xception, are freely available to solve different classification problems. Two approaches were used to apply transfer learning with a pre-trained network:

- The weights fine-tuning: we use the pre-trained model to initialize the weights, then train the model, in which we train all the network's layers, or just some layers, on the new task to fine-tune the weights (see Figure 4.21).
- The off-the-shelf CNN features: we use the pre-trained model trained on a specific task to extract features from another task without retraining the network on the new task (see Figure 4.21).

Generally, in the literature, the researchers used the off-the-shelf CNN features and trained only the classifier; if the new small dataset was similar to the dataset used in the pre-trained stage. In the case of a new small dataset not similar to the dataset used to pre-train the model, the researchers freeze some pre-trained feature extraction layers and train the rest layers with the classifier. In the case of a new huge dataset similar to the dataset used in the pre-trained stage, the researcher used the pre-trained weights to initialize the new weights and then fine-tuned all the weights on the new dataset. After several experiments, Razavian et al. [200] find that the off-the-shelf features are very interesting for many computer vision tasks, such as image classification, object detection, and visual instance retrieval. Moreover, the authors declared: "The results strongly suggest that features obtained from deep learning with convolutional nets should be the primary candidate in most visual recognition tasks."

4.6 Conclusion

The deep CNN models achieved high success in several computer vision tasks in the last decade. These models are fascinating compared to the handcraft methods; they fuse two main steps used in the classical techniques (feature extraction and classification) in only one step.

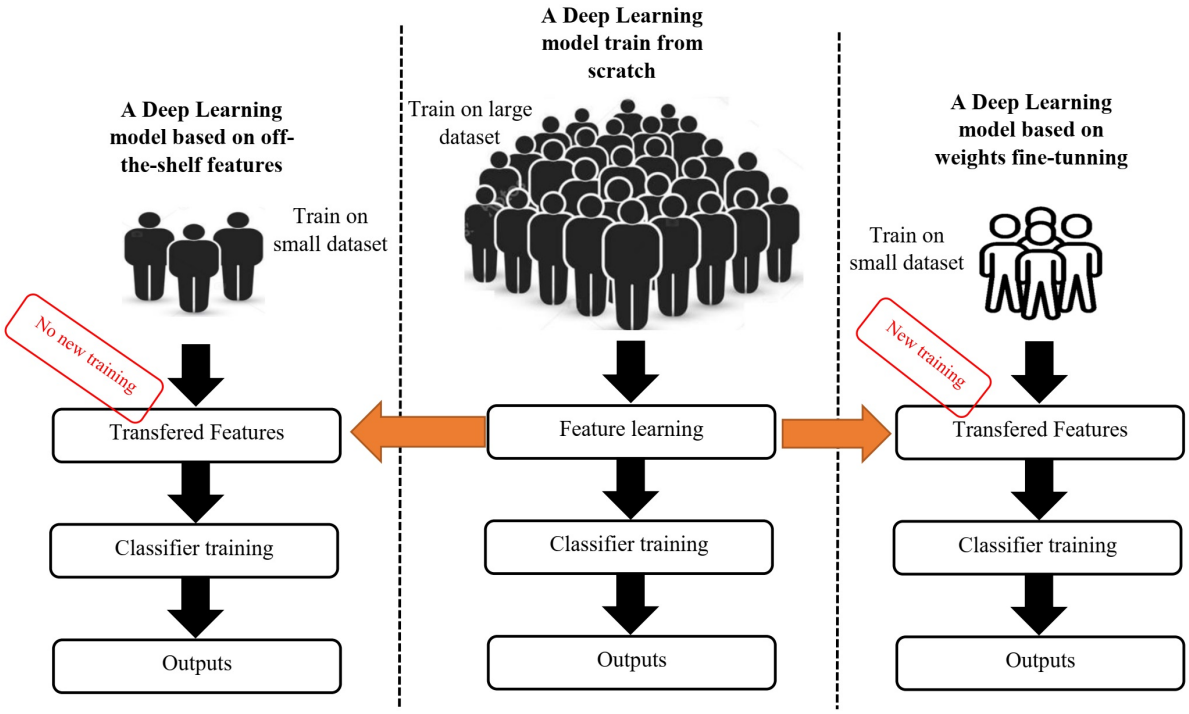


Figure 4.21: Transfer Learning approaches.

This Chapter presented the basics of CNN architecture after introducing the idea of ANN architecture inspired by the human brain neurons. In addition, it reviewed four deep CNN architectures famous in the object classifications task (AlexNet, VGGnet, ResNet, Inception, and Xception). It also presented a deep CNN model called YOLO applied for object detection. Furthermore, it reviewed Data Augmentation and Transfer Learning techniques applied in the literature to improve the effectiveness of the Deep CNN models. This Chapter is introduced to serve the following contribution Chapters by preparing their deep learning material.

In the next Chapter, we will see the first contribution Chapter, which proposes a face recognition approach based on Deep CNN architecture.

Part II

Contributions

CHAPTER 5

FACE RECOGNITION SYSTEM

"Facial recognition software is already quite accurate in measuring unchanging and unique ratios between facial features that identify you as you. It's like a fingerprint."

Jan Chipchase, Computer Scientist.

5.1 Introduction

Face recognition is a famous biometric modality that has received extraordinary attention in the research community for the last decades [44]. It offers several characteristics, for example, but not limited: it can be applied without the user's knowledge and cooperation. Furthermore, the facial trait captured from a distance without touching the sensors, which can limit the potential spread of the COVID-19 pandemic and some other diseases by touch. In addition, face recognition systems are characterized by low cost and high accuracy.

The acceptance of users to exploit their facial traits in the recognition process helped the success of the face recognition systems, where several face recognition systems were proposed in the literature and have been used by many applications like e-learning, ATM cash dispensers, government agencies, and many daily human applications. However, most of these systems extract features from all the regions of the face, which reduces the effectiveness of these systems because many face regions do not contain any distinctive descriptor. In addition, the face images captured under uncontrolled conditions like poses, occlusions, facial expressions, and especially the illumination variation reduce the accuracy rate dramatically in the proposed systems.

This Chapter presents the contributions of this PhD thesis in the face recognition field. Below, we summarized these contributions:

I. In the first contribution, we introduced a new facial recognition system relying on handcraft methods. The proposed system applied the Scale-Invariant Feature Transform (SIFT [18]) method to detect the discriminant regions containing critical information for distinguishing between faces. Then, the proposed system extracts feature using the Adaptive Local Ternary Patterns (ALTP [19]) and apply the KNN in the classification task. The results of this proposed system have been published in [202].

II. In the second contribution, we proposed a facial recognition system relying on Deep Learning to achieve more accuracy. The system extract features based on the AlexNet-v2 model pre-trained on the ImageNet database [198]. Then, it used the PCA to get the relevant descriptors and applied a robust SVM algorithm called Linear Support Vector Classification (LinearSVC [1]) in the classification stage. The performance of this system has been evaluated and published in [203] .

III. We proposed a novel robust face recognition system under illumination variation in the third contribution. The system used four convolutional blocks from the VGG16 model pre-trained on the ImageNet database to extract features. Then, it used a Global Average Polling (GAP) for the dimensionality reduction and the PCA to keep the relevant features. Finally, In the classification step, the robust-illumination system used the LinearSVC. The details of this system have been published in [204].

This Chapter presents the architecture of each system, followed by experiments on benchmark datasets and comparisons with state-of-the-art methods to prove the efficiency of our systems. In the end, we added a conclusion to summarise the Chapter.

5.2 Contribution I: A robust face recognition system based on handcraft methods

This section presents our proposed face recognition system that uses the SIFT technique to detect discriminant regions and the ALTP to extract features. We chose ALTP because it is specifically designed to handle variations in pose and illumination [19]. Then, it evaluates the performance of our proposed system on two benchmark datasets (ORL and FERET). At the end of the section, we compared our results with state-of-the-art approaches to prove the effectiveness of our proposed handcraft system.

[1]"sklearn.svm.LinearSVC — scikit-learn 0.24.2 documentation", https://scikitlearn.org/stable/modules/generated/sklearn.svm.LinearSVC.html, Accessed on 14/08/2022

5.2.1 The proposed system

We proposed a face recognition system based on hybrid handcraft approaches, combining SIFT key-point and ALTP texture feature extraction techniques. Our proposed system detects the discriminant regions from the pre-processed face image. Then, it extracted local features from the detected regions and calculated the ALTP histograms to use them in the matching stage. Algorithm 1 shows the pre-processing step of our proposed system. An overview of the proposed handcraft face recognition system is illustrated in Figure 5.1.

Algorithm 1 Pre-processing Algorithm

Input: A face dataset.
Output: Training and testing subsets.

1. Load the dataset: load the images and their labels.
2. Convert the images to grayscale: Convert all images to grayscale
3. Resize the images: resize the images to the size of 128 × 128 pixels
4. Split the dataset: Split the dataset into training and testing sets.
5. Save the pre-processed images: Save the pre-processed training and testing images and their labels to separate files

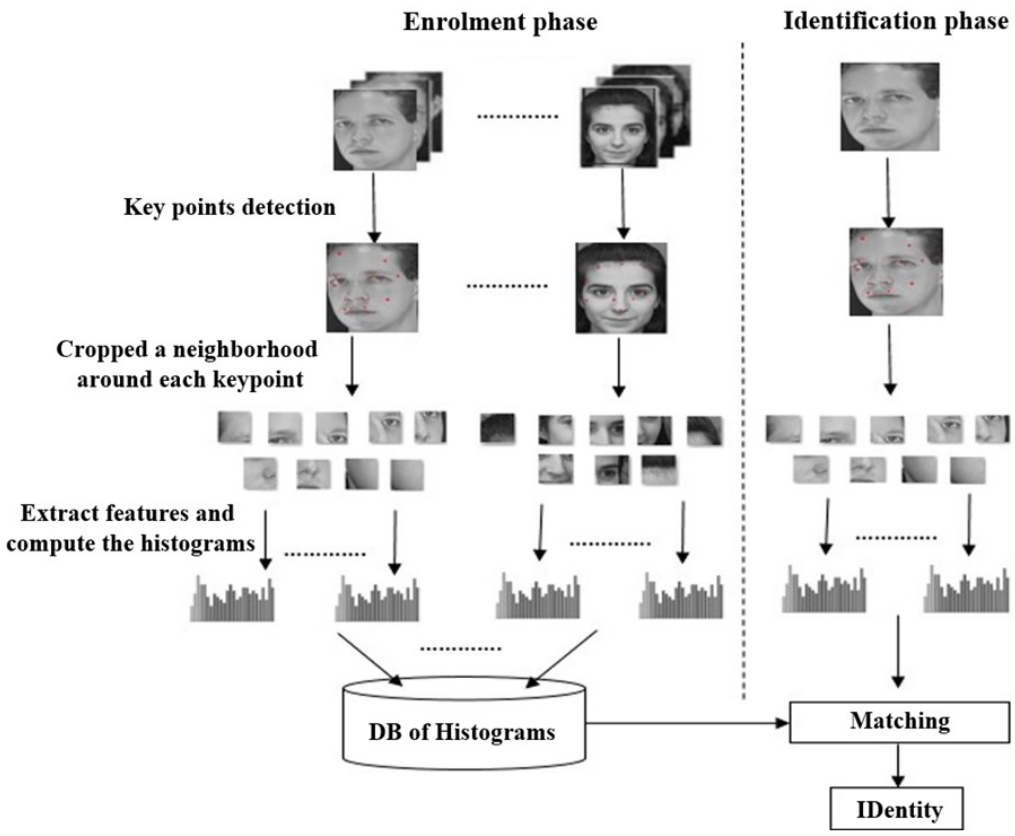


Figure 5.1: Overview of the proposed handcraft face recognition system, adapted from [202]

Our proposed system detects key points from the face images relying on Scale-Invariant Feature Transform (SIFT [18]), and then it extracts a region of 32×32 pixels around each key point. In addition, the system extracts local descriptors from the detected regions based on Adaptive Local Ternary Patterns (ALTP [19]). After that, it counted the ALTP histogram of each region. Furthermore, for the similarity, our proposed handcraft system applies the chi-square distance; it checks the similarity between the ALTP histograms stored in the enrolment stage and the ALTP histograms of the testing image. More details about our proposed system are illustrated in Figure 5.2.

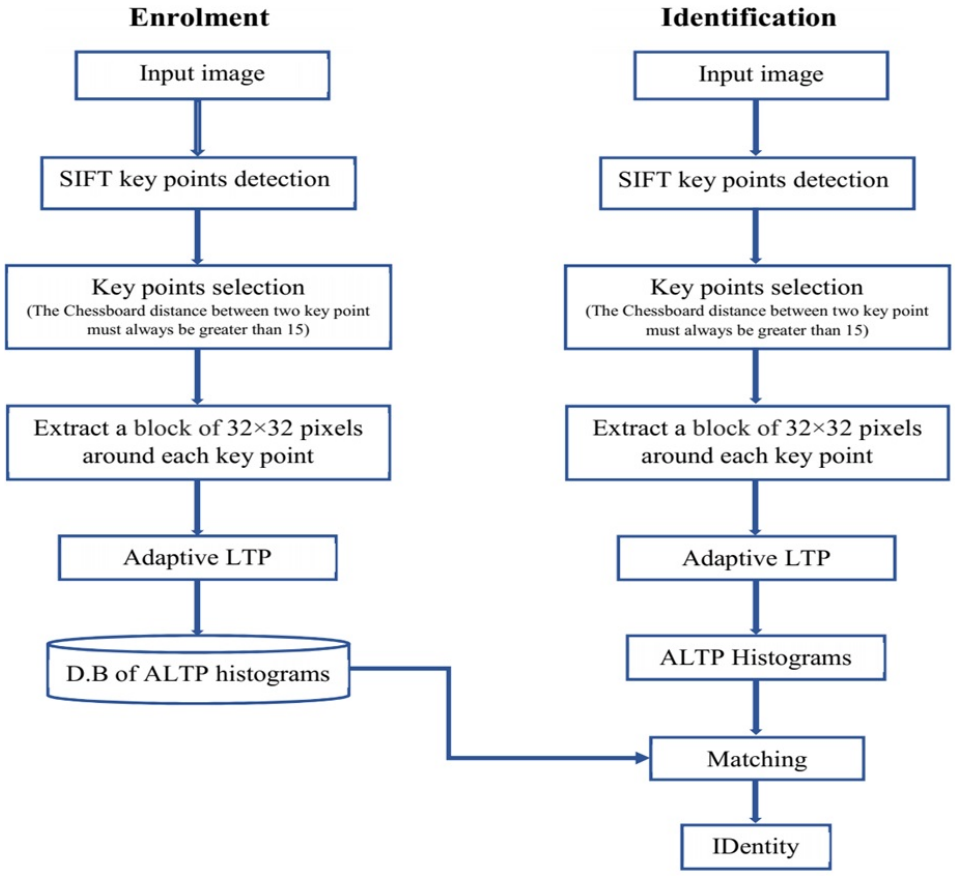


Figure 5.2: Diagram of the proposed handcraft face recognition system, adapted from [202]

Our proposed system is built relying on three principal modules that are briefly presented below:

I. Key points detection and discriminant regions extraction: Our system used SIFT technique to detect key points; we chose this technique because it is robust to noise and various image transformations, including rotating and scaling [18]. SIFT key-point detection technique is based on the three following steps:

a. Build the Difference of Gaussian Pyramid.

b. Detect the local minima and maxima in the created pyramid

c. Illuminate the key points that have Low-Contrast

Then, it extracts a block of 32×32 pixels around each key point, in which the chessboard distance between the selected points is greater than 15 pixels. Figure 5.3 illustrates an example of the key points detected by SIFT. For more detail about SIFT key point extraction (see Chapter 3 section 3.2.1.2).

Figure 5.3: Key points localized by SIFT technique

II. Count the ALTP histograms : Our system used the ALTP to extract local features from the extracted blocks. The ALTP is effective against variations in pose and illumination [19]. In addition, it is more robust to the influence of noise than LTP [100], where it selects an automatic threshold based on Waber's Law to build the ALTP codes. (More detail about ALTP is presented in Chapter 3 section 3.2.1.2)

After computing the ALTP Codes (upper code and lower code), we generate the histogram of each code and concatenate them to obtain the ALTP histogram.

III. Similarity check: In the similarity checker step, we used the chi-square distance, a robust technique that obtained good results in similarity assessment between face feature histograms [205]. We calculate the chi-square distance between each ALTP histogram of the testing image and the histograms of training images, and we keep the minimum distance. Then we compute the sum of the minimum distances to define the person that has a minimum sum. The following equation shows how we calculate the chi-square distance between two ALTP histograms (x and y), where n is the length of the histogram.

$$X^2 = \frac{1}{2}\sum_{i=0}^{n}\frac{(x_i - y_i)^2}{(x_i + y_i)} \tag{5.1}$$

5.2.2 Experiments

In this work, we proposed a hybrid facial recognition system relying on two handcraft methods (SIFT and ALTP). To prove the robustness of our system, we evaluate the performance of the proposed system on two popular benchmark face datasets collected under different conditions. We conducted experiments based on the Five-fold cross-validation protocol for the efficient use of the available data; because this protocol divided the face dataset into five partitions, as all partitions are used both for training and validation purposes. The Five-fold cross-validation protocol obtains representative results; each fold serves as a test set once, while the other four folds serve as the training set.

In this sub-section, we describe the datasets used in our experiment. Then we compare the accuracy rate of our proposed system with robust handcraft methods used in the literature for facial recognition. At the end of the sub-section, we compare our results with current state-of-the-art methods.

5.2.2.1 Datasets

We used the ORL and the FERET databases to evaluate the effectiveness of our system. Below, we introduce the two datasets briefly.

I. The ORL database: The ORL database [2] (also known as the AT&T face database) is a famous face database collected by AT&T Laboratories of Cambridge University between April 1992 and April 1994. The database contains 400 face images captured from 40 persons, containing ten different images for each person. The face images are captured at different times with variations in lighting, facial expressions (smiling / not smiling, open/closed eyes), and accessories (glasses/ no glasses). The researchers used a dark background in the lab to collect the face images, where the size of the images is 112×92 pixels. Figure 5.4 shows a few sample images from the ORL database.

II. The FERET database: The FERET database [206] is a face benchmark used to evaluate the new face recognition algorithms. It was created by the Facial Recognition Technology (FERET) program that the DOD Counterdrug Technology Program has sponsored. Our experiments used a subset comprising 1400 face images captured from 200 persons, in which each individual was captured from seven different poses. In addition, the subset contains variations in illumination and expression. The names of the images used in our experiment are marked with: "bk", "bj", "bg", "bf"," be", "bd", and "ba". Simples from the FERET database are illustrated in Figure 5.5.

[2] "AT&T laboratories of cambridge university ,the database of faces," http://cam-orl.co.uk/facedatabase.html, Accessed on 12/09/2022

Figure 5.4: Images selected from the ORL database

Figure 5.5: Seven face images selected from the FERET subset used in our experiments.

5.2.2.2 Experimental Results

We evaluate our proposed handcraft system on the ORL dataset to prove its robustness against lighting, facial, and accessories variations, in which we applied a five-fold cross-validation protocol. We divide the ORL database into five equal parts. The evaluation is then repeated five times; in each, we use one part as a testing set and the rest four parts as a training set. In the end, we calculate the average accuracy by taking the mean of the five obtained accuracy rates.

Our system achieved an extraordinary accuracy rate that outperformed some famous methods, where the proposed facial recognition system reached 99.75% on the ORL database. However, the methods used for comparison with us achieved limited accuracies, in which the LTP and SIFT achieved an accuracy rate that outperformed the LBP and ALTP. Still, this face recognition accuracy is lower compared to ours. Figure 5.6 shows the effectiveness of our face recognition system compared to LTP, LTP, ALTP, and SIFT methods on the ORL database.

In addition, we used the FERET benchmark datasets to evaluate the performance of our face recognition system. We achieved remarkable accuracy on the FERET dataset compared to

the other methods, where we reached 95.12% by applying the five-fold cross-validation protocol. However, the LBP, LTP, ALTP, and SIFT achieved worse results than our system, where the best accuracy rate was achieved by ALTP (81.95%). Figure 5.6 illustrates the robustness of our proposed handcraft system on the FERET compared to other methods.

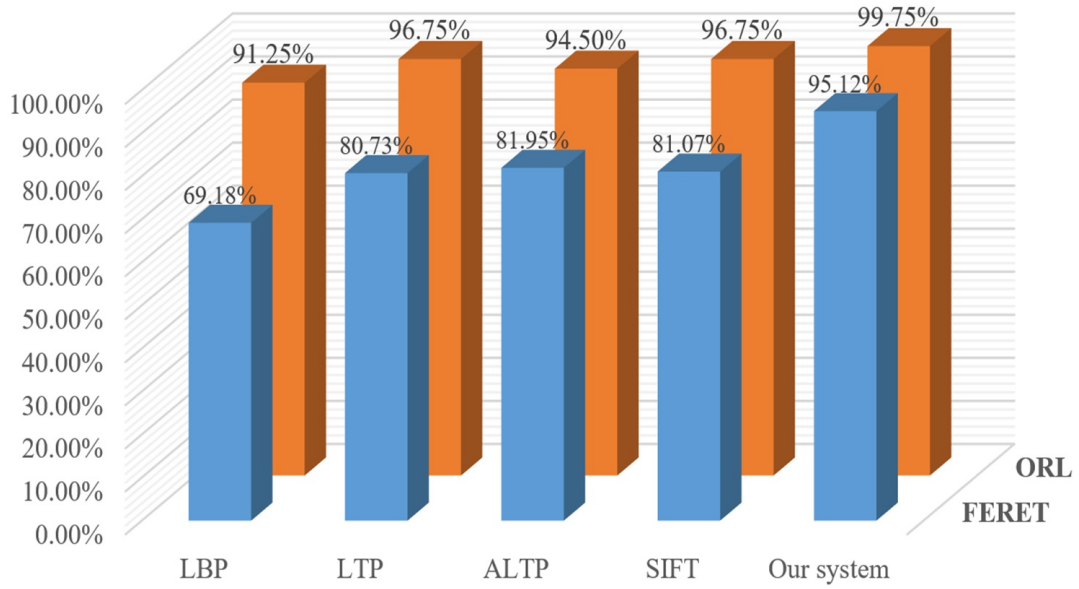


Figure 5.6: The average accuracy rate of our proposed system compared to the accuracies obtained by some popular handicraft techniques on the ORL and FERET databases.

5.2.3 Comparison with state-of-the-art methods

In this sub-section, we compare the accuracy rate of our proposed handcraft system with state-of-the-art face recognition systems evaluated on the ORL and FERET databases. We started the comparison with the approach evaluated on the ORL database, and then we moved to the methods tested on the FERET database.

By applying the five-fold cross-validation protocol on the ORL database, our proposed handcraft system achieved an accuracy face recognition rate of 99.50%. The obtained result outperformed the accuracies of the recent works presented in Table 5.1.

In addition, we achieved an encouraged accuracy rate on the FERET database, in which we reached 95.12% by applying a five-fold cross-validation protocol. Our system accuracy rate on the FERET database surpassed current approaches on this database (See Table 5.1). Still, the obtained accuracy is lower than our ambitions. The proposed handcraft system achieved limited results on the FERET dataset compared to the ORL dataset, because the first dataset is three times greater than the second. In addition, the FERET datasets were captured under more complex conditions. Recently Deep Learning could overcome the limitation of the handcraft

methods, such as recognizing images taken in non-controlled environments and working with big datasets, which Deep Learning achieved high success in several challenging computer vision tasks [14, 207]. We will use Deep Learning to get a more accurate recognition rate in the second proposed face recognition system.

Dataset	Year	Method	Protocol	Accuracy
ORL	2020	Ouyang et al. [114]	Train: 80%, Test: 20%	97.22%
	2019	Ouslimani et al. [115]	Train: 80%, Test: 20%	98.61%
	2020	Sapijaszko and Mikhael [116]	Train: 80%, Test: 20%	98.80%
	2022	Elaggoune et al. [129]	Train: 70%, Test: 30%	99.50%
	2022	Elaggoune et al. [129]	Train: 70%, Test: 30%	99.17%
	2019	Min et al. [130]	Train: 10%, Test: 90%	97.77%
	2022	Hosgurmath et al. [131]	Train: 80%, Test: 20%	93.10%
	2022	Tamilselvi and Karthikeyan [3]	/	97%
	2022	Yallamandaiah and Purnachand [133]	Train: 70 %, Test: 30 %	98.48%
		Our proposed system	**5-fold cross-validation**	**99.75%**
FERET	2019	Ouslimani et al. [115]	Train: 71%, Test: 29%	93.46%
	2020	Qin et al. [117]	Train: 71%, Test: 29%	76.50%
	2020	Qin et al. [117]	Train: 71%, Test: 29%	78.00%
	2019	Min et al. [130]	Train: 14%, Test: 86%	93.04%
	2018	Zeng el al. [136]	Train: 14%, Test: 86%	93.90%
		Our proposed system	**5-fold cross-validation**	**95.12%**

Table 5.1: The average accuracy of our proposed handcraft face recognition system compared to state-of-the-art face recognition systems.

5.3 Contribution II: An effective face recognition system based on Deep Learning

Deep Learning, particularly the Deep Convolutional Neural Network, achieved high success in biometrics, especially in the face recognition field [208].

In this section, we will apply the Deep Convolutional Neural Network to extract discriminant features from the face modality to improve the face recognition accuracy rate. We divide this section into three sub-section. We present the proposed method in the first sub-section and the experiment results described in the second part. Finally, we compared the results of our proposed deep Learning face recognition system with state-of-the-art methods.

5.3.1 The proposed system

The feature extraction stage is one of the most critical tasks in the face recognition field. The AlexNet model is the first deep Learning model that won the ILSVRC 2012 competition [107]. It achieved the best accuracy rate on the ImageNet database, opening the door to the success of Deep Learning in image recognition.

Our proposed face recognition system extract features based on the second version of the AlexNet model pre-trained on the ImageNet dataset. The ImageNet dataset has 1.4 million images divided into 1000 classes; 17% of the ImageNet's images contain at least one face [209]. We used a pre-trained model because it can obtain high accuracy without requiring large amounts of face image training data. Our system used the Alexnet-v2 model without the fully connected layers and without fine-tuning its weights on the new databases to improve performance and save computational resources and time.

The AlexNet-v2 model includes only three convolutional blocks. The first and the second have only one convolutional layer followed by a max-pooling layer, and the third block has three convolutional layers followed by a max-pooling layer. Our proposed system used these three convolutional blocks to extract features and applied the LinearSVC in the classification stage, where it applied the PCA to reduce the dimensionality and keep relevant features before applying the LinearSVC classifier. Our system architecture is shown at the bottom of Figure 5.7, and the pre-trained Alexnet-v2 model is shown at the top of Figure 5.7.

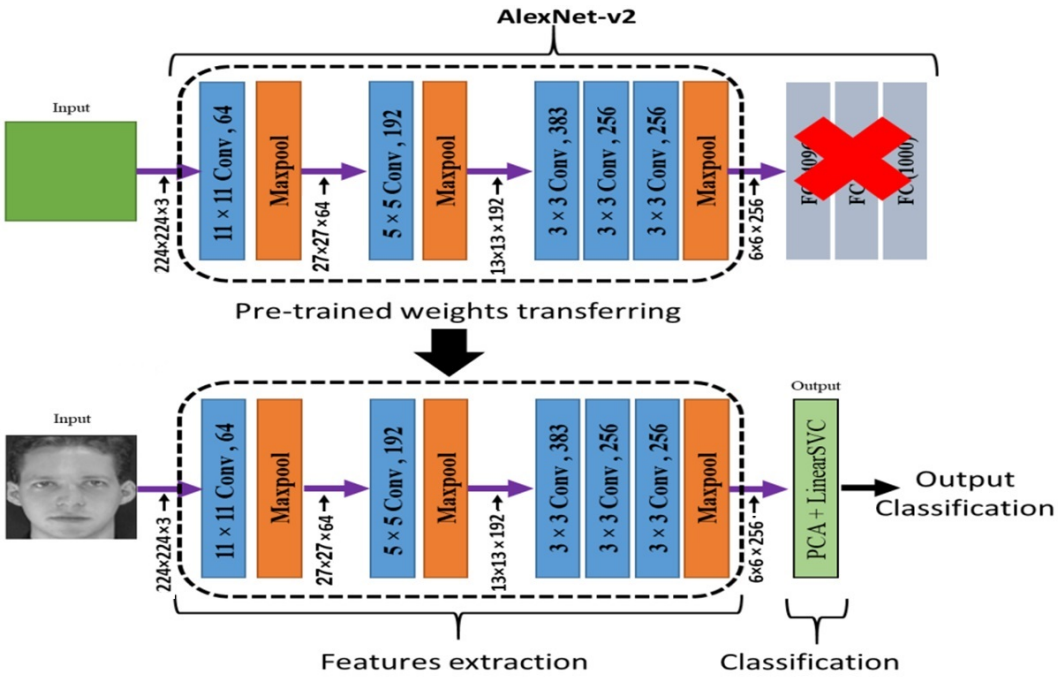


Figure 5.7: The architecture of our proposed face recognition system based on Deep Learning.

Algorithm 2 shows the structure of our proposed face recognition system based on Deep Learning.

Algorithm 2 The face recognition system building by Deep Learning.

Input: A face dataset.
Output: The system's performance (accuracy rate).

1. Load the images and their labels.
2. Resize the images to the size of 224×224×3 pixels.
3. Split the dataset into training and testing sets.
4. Use the pre-trained AlexNetv2 model to extract features from each training image.
5. Apply PCA to the training feature vectors to reduce their dimensionality.
6. Train the LinearSVC classifier using the reduced training feature vectors and their corresponding labels.
7. Use the pre-trained AlexNetv2 model to extract features from each testing image.
8. Apply PCA to the testing feature vectors with the same parameters as the training feature vectors.
9. Use the trained SVM classifier to predict the labels for the reduced dimensional testing feature vectors.
10. Compare the predicted labels to the true labels, and calculate the accuracy rate of the proposed system by dividing the number of correct predictions by the total number of the testing images.
11. Return the performance of the system.

5.3.2 Experiments

To prove the effectiveness and robustness of our proposed face recognition, we used the two face benchmarks that were used to evaluate the previously proposed handcraft system. We started the experiments based on the ORL database and then conducted the experiments on the FERET dataset. In addition, we compared our accuracy rate with relevant deep CNN models trained from scratch. Furthermore, we compared our system accuracy using ImageNet pre-trained models and CNN models trained from scratch with Data Augmentation. At the end of this sub-section, we illustrate the superiority of our proposed face recognition system compared to state-of-the-art approaches.

5.3.2.1 Experimental Results

We trained the Xception, Inception-v3, ResNet50, VGG16, and AlexNet-v2 from scratch on the ORL database. Then we removed each model's classification layers and applied PCA to reduce the dimensionality and LinearSVC for the classification task. The five models achieved weak accuracies; the best face recognition accuracy did not surpass 92%.

We augmented the database using Zoom, Flip, and Random Rotation Augmentation techniques to improve the accuracies because the Deep CNN model has hangry to the data. The

accuracy rate of the models has been enhanced, where the Inception-v3 reached 97.50%. Still, this accuracy rate is lower compared to the accuracy achieved by our proposed system (100%).

Transfer learning is an effective technique used in the literature to enhance the performance of the DL models on tasks with small datasets. The five models pre-trained on ImageNet achieved high results on the ORL database, where the VGG16 reached our accuracy rate (100%). However, the VGG16 is very complex compared to our proposed system. It has 13 convolutional layers, but our system contains only three convolutional layers. Figure 5.8 illustrates the superiority of our proposed system compared to the different Deep Learning models trained from scratch, using Data Augmentation and pre-trained architectures.

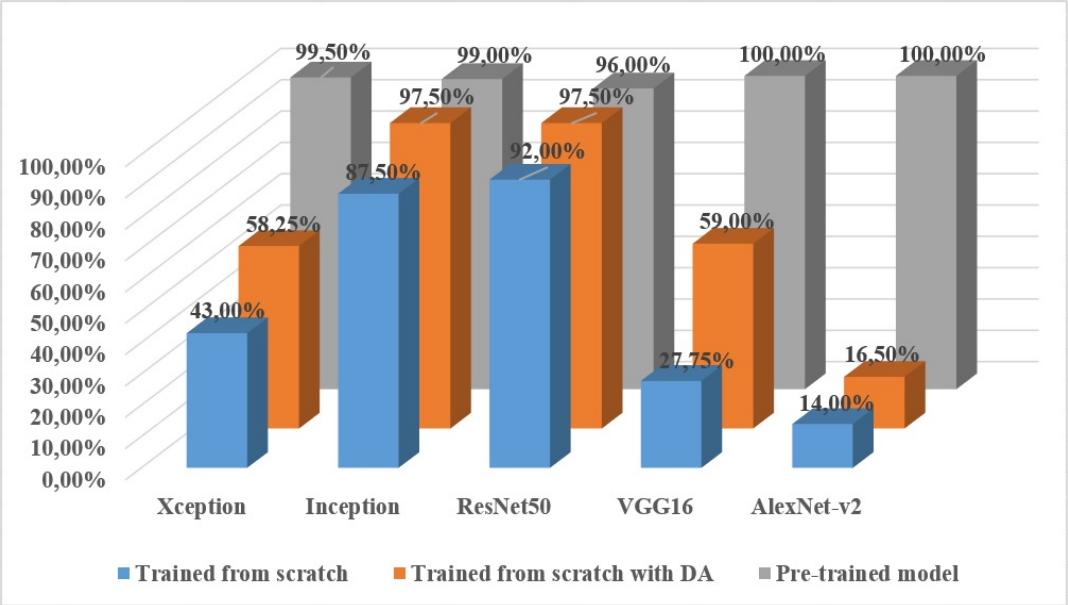


Figure 5.8: The accuracy rate of our proposed Deep learning face recognition system (pre-trained AlexNet-v2+PCA+LinearSVC) compared to the accuracies of various deep learning models on the ORL database.

The models trained from scratch on the FERET dataset achieved lower accuracies, except the Xception model, which reached an accuracy rate of 95.50% that surpassed our proposed handcraft system. Still, the accuracy rate of the Xception model trained from scratch is lower compared to the accuracy rate of our proposed deep learning system, reaching 99.89%. Figure 5.9 illustrates that the Data Augmentation improved the recognition accuracy of the different models. Still, the new accuracies are lower compared to our face recognition system.

The ImageNet pre-trained models achieved remarkable results that outperformed those achieved by the models trained from scratch and with Data Augmentation. However, our proposed face recognition system achieved the best accuracy on the FERET database, reached to 99.89% by applying the five-fold cross-validation protocol (See figure 5.9).

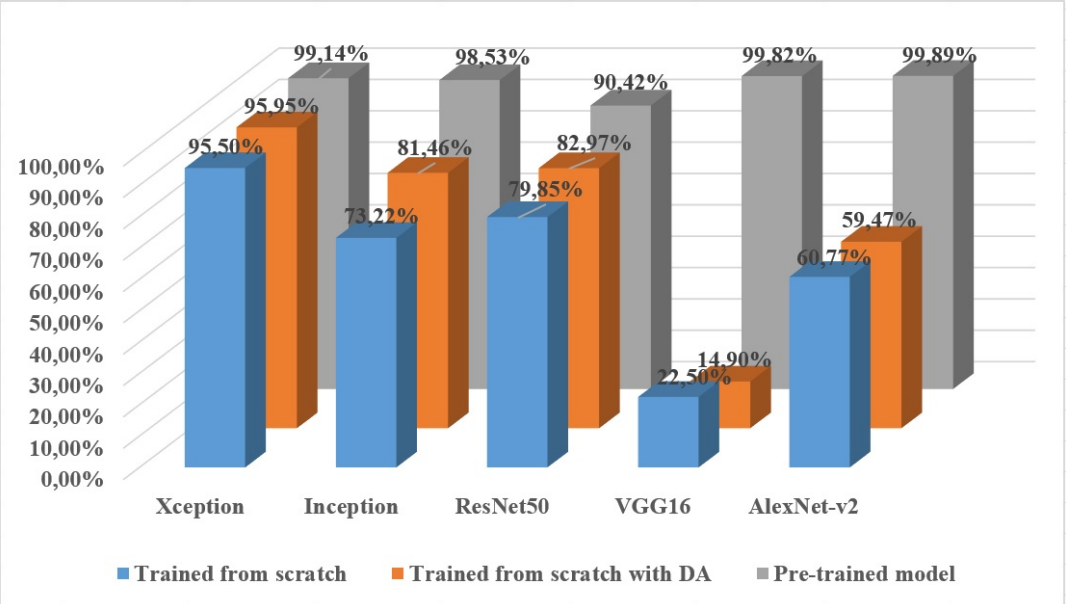


Figure 5.9: Comparision of the accuracy rate of our proposed Deep learning face recognition (pre-trained AlexNet-v2+PCA+LinearSVC) with the accuracies of various deep learning models on the FERET dataset

5.3.3 Comparison with state-of-the-art methods

Our proposed face recognition system based on Deep Learning proved its robustness and effectiveness compared to state-of-the-art methods on the ORL and FERET databases. The accuracy rate of the proposed system reached 100% on the ORL dataset by applying five-fold cross-validation; this accuracy outperforms all the recent methods presented in Table 5.2.

Moreover, our proposal achieved excellent accuracy on the FERET dataset, reaching 99.89% by applying the five-fold cross-validation. To the best of our knowledge, this accuracy is state-of-the-art on the FERET database.

5.4 Contribution III: An Illumination-Robust facial recognition system relying on Deep learning

The face images captured in non-cooperative environments, such as variations of illumination, poses, and expressions, pose a considerable challenge in face recognition, where the lighting variation affects the performance of the facial recognition systems more severely [95].

In this section, we proposed a face recognition system robust to illumination variation. We present our proposal's architecture and evaluate its performance on two benchmarks collected under lighting variation and poor illumination. In addition, we compared state-of-the-art approaches to our face recognition accuracies.

Dataset	Year	Method	Protocol	Accuracy
ORL	2020	Ouyang et al. [114]	Train: 80%, Test: 20%	97.22%
	2019	Ouslimani et al. [115]	Train: 80%, Test: 20%	98.61%
	2020	Sapijaszko and Mikhael [116]	Train: 80%, Test: 20%	98.80%
	2022	Elaggoune et al. [129]	Train: 70%, Test: 30%	99.50%
	2022	Elaggoune et al. [129]	Train: 70%, Test: 30%	99.17%
	2019	Min et al. [130]	Train: 10%, Test: 90%	97.77%
	2022	Hosgurmath et al. [131]	Train: 80%, Test: 20%	93.10%
	2022	Tamilselvi and Karthikeyan [3]	/	97%
	2022	Yallamandaiah and Purnachand [133]	Train: 70%, Test: 30%	98.48%
		Our proposed handcraft system	**5-fold cross-validation**	**99.75%**
		Our proposed DL system	**5-fold cross-validation**	**100%**
FERET	2019	Ouslimani et al. [115]	Train: 71%, Test: 29%	93.46%
	2020	Qin et al. [117]	Train: 71%, Test: 29%	76.50%
	2020	Qin et al. [117]	Train: 71%, Test: 29%	78.00%
	2019	Min et al. [130]	Train: 14%, Test: 86%	93.04%
	2018	Zeng el al. [136]	Train: 14%, Test: 86%	93.90%
		Our proposed handcraft system	**5-fold cross-validation**	**95.12%**
		Our proposed DL system	**5-fold cross-validation**	**99.89%**
		Our proposed DL system	**Train: 14%, Test: 86%**	**98.50%**

Table 5.2: The average accuracy of our proposed Deep learning face recognition system compared to state-of-the-art face recognition systems.

5.4.1 The proposed system

Our proposed system used a part of the VGG16 model pre-trained on the ImageNet dataset to extract features from the face images.

VGG16 model is a robust Deep Learning model that achieved good results in several image recognition tasks, which is the runner-up of the ILSVRC 2014. However, the big problem of the VGG16 model is that it requires substantial computational resources to train its parameters because it is built based on about 138 million parameters. To overcome this problem, we proposed a features extraction model that used only four convolutional blocks and has only about 7 million parameters. Furthermore, to save training time, the proposed system used the VGG16 model pre-trained on the ImageNet database without retraining it on the new databases, in which 17% of the ImageNet's images contain face images collected under illumination variations.

After extracting features, we applied a Global Average Pooling (GAP) layer to reduce the special size of the extracted features. Then, PCA is applied to remove noisy or redundant features and keep relevant features. In the classification stage, we used a fast and robust SVM algorithm called LinearSVC. More detail about our proposed architecture illustrates in Figure 5.10 .

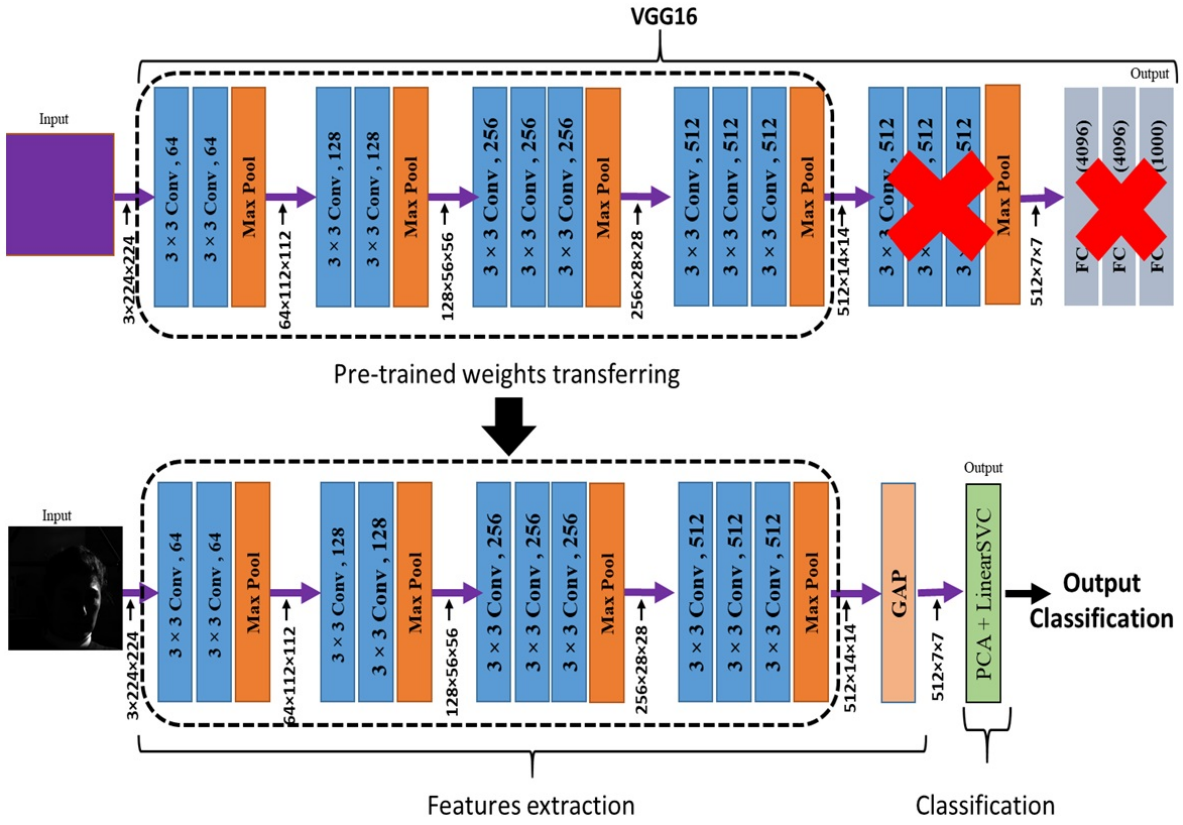


Figure 5.10: The architecture of our proposed face recognition system under lighting variation based on Deep Learning, adapted from [204]

5.4.2 Experiments

To prove the robustness of our proposed face recognition system under illumination variation, we conducted many experiments on two benchmark face datasets that contain images captured under lighting variation and poor illumination. We briefly reviewed the two datasets before starting the experiments.

5.4.2.1 Datasets

In our experiment, we used the Extended Yale B and AR face databases that were exploited in the literature to evaluate the performance of the face recognition system under lighting variation.

Below, we describe the component of each database.

I. The Extended Yale B Database: The Extended Yale B database [210] contains 16128 face images of 28 individuals collected under nine poses and 64 lighting conditions (28×9×64=16128).

We split this database into five subsets in our experiments based on the angle between the direction of the illumination source and the central camera axis. Table 5.3 illustrates how we split the 64 illumination conditions into five subsets (7,12,12,14,19) and the number of samples in each subset (number of samples= number of illumination conditions × number of poses). Figure 5.11 shows two face images from each subset.

Subsets	illumination angle in degrees	Number of faces
One	0-12	63
Two	13-25	108
Three	26-50	108
Four	51-77	126
Five	>77	171

Table 5.3: The number of face images and the Illumination angles in each subset

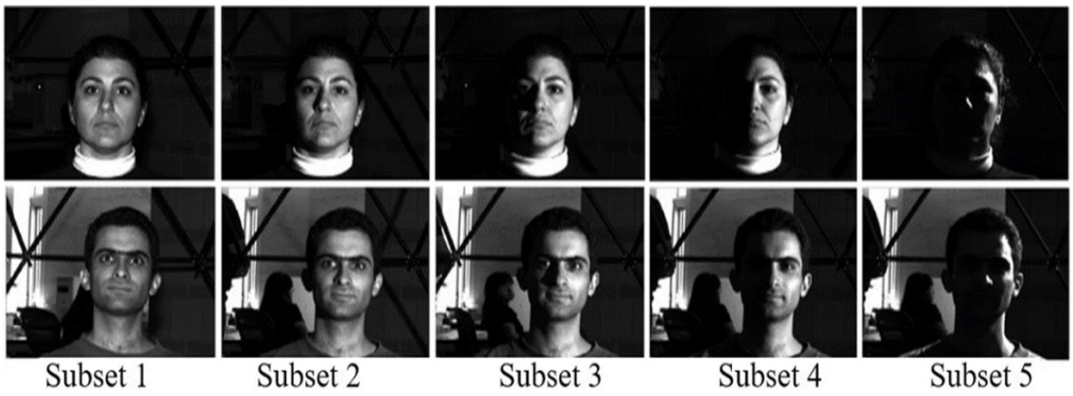


Figure 5.11: Images from the Extended Yale B dataset split based on illumination conditions.

II. The AR Face Database: This database [211] comprises four thousand images captured from 162 individuals (56 women and 70 men). The database was collected under uncontrolled conditions like occlusion and the variation in facial expression and illumination.

In our experiment, we used a dataset comprising 1400 images collected from 100 persons; half of them are women, and each human has 14 face images. The selected dataset contained variations in illumination and facial expression. Figure 5.12 shows samples from the AR database.

Figure 5.12: Face images of a person included in the AR database

5.4.2.2 Experimental Results

To prove the effectiveness of our proposed system, we conducted many experiments on the Extended Yale B and the AR face databases. Then we compared our results with current state-of-the-art techniques.

We started with Extended Yale B Database. We calculated the accurate recognition rate of our system, where Subset 1 is used for training because it is captured under good illumination conditions. For the testing, we select a subset collected under bad illumination conditions (Subset 3, Subset 4, or Subset 5). Table 5.4 shows the extraordinary accuracy rate achieved by our proposed system.

	Subset 3	Subset 4	Subset 5	Average
Accuracy rate	100%	99.63%	98.33%	99.32%

Table 5.4: The accuracy of our face recognition system on the Extended Yale B database, where Subset 1 is used for training

To illustrate the robustness of our accuracies compared to different Deep CNN models, we trained relevant models from scratch on the Extended Yale B database. The models achieved a lower accuracy rate, particularly on subset 5, which contains poor illumination. The inception-v3 achieved a good average compared to the models trained from scratch, reaching 93.38%. Still, this accuracy is significantly worse than our result (See Table 5.5).

Moreover, we applied Data Augmentation to enhance the performance of the models trained from scratch, but the results were disappointing. The best average accuracy rate is 92.76%, achieved by Inception-v3.

Compared to the models trained from scratch, the models pre-trained on the ImageNet dataset achieved acceptable results on the Extended Yale B database, where AlexNet achieved an average accuracy rate of 96.02%. Still, this model achieved a worse recognition performance on

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our system
Subset 3	99.97%	99.40%	3.57%	91.10%	84.66%	**100%**
Subset 4	92.15%	94.64%	3.57%	71.54%	67.80%	**99.63%**
Subset 5	81.47%	86.09%	3.57%	52.11%	54.28%	**98.33%**
Average	91.20%	93.38%	3.57%	71.58%	68.91%	**99.32%**

Table 5.5: The accuracy rate of the proposed system on the Extended Yale B database compared to relevant models trained from scratch.

Subset 5, which was captured under poor illumination (See Table 5.6). In addition, the accuracy rate of the pre-trained AlexNet model is lower than ours.

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our approach
Subset 3	98.84%	99.40%	93.19%	99.40%	99.97%	**100%**
Subset 4	96.63%	96.83%	74.83%	90.99%	97.28%	**99.63%**
Subset 5	85.99%	87.07%	55.43%	86.57%	90.81%	**98.33%**
Average	93.82%	94.43%	74.48%	92.32%	96.02%	**99.32%**

Table 5.6: The recognition accuracy rate of our proposed system compared to the different ImageNet pre-trained models on the Extended Yale B.

In addition, we used the AR face database collected under lighting variation to prove the performance of our proposed system. The high face recognition accuracy rate achieved by applying the two-fold cross-validation protocol demonstrates the robustness of our system (see Table 5.7). We applied the two-fold cross-validation for a significant comparison because most of the systems evaluated on the AR database split it into two equal parts, one for the training and the other for the testing.

	Fold 1	Fold 2	Average
Accuracy	99.86%	99.71%	99.79%

Table 5.7: Our accuracy rate on the AR face dataset.

The Inception-v3 and Xception architectures trained from scratch on the AR dataset achieved high performance, where the first reached 97% and the second achieved 99.36%. Still, the accuracy rate of the two models is lower than ours (see Table 5.8). In addition, The ResNet50 and the AlexNet-v2 trained from scratch produced poor results due to the curse of overfitting caused by the small size of the training set, which consists of 100 classes, each having only seven samples.

Furthermore, we trained the models from scratch with Data Augmentation on the AR database, but the result was disappointing; the models did not surpass the accuracy rate achieved

in the previous experiment, where the Xception achieved an average accuracy rate of 98.87% and the Inception obtained 96.12%.

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our approach
Fold 1	99.14%	94.57%	1.00%	71.54%	1.00%	**99.86%**
Fold 2	99.57%	99.43%	1.24%	52.11%	1.00%	**99.71%**
Average	99.36%	97.00%	1.12%	71.58%	1.00%	**99.79%**

Table 5.8: The accuracy rate of our proposed system compared to relevant models trained from scratch on the AR database.

Moreover, the accuracy rate achieved by pre-trained Xception, Inception-v3, ResNet50, VGG16, and AlexNet-v2 models is significantly worse than the result from scratch training. The VGG16 model has achieved the best result (93.43%). However, its accuracy is considerably lower than ours (see Table 5.9).

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our approach
Fold 1	88.86%	88.29%	12.29%	93.29%	1.00%	**99.86%**
Fold 2	87.00%	91.57%	9.00%	93.57%	1.00%	**99.71%**
Average	87.93%	89.93%	10.65%	93.43%	1.00%	**99.79%**

Table 5.9: Our accuracy on the AR dataset compared to the different models pre-trained on ImageNet.

5.4.3 Comparison with state-of-the-art methods

To prove the robustness of our system against illumination variation, in this sub-section, we compared our accuracies with recently proposed robust-illumination face recognition systems that were evaluated on the AR and Extended Yale B databases.

By using subset 1 of the Extended Yale B database for training and subset 3, 4, and 5 for testing, our system achieved an excellent accuracy rate that surpassed state-of-the-art methods [118–120] applied to the same scenario (Figure 5.13 illustrates that).

In addition, by applying the five-fold cross-validation protocol on the Extended Yale B, our approach achieved 100% on subsets 3 and 4. Furthermore, it reached 99.98% on subset 5. These accuracies outperformed the accuracy achieved by IRDRCNN+RGHF [134] and ARCNN [1] (See Figure 5.14)

Moreover, by splitting each subset from the Extended Yale B database into 20% for training and 80% for testing, our proposed system accuracy reached 100% on subsets 3, 4, and 5.

Figure 5.15 shows the superiority of our accuracies compared to Inception-v3+M-CLAHE [2], ResNet50+M-CLAHE [2], and VGG16+M-CLAHE [2].

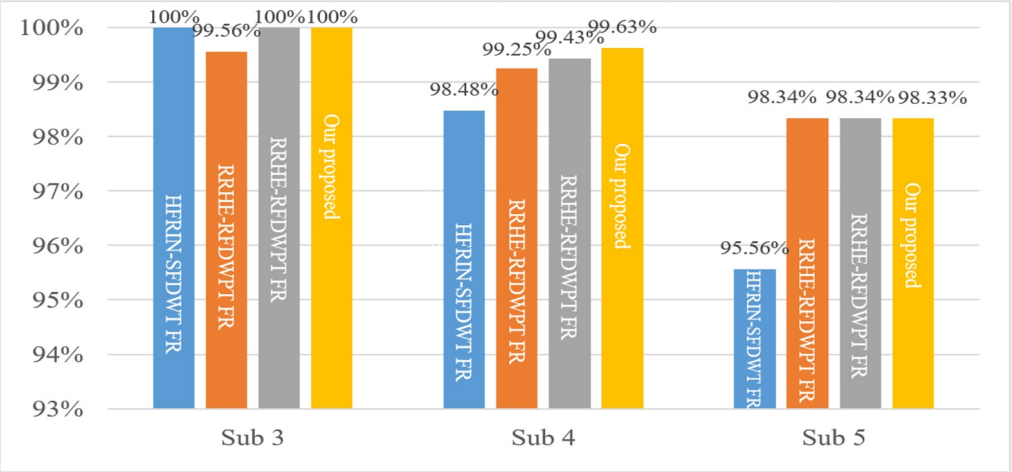


Figure 5.13: Our accuracy rate compared to some state-of-the-art approaches (HFRIN-SFDWT FR [118] , RR-CS IWT + FLDA [119] , RRHE-RFDWPT FR [120]) on the Extended Yale B Database, where Subset 1 is used for training.

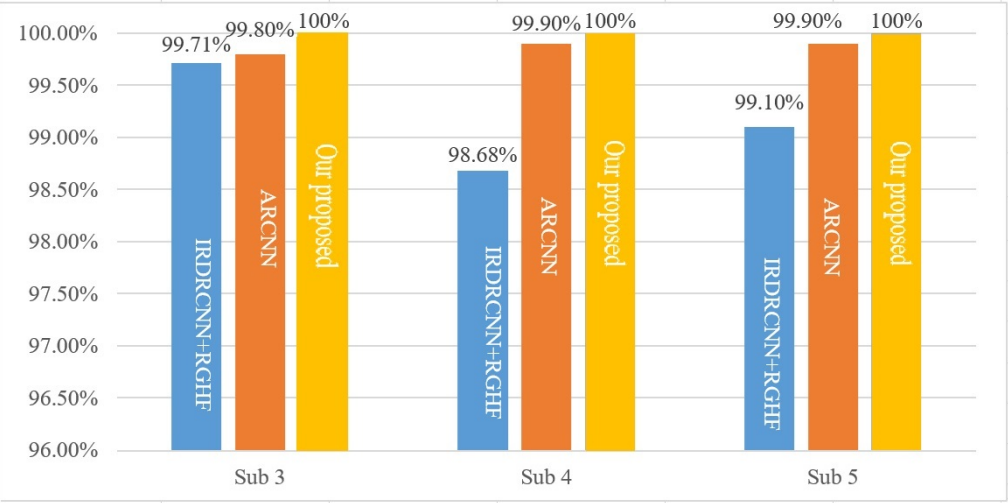


Figure 5.14: The accuracy of some state-of-the-art approaches (IRDRCNN+RGHF [134] and ARCNN [1]) compared to ours on the Extended Yale B dataset, where we use a 5-fold CV protocol.

Furthermore, Our system achieved an excellent accuracy rate on the AR dataset, which reached 99.79 by applying the two-fold cross-validation protocol, 100% by splitting the data into 50% for training and 50% for testing, and 99.89% by splitting the data into 64% for testing and 36% for training. Table 5.10 illustrates the superiority of our system's accuracy rate compared to the results achieved by current state-of-the-art approaches on the same dataset.

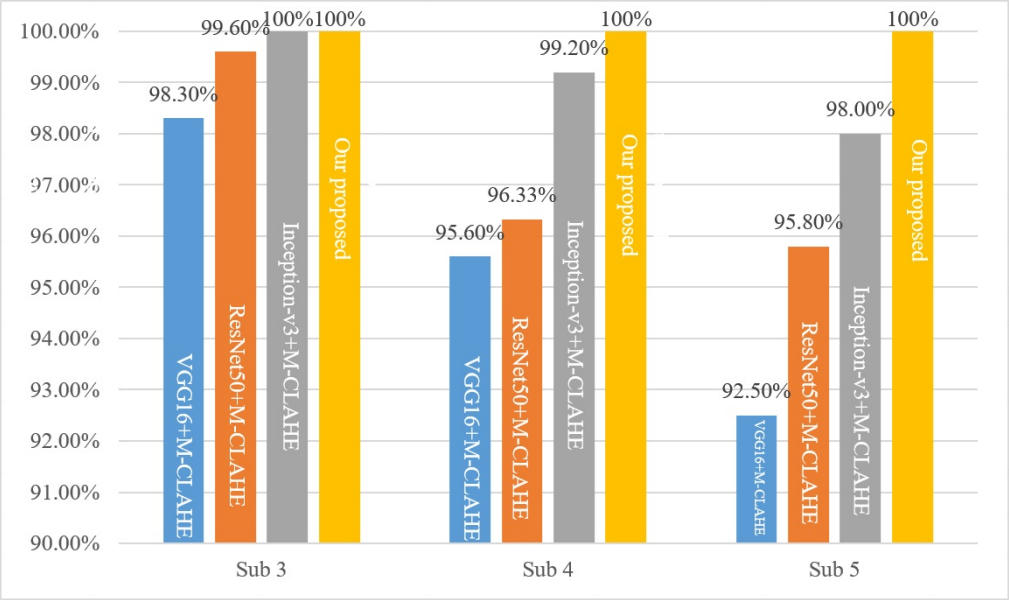


Figure 5.15: Some state-of-the-art methods' face recognition accuracy (Inception-v3+M-CLAHE , ResNet50+M-CLAHE , and VGG16+M-CLAHE [2]) on the Extended Yale B Database compared to our accuracy, in which we used 80% for testing and the rest for training.

Year	Method	Protocol	Average
2018	Wen et al. [121]	Train: 50%, Test: 50%	93.01%
2019	Yuan et al. [122]	Train: 36%, Test: 64%	99.18%
2019	Yuan et al. [122]	Train: 36%, Test: 64%	95.54%
2021	Hu et al. [123]	Train: 50%, Test: 50%	98.81%
2018	Ayyavoo and Suseela [124]	Train: 36%, Test: 64%	98.78%
2018	Dahmouni el al. [125]	Train: 50%, Test: 50%	98.09%
2018	Dahmouni el al. [125]	Train: 50%, Test: 50%	98.45%
	Our approach	**Train: 36%, Test: 64%**	**99.89%**
		Train: 50%, Test: 50%	**100%**
		2-fold cross-validation	**99.79%**

Table 5.10: Some state-of-the-art approaches' face recognition accuracy compared to our performance on the AR dataset.

5.5 Conclusion

In this Chapter, we proposed three systems used for facial recognition; the first relies on handcraft methods, the second uses Deep Learning, and the third is proposed to recognize face images captured under illumination variation and poor lighting.

In the first contribution, we proposed a hybrid face recognition system based on SIFT and ALTP. Our system extracts features from the relevant regions, which applies SIFT to detect the critical regions, and then uses ALTP to extract discriminant features from the detected regions. For the classification stage, our system used KNN based on chi-square distance. The system achieved acceptable results, reaching 99.75% on the ORL database and 95.12% on the FERET dataset.

In the second contribution, we proposed a new system based on deep CNN to achieve more accuracy rate. The system applied the AlexNet-v2 model pre-trained on the ImageNet dataset to extract features from the face images. PCA was applied for dimensionality reduction, and LinearSVC was used for the classification. This system achieved a new state-of-the-art accuracy rate, reaching 99.89% on the FERET dataset and 100% on the ORL database.

In the third contribution, we proposed a robust face recognition system for the facials collected under illumination variation and poor lighting. The system extract features relying on four convolutional blocks from the VGG16 model pre-trained on ImageNet. Then it added a GAP layer to reduce the dimensionality and PCA to keep relevant components. Finally, for the classification stage, the system used LinearSVC. The system achieved extraordinary accuracy compared to the state-of-the-art method, where an average accuracy rate reached 99.32% on the Extended Yale B database and 99.79% on the AR face database.

CHAPTER 6

IRIS RECOGNITION SYSTEM

"The striking thing about an iris pattern is that it contains a great deal of randomness, making each one very unique."

John Daugman, Computer Scientist.

6.1 Introduction

Iris recognition is one of the most efficient and reliable modalities in the biometric field. It received extraordinary attention in the research community in the recent three decades, where it is considered one of the most promising biometric traits. The iris recognition offers several characteristics, for example, but not limited to: the iris texture contains distinctive descriptors that can be used to recognize persons, where the iris trait achieved high performance even to differentiate between identical twins [4]. In addition, the iris texture is invariant over a person's lifetime [212]. Furthermore, with the development of cameras, we can capture the iris region without touching the sensors, limiting the potential spread of the COVID-19 pandemic and some other diseases by touch.

The iris recognition technology has been used in several tasks that require a high level of security, such as border security, airport, government, and military agencies. Many works have been published in recent years to develop robustness systems that can extract effective features from iris images [153, 154, 213] . However, most of the proposed systems achieved a lower accuracy rate with the images captured under non-cooperative conditions, such as

lighting variation, specular reflection, blur, pupil constriction/ dilation, and occlusion due to eyelid/glasses/eyelash [214]. In addition, most of the proposed systems extract features after segmenting and normalizing the iris region, which affects the features' effectiveness and consumes computational resources. Furthermore, some previous papers applied famous deep CNN models to build robust iris recognition systems [154, 157]. However, these models contain tens to hundreds of millions of parameters. But, there is not a massive iris database to train this considerable number of parameters, where the most extensive available database contains only 117,503 images [215].

This Chapter presents the contributions of this Phdthesis in the iris recognition field. We proposed a robust iris recognition system that used Yolov4-tiny to detect the iris region. Then, the system extracts features from the detected region without iris segmenting and normalizing. In addition, we proposed an effective deep CNN model that reduces the architecture of the Inception-v3 model and accelerates the feature extraction task. The proposed model used only five inception modules from the ImageNet pre-trained Inception-v3 model, and without fine-tuning them on the iris database. Therefore, the proposed CNN model saves training time and does not require massive data. Moreover, the system used the PCA to keep relevant iris features and the LinearSVC in the classification stage. The performance of this system has been evaluated and published in [216].

Chapter 6 presents the architecture of our proposed iris recognition system, then conducts multiple experiments on many benchmark datasets that were captured under challenging conditions. In addition, this Chapter compared our results with state-of-the-art methods to prove the efficiency of the proposed system. Finally, at the end, we added a conclusion to summarise this Chapter.

6.2 Contribution IV: A robust iris recognition system based on the pre-trained Inception-v3 model

We proposed an effective iris recognition system that used the Yolov4-tiny model to detect the iris region. A new proposed CNN model inspired by the pre-trained Inception-v3 model is used to extract features, PCA to reduce the dimensionality, and LinearSVC for the classification task.

This section presents the detail of the proposed iris recognition system. Then, it evaluates his performances on four benchmark datasets. At the end of the section, we compared our accuracy rate with state-of-the-art methods to prove the effectiveness of our proposed system.

6.2.1 The proposed system

We proposed a robust iris recognition system relying on the fine-tuned YOLOv4-tiny and the pre-trained Inception-v3 models; the first model uses to detect the iris region, and the second applies to extract features. The architecture of the system is presented in Figure 6.1.

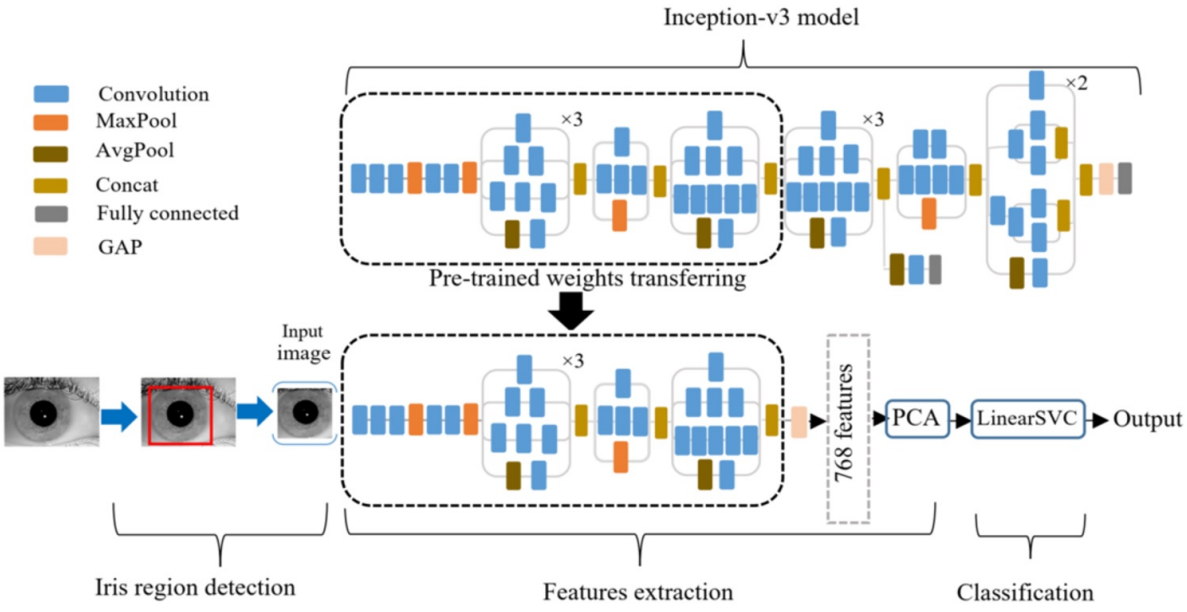


Figure 6.1: The architecture of our proposed iris recognition system based on the pre-trained Inception-v3 model, adapted from [216].

To clearly present our proposed iris recognition system, we divide this section into three sub-sections: we present the iris detection stage in the first sub-section. Then, the second sub-section introduced our proposed feature extraction model. Finally, in the third sub-section, we discuss the classification task.

6.2.1.1 Iris detection

In this step, we used a fine-tuned Yolov4-tiny model to detect the iris region before inputting it into our proposed feature extraction model. We extract the features from the detected iris region without pre-processing it.

To fine-tune the MS COCO pre-trained Yolo4-tiny model weights, we randomly selected 400 iris images from CASIA-Iris-V1, IITD, CASIA-Iris-Interval, and CASIA-Iris-Thousand databases (100 images from each database). Then, we manually annotated the iris regions bounding boxes of the chosen images before using them to train the Yolov4-tiny model. Some examples of iris region detection obtained by the fine-tuned Yolov4-tiny model are shown in Figure 6.2.

6.2.1.2 Feature extraction

The inception-v3 model achieved high success in the last years, with an accuracy rate reaching 94.4% on the ILSVRC 2015. Recently, this famous model has been applied to many classification problems and obtained good results [217, 218].

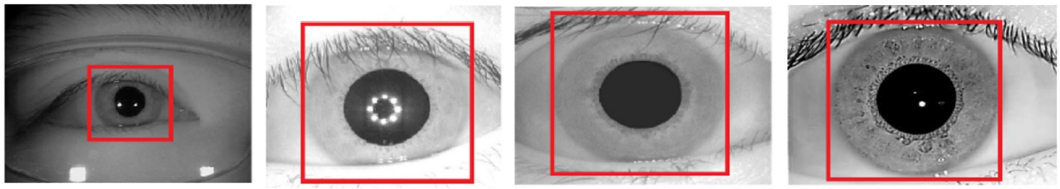


Figure 6.2: Examples of iris region localization

Our system extracts iris features based on only a part of the pre-trained Inception-v3 model. The system used off-the-shelf weights trained on the ImageNet dataset without fine-tuning them on the iris dataset.

The Inception-v3 model is built based on 11 inception modules (see Figure 6.1), containing more than 23 million parameters. Our system proposed a new deep CNN architecture inspired by the pre-trained Inception-v3 model to reduce the inception-v3 model's complexity. The new architecture used only five inception modules and was built based on only 7 million parameters. We selected five inception modules because we tried to use less than five, but the system's accuracy rate was dramatically degraded in this case. In addition, the proposed system added a GAP layer after the fifth module to reduce the features' size from 37.632 to 768. Furthermore, the system applied PCA with default parameters to keep relevant features.

6.2.1.3 Classification

In the classification task, our proposed iris recognition system applied a simple and fast SVM algorithm called LinearSVC. This SVM algorithm has been chosen as the most appropriate Machine Learning technique to solve classification problems [219]. In addition, it has recently been used to solve several classification problems and achieved high success [220, 221].

We trained the LinearSVC classifier using the training images' relevant features and corresponding labels. Then the system used the trained classifier to predict the labels of the testing iris images based on the features extracted from them.

6.2.2 Experiments

To prove the robustness of our proposed iris recognition system, we conducted many experiments on four benchmark iris datasets captured under unconstrained conditions; we present the used datasets in sub-section 6.2.2.1.

The experiments were carried out on the Google Colab [1] environment, using Keras [222] and Pytorch [223] frameworks.

[1] "Google Colab", https://colab.research.google.com

We calculated the accuracy rate of our proposed system on the IITD database by applying the train/test split and cross-validation protocols. Then, we compared our results obtained by the five-fold cross-validation protocol to famous deep CNN models trained on the unprocessed iris region images. For a meaningful comparison, we trained the deep CNN models from scratch. Then, we used Data Augmentation (DA) to improve the accuracy rate. In addition, we conducted a third experiment in which we trained the classification layers of the pre-trained models.

We used the different Deep CNN models without any modifications, in which we changed only the output neuron number. Figure 6.3 (b) shows an iris region that has not been subjected to any pre-processing. Figure 6.4 shows samples after applying DA on iris region images from the IITD database.

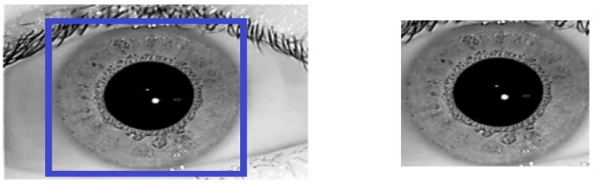


Figure 6.3: Iris region detection, (a) the detected iris region and (b) the iris region image.

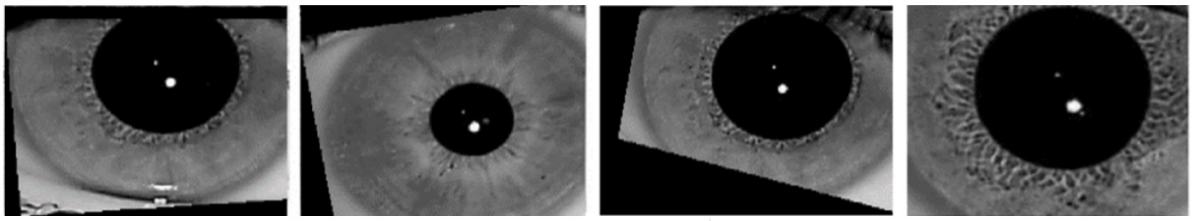
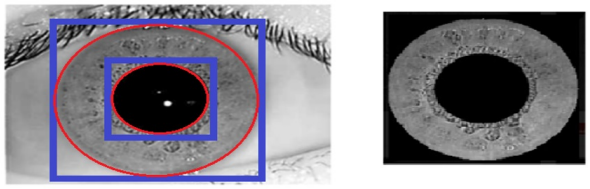


Figure 6.4: sample iris region images from the IITD dataset after applying DA

In addition, to demonstrate why our system used the iris images without segmenting and normalizing them, we conducted two additional experiments.

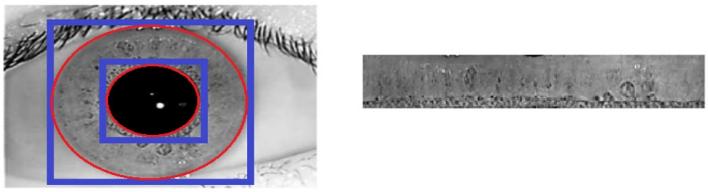
In the first experiment, we segmented the iris region based on the fine-tuned Yolov4-tiny, in which we detected the iris's external and internal boundaries. Then we replaced the pixels found outside the iris boundaries with uniform intensity. Figure 6.5 (b) shows a segmented iris region. We used the obtained segmented iris region to compare the accuracy rate of our proposed CNN model followed by PCA and LinearSVC with some Deep CNN models.

In the second experiment, we normalized the iris region based on the robber sheet model and the iris region boundaries detected by the Yolov4-tiny. Figure 6.6 (b) shows a normalized iris region. Then we used the obtained normalized iris regions to compare the accuracy rate achieved by our CNN followed by PCA and LinearSVC with different Deep CNN models.



(a) (b)

Figure 6.5: Iris region segmentation, (a) iris region boundaries detection, and (b) the segmented iris region image.



(a) (b)

Figure 6.6: Iris region normalization, (a) iris region boundaries detection, and (b) the normalized iris region image.

In the end, we compare the performance of our system with the best accuracies achieved by the segmented and normalized images.

We repeat the same experiments with CASIA-Iris-V1, CASIA-Iris-Inteval, and CASIA-Iris-Thousand datasets. Before we begin, we briefly review the databases used in our experiments.

6.2.2.1 Datasets

We used four benchmark datasets to evaluate the performance of our proposed iris recognition system. Below, we present these iris image databases and show samples from each database.

I. The IITD database: The IITD [2] database was captured at IIT Delhi, New Delhi, India, from the staff and students between Jan and July 2007. The database composes 2240 iris images of 224 persons, their age being around 14-55 years, where each person has ten samples. The database has been collected under non-cooperative conditions in Biometrics Research Laboratory, in which the images contain specular reflections, blur, eyelids, and eyelashes occlusion. The database used JPC1000, digital CMOS, and JIRIS cameras to collect iris images of 320x240 pixels. Figure 6.7 shows some iris images from the IITD database.

[2]"IIT Delhi Iris database", available online: https://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Iris.htm, accessed on 12 March 2022

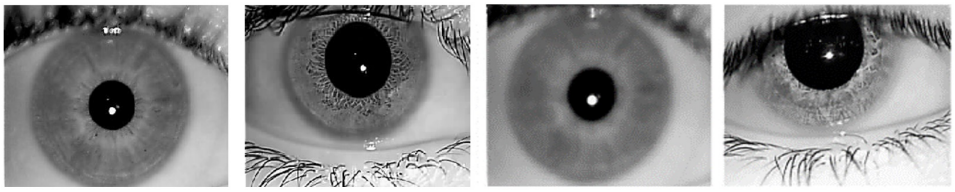


Figure 6.7: Sample iris images of the IITD database

II. The CASIA-Iris-V1 database: The CASIA-Iris-V1 [3] is the first iris dataset created by the Institute of Automation of the Chinese Academy of Sciences (CASIA). This database contains 756 iris images captured from 108 eyes using a custom NIR camera. The database captured seven images from each eye in two sessions; three were collected in the first and four in the second. The database creators replaced the pupil region with a uniform intensity circle to mask the specular reflection. This database's most critical iris recognition challenges are the occlusion by the eyelashes and the eyelids. Figure 6.8 shows some samples from the CASIA-Iris-Interval database.

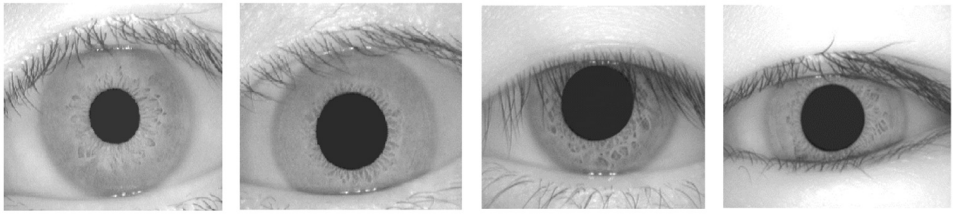


Figure 6.8: Sample iris images of the CASIA-Iris-v1 database

III. The CASIA-Iris-Interval database: The Casia-Iris-Interval [4] database contains 2,655 iris images of 249 persons captured from 395 eyes, where the size of these images being 320 × 280 pixels. The irises are collected under intense illumination to get a rich texture. The challenges presented in this database are pupil constriction/dilation, specular reflection, and eyelashes/eyelids occlusion. The Casia-Iris-Interval database has 395 classes containing different sample numbers, in which most of the systems evaluated on this database chose subsets containing seven samples in each class [7, 144, 153, 154]. Our experiment used a dataset with 153 classes, each containing 7 samples. Figure 6.9 shows some iris images from the CASIA-Iris-Interval database.

[3] "CASIA-Iris-V1 database", available online: http://biometrics.idealtest.org, accessed on 12 March 2022

[4] "The CASIA-Interval database", available online: http://biometrics.idealtest.org/dbDetailForUser.do?id=4 , accessed on 12 March 2022.

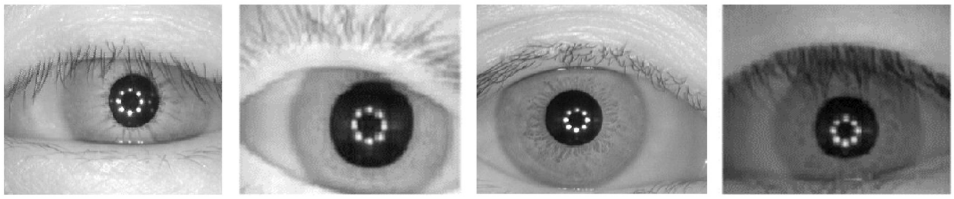


Figure 6.9: Sample iris images of the CASIA-Iris-Interval database

IV. The CASIA-Iris-Thousand database: The CASIA-Iris-Thousand [5] is the first iris database collected from one thousand persons and published with open access. The database contained 20,000 iris images captured from 2000 eyes; each eye has ten samples. The IrisKing manufactured the IKEMB-100 camera that was used to collect this database.

The CASIA-Iris-Thousand was highly used in the literature to evaluate the performance of the new proposed iris recognition systems. This database contains highly challenging iris recognition tasks, such as eyeglasses occlusion, specular reflections, and pupil dilation/ constriction. Figure 6.10 shows some samples from the CASIA-Iris-Thousand database.

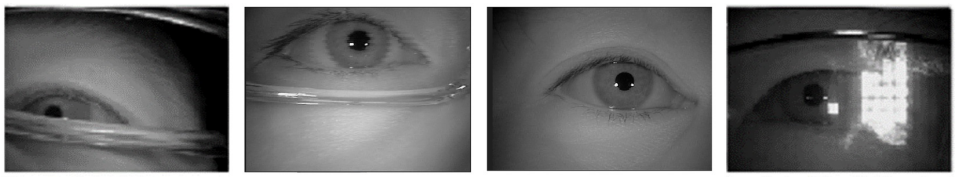


Figure 6.10: Sample iris images of the CASIA-Iris-Thousand database

Table 6.1 presents an overview and a comparison between the four iris databases used in this Chapter.

Dataset	Blur	Reflections	Illumination	Occlusions	Classes' number	Samples in each class	Resolution
CASIA-Iris-Thousand	/	+	++	++	2000	10	640×480
CASIA-Iris-Interval	/	++	+	+	153	7	320×280
CASIA- Iris-V1	/	/	/	+	108	7	384×256
IITD	+	+	/	+	224	10	320×240

Table 6.1: Overview of the four iris databases used in our experiments, in which" +" signifier the presence of challenge,"/" the absence of challenge, and" ++" the challenge degree' is high.

[5]"The CASIA-Interval database", available online: http://biometrics.idealtest.org/dbDetailForUser.do?id=4 , accessed on 12 March 2022.

6.2.2.2 Experimental results on the IITD database

To prove the performance of our proposed system, we conducted many experiments on the IITD iris database that was collected under non-ideal conditions such as blur, reflections, and occlusion.

Table 6.2 illustrates the accuracy rate achieved by our iris recognition system. We reached 99.73% by dividing the dataset into 50% for training and 50 for testing. In addition, by applying a five-fold Cross Validation protocol, we achieved an average accuracy rate of 99.91% with a Standard Deviation (SD) of ± 0.20%. This lower SD indicates that the system is performing consistently well across the different test sets, achieving five accuracies near the average accuracy (99.91%).

Protocol	Accuracy rate
Train: 50%, Test: 50%	99.73%
5-fold CV	99.91% ± 0.20%

Table 6.2: The obtained accuracy of our proposed iris recognition system on the IITD database.

Our system could not recognize only two irises among 2240 iris images. The first image contains a very blurred iris, and the top eyelashes almost entirely occlude the second image. Figure 6.11 shows the two iris images.

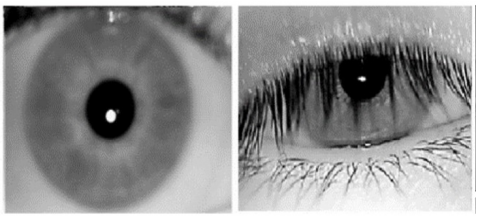


Figure 6.11: The IITD iris images that were unable to recognize by our system

To demonstrate the robustness of our system, we compare our accuracy rate to the results achieved by different Deep CNN models, in which we used the unprocessed iris regions extracted from the IITD images based on Yolov4-tiny.

We applied three scenarios in this comparison. First, we train the deep CNN models from scratch. In the second, we applied Data Augmentation to improve the accuracy rate. In the third, we used pre-trained models to overcome the curse of overfitting. The Xception, Inception-v3, ResNet50, and AlexNet-v2 models trained from scratch achieved good results on the IITD database, surpassing 97%. Still, the accuracy rate achieved by our system outperformed all these models (see Table 6.3).

We applied Data Augmentation to improve the accuracy rate of the different models. Still, the accuracy rate achieved after augmenting the data is lower than the proposed system (See

Table 6.3). The ImageNet pre-trained models achieved a high result that surpassed 99%, but our accuracy rate exceeded all the accuracies obtained by these pre-trained models (See Table 6.3).

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our system
Train from scratch	99.64% ± 0.12%	97.25% ± 1.34%	97.48% ± 1.91%	0.45% ± 0.00%	97.68% ± 1.38%	
Train from scratch with DA	98.08% ± 0.8%	98.93% ± 0.40%	95.09% ± 1.92%	0.45% ± 0.00%	94.42% ± 1.26%	**99.91% ± 0.20%**
Pre-trained models	99.24% ± 0.46%	99.29% ± 0.75%	70.49% ± 8.45%	99.11% ± 0.57%	97.76% ± 1.06%	

Table 6.3: The accuracy rate of our system compared with different Deep CNN models on the IITD database using unprocessed iris region.

We used Yolov4-tiny to detect the iris and the pupil region, then segmented the iris region (see Figure 6.5). The ResNet50 trained from scratch on the segmented iris images achieved an accuracy rate of 98.51%. Still, this accuracy is lower than the accuracy achieved by the proposed CNN followed by PCA and LinearSVC; we reached 99.33% with a very low Standard Deviation ± 0.46% (see Table 6.4). The Data Augmentation enhanced the results of the Xception and inception-v3, but our accuracy rate outperformed the improved results. In addition, the pre-trained models could not surpass our recognition accuracy rate, in which the best model is the Xception, reaching 97.28%. Table 6.4 illustrates that.

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our system
Train from scratch	96.43% ± 2.08%	96.65% ± 1.61%	98.51% ± 1.41%	0.45% ± 0.00%	94.05% ± 1.49%	
Train from scratch with DA	97.93% ± 1.21%	97.50% ± 2.01%	96.70% ± 1.12%	0.45% ± 0.00%	93.80% ± 3.09%	**99.33% ± 0.46%**
Pre-trained models	97.28% ± 1.17%	97.10% ± 1.66%	66.61% ± 2.90%	81.03% ± 8.11%	95.22% ± 2.67%	

Table 6.4: The accuracy rate of our CNN compared with different Deep CNN models on the IITD database based on the segmented iris region.

We used Yolov4-tiny to detect the iris and the pupil region, then normalized the iris region based on Daugman's rubber sheet model (see Figure 6.6). The Xception, Inception-v3, and ResNet50 trained from scratch on the normalized IITD iris database achieved good results, in which the Xception reached 98.26%. The Inception-v3 augmented its accuracy rate by applying Data Augmentation to 98.08%. In addition, the pre-trained Xception and Inception-v3 models achieved an acceptable result that surpassed 95%. However, our CNN model followed by PCA

and LinearSVC achieved an accuracy rate of 99.38%; it surpassed all the models in the three training scenarios, in which the Xception trained from scratch achieved the best average accuracy rate (98.26%). Table 6.5 illustrates that.

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our system
Train from scratch	98.26% ± 0.45%	97.63% ±1.47%	97.99% ±1.20%	0.45% ±0.00%	95.09% ±2.36%	
Train from scratch with DA	97.81% ±0.97%	98.08% ±1.11%	92.95% ±4.83%	0.45% ±0.00%	57.05% ±13.77%	**99.38% ±0.48%**
Pre-trained models	95.53% ±1.13%	95.22% ±3.13%	23.13% ±8.31%	35.31% ±47.76%	89.24% ±3.88%	

Table 6.5: The accuracy rate of our CNN compared with different Deep CNN models on the IITD database based on the normalized iris region.

The previous experiments demonstrated that using an unprocessed iris region is an effective way to obtain high accuracy. The segmentation and the normalization of the iris region affect the quality of the iris region, where the best accuracy rate has been achieved using the unprocessed iris region. Our proposed CNN achieved an accuracy rate of 99.33% using segmented iris region and 99.38% with normalized iris region. The obtained results surpassed the different Deep CNN models. Still, the accuracy rate achieved by our CNN using an unprocessed iris region (our system) surpassed all the results, reaching 99.91%. Figure 6.12 illustrates that.

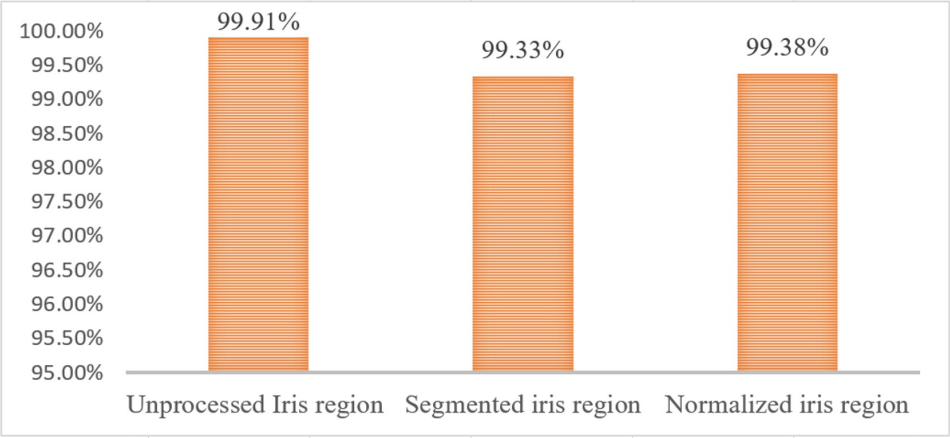


Figure 6.12: The accuracy rate of our proposed CNN+ PCA+ LinearSVC on the IITD database.

6.2.2.3 Experimental results on the CASIA-Iris-v1 database

We conducted many experiments on the CASIA-Iris-V1 database to prove the robustness of our system. We calculated the accuracy rate of our proposed system by applying two different protocols. By applying a split train/test protocol, we achieved 100% when we used 80% for training and 20% for testing. Also, we obtained 99.69% when we used four images for training and three for testing. In addition, Our system obtained a good accuracy rate, reaching 99.60% by applying a five-fold cross-validation protocol. Table 6.6 shows the results achieved by our system.

Protocol	Accuracy rate
Train:80%, Test: 20%	100 %
Train:4/7 , Test: 3/7	99.69 %
5-fold CV	99.60 % ± 0.36 %

Table 6.6: The obtained accuracy of our proposed iris recognition system on the CASIA-Iris-V1 database.

We conducted many experiments on the CASIA-Iris-v1 database using different Deep CNN models to prove the robustness of our proposed system. Various training scenarios have been applied in our experiments; the Xception model was trained from scratch on the unprocessed iris regions and achieved a high accuracy rate reaching 98.73%, which surpassed Inception-v3, ResNet50, VGG16, and AlexNet-v2 models (see Table 6.7). Still, it could not outperform our system's accuracy rate of 99.60% on the CASIA-Iris-V1. We applied Data Augmentation to achieve more accuracy, but the obtained results were lower compared to the models trained from scratch. In addition, Table 6.7 illustrates that the pre-trained models could not surpass the recognition rate achieved by our proposed system.

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our system
Train from scratch	98.73% ±0.14%	95.23% ±2.90%	0.66% ±0.00%	0.66% ±0.00%	0.66% ±0.00%	
Train from scratch with DA	97.09% ±0.11%	84.41% ±5.66%	±0.66% ±0.00%	0.66% ±0.00%	0.66% ±0.00%	**99.60% ±0.36%**
Pre-trained models	96.89% ±0.23%	95.36% ±2.78%	12.96% ±4.04%	87.03% ±4.49%	94.71% ±0.81%	

Table 6.7: The accuracy rate of our system compared with different Deep CNN models on the CASIA-Iris-V1 database using unprocessed iris region.

Moreover, we conducted some experiments on the CASIA-Iris-v1 database after segmenting its iris images. The Xception model trained from scratch achieved the best accuracy rate, reaching 98.16%; the obtained result is lower than our proposed CNN, which achieved 98.54%. The Data Augmentation technique could not improve the accuracy rate of the different deep CNN models,

in which the best accuracy rate obtained is 96.22%. In addition, the pre-trained Deep CNN models achieved acceptable results, but our accuracy rate is always the best. Table 6.8 illustrates the superiority of our proposed CNN on the CASIA-Iris-V1 after segmenting the iris images.

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our system
Train from scratch	98.16% ±0.13%	92.73% ±6.77%	0.66% ±0.00%	0.92% ±0.36%	31.15% ±41.92%	
Train from scratch with DA	96.22% ±0.18%	93.78% ±4.16%	0.79% ±0.30%	0.66% ±0.00%	0.66% ±0.00%	**98.54% ± 0.86%**
Pre-trained models	96.17% ±0.12%	92.36% ±3.46%	40.29% ±3.33%	92.86% ±2.01%	84.53% ±3.35%	

Table 6.8: The accuracy rate of our CNN compared with different Deep CNN models on the CASIA-Iris-V1 database based on the segmented iris region.

The different Deep CNN models trained from scratch on the normalized CASIA-Iris-V1 could not outperform our system's accuracy rate of 98.41%, in which the best accuracy rate has been achieved by the Xception model trained from scratch (98.07%). In addition, the Data Augmentation technique and the pre-trained models could not obtain an accuracy rate that outperforms the previous results. Table 6.9 illustrates our proposed CNN's superiority on the normalized CASIA-Iris-V1 database.

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our system
Train from scratch	98.07% ±0.10%	90.56% ±6.00%	0.66% ±0.00%	0.66% ±0.00%	81.08% ±3.79%	
Train from scratch with DA	95.10% ±0.22%	85.47% ±3.50%	0.79% ±0.30%	0.66% ±0.00%	16.85% ±34.35%	**98.41% ±0.36%**
Pre-trained models	95.98% ±0.20%	86.85% ±2.68%	4.75% ±2.83%	83.73% ±1.23%	84.92% ±3.09%	

Table 6.9: The accuracy rate of our CNN compared with different Deep CNN models on the CASIA-Iris-V1 database based on the normalized iris region.

The previous experiments prove the robustness of our proposed CNN model compared to the different Deep CNN models in the various training scenarios. Moreover, Figure 6.13 illustrates the effectiveness of the features extracted from the unprocessed iris region compared to the segmented and normalized irises, in which the proposed CNN achieved lower results with the segmented and normalized irises compared with the unprocessed iris that reached 99.60%.

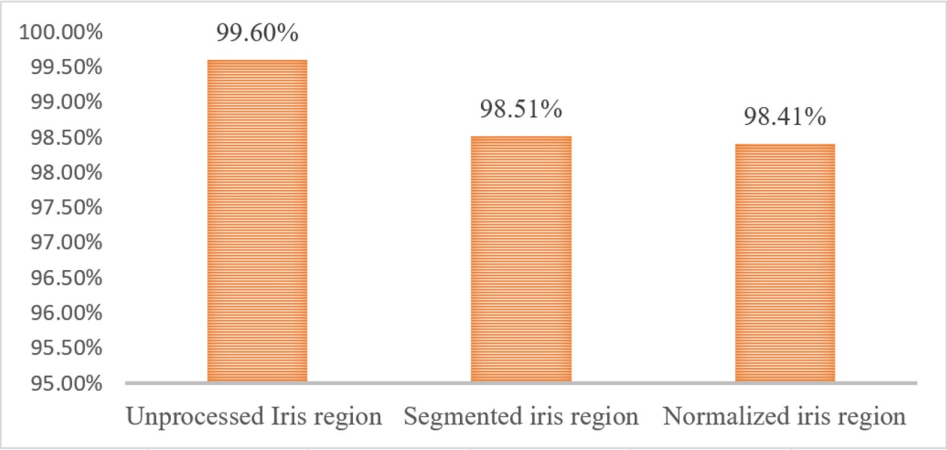


Figure 6.13: The accuracy rate of our proposed CNN+ PCA+ LinearSVC on the CASIA-Iris-V1 database.

6.2.2.4 Experimental results on the CASIA-Iris-Interval database

We evaluate our iris recognition system on the CASIA-Iris-Interval database to prove its robustness against reflection, occlusion, and illumination variation. The proposed iris recognition system achieved a high accuracy rate of 100% by using 6 images for training and 1 for testing. Furthermore, it obtained 100% using 4 images for training and 3 for testing. In addition, our system achieved 99.91% by using a five-fold cross-validation protocol with a Standard Deviation of ± 0.21% (see Table 6.10).

Protocol	Accuracy rate
Train:6/7 , Test: 1/7	100%
Train:4/7 , Test: 3/7	100%
5-fold CV	99.91% ± 0.21%

Table 6.10: The obtained accuracy of our proposed iris recognition system on the CASIA-Iris-Interval database

We conducted several experiments to prove the robustness of our proposed CNN, in which we trained different Deep CNN models in different scenarios and compared their results to ours.

On the CASIA-Iris-Interval, the Xception model trained from scratch based on unprocessed iris regions obtained a high accuracy rate reaching 99.07%. Still, our proposed system surpassed this accuracy rate, achieving 99.91% (± 0.21%). The Data Augmentation technique and the pre-trained models could not get an accuracy rate that exceeded 97.29% and 96.36%, respectively. Table 6.11 shows the performance of our system on the Casia-Iris-Interval compared with different Deep CNN models.

In addition, we conducted other experiments on the segmented iris regions, in which we used

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our system
Train from scratch	99.07% ±0.34	94.77% ±2.95%	0.47% ±0.00%	0.56% ±0.21%	89.29% ±2.05%	
Train from scratch with DA	97.29% ±0.75%	96.12% ±0.99%	0.47% ±0.00%	0.47% ±0.00%	0.47% ±0.00%	**99.91% ±0.21%**
Pre-trained models	96.36% ±1.16%	96.51% ±1.36%	23.00% ±10.27%	94.68% ±0.91%	95.80% ±1.00%	

Table 6.11: The system's accuracy rate compared with different Deep CNN models on the CASIA-Iris-Interval database using unprocessed iris region.

Yolov4-tiny to detect the iris boundaries and then segmented the iris region. Once again, the Xception model trained from scratch achieved the best accuracy rate, reaching 98.76%. Still, our CNN surpassed this result, reaching 99.35% (see Table 6.12). The Data Augmentation technique and the pre-trained models could not be surpassed 98.54% on the segmented CASIA-Iris-Interval database; table 6.12 illustrates that.

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our system
Train from scratch	98.76% ±0.34%	96.25% ±1.11%	0.47% ±0.00%	0.56% ±0.21%	29.25% ±39.44%	
Train from scratch with DA	98.54% ±0.52%	96.32% ±3.46%	0.47% ±0.00%	0.47% ±0.00%	0.56% ±0.21%	**99.35% ±0.26%**
Pre-trained models	94.17% ±3.12%	89.78% ±27.27%	45.53% ±14.23%	64.90% ±10.09%	85.34% ±1.84%	

Table 6.12: The accuracy rate of our CNN compared with different Deep CNN models on the CASIA-Iris-Interval database based on the segmented iris region.

Moreover, we conducted further experiments on the normalized CASIA-Iris-Interval database. In the normalization stage, we applied Daugman's rubber sheet model based on the iris boundaries localized by Yolov4-tiny, where this normalization model is applied in the literature and achieved good results [7, 144]. Our proposed system obtained 99.16% in this scenario. The Xception trained from scratch achieved 97.54%, and the Inception-v3 trained on the normalized images based on Data Augmentation reached 92.40%. Table 6.13 illustrates the superiority of our proposed CNN on the normalized CASIA-Iris-Interval database compared with different Deep CNN models.

In the previous experiments, our proposed CNN outperformed all the CNN models in the different scenarios. Our proposed system that extracts features from the unprocessed iris region based on the proposed CNN achieved an excellent result, reaching 99.91%, with a very low Standard Deviation value of ±0.39%. Figure 6.14 illustrates the effectiveness of the features

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our system
Train from scratch	97.54% ±1.12%	90.73% ±5.26%	0.47% ±0.00%	0.47% ±0.00%	85.81% ±3.91%	
Train from scratch with DA	96.08% ±1.42%	92.40% ±3.27%	0.47% ±0.00%	0.47% ±0.00%	0.47% ±0.00%	**99.16% ±0.39%**
Pre-trained models	92.49% ±3.44%	85.70% ±2.56%	1.70% ±0.58%	62.00% ±2.24%	78.34% ±2.21%	

Table 6.13: The accuracy rate of our CNN compared with different Deep CNN models on the CASIA-Iris-Interval database based on the normalized iris region.

extracted from the unprocessed iris region compared to the segmented and normalized irises.

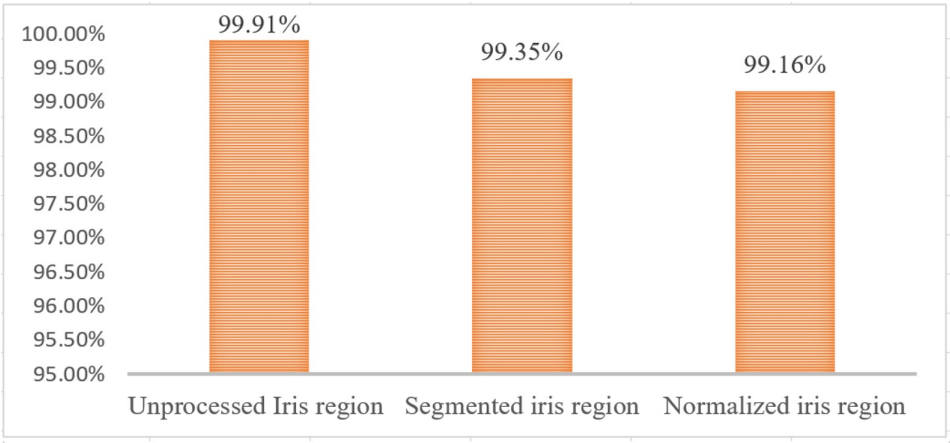


Figure 6.14: The accuracy rate of our proposed CNN+ PCA+ LinearSVC on the CASIA-Iris-Interval database.

6.2.2.5 Experimental results on the CASIA-Iris-Thousand database

The CASIA-Iris-Thousand database is an iris images database containing 20.000 images, captured under non-cooperative conditions. The database has lighting reflections, illumination variations, and a high degree of occlusion. Our system proves its effectiveness in this database, reaching 99.05% by splitting the data into 70% for training and 30% for testing. Furthermore, it obtained 99.19% by applying a five-fold cross-validation protocol, with a very low Standard Deviation (± 0.14%), which indicates that the five accuracies achieved by the cross-validation process are very near to the average accuracy (99.19%), Table 6.14 illustrates that.

The Xception and Inception-v3 trained from scratch on the segmented CASIA-Iris-Thousand database achieved an acceptable result: the first model reached 96.12% and the second 95.75%. Still, these results are lower than those obtained by our proposed system (99.19% ± 0.14%).

Protocol	Accuracy rate
Train:70%, Test: 30%	99.05 %
5-fold CV	99.19% ± 0.14%

Table 6.14: The obtained accuracy of our proposed iris recognition system on the CASIA-Iris-Thousand database.

The Data Augmentation technique improves the accuracy rate of the previous models. But, our accuracy rate surpassed the enhanced results (see Table 6.15). Also, the pre-trained models could not achieve a good result; the best accuracy rate was 88.73%. Table 6.15 illustrates the superiority of our recognition accuracy compared to the results of the different models in the three scenarios.

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our system
Train from scratch	96.12% ±0.49%	95.75% ±0.31%	90.74% ±2.38%	0.05% ±0.00%	0.05% ±0.00%	
Train from scratch with DA	96.72% ±0.35%	96.59% ±0.96%	94.48% ±0.79%	0.05% ±0.00%	0.05% ±0.00%	**99.19% ±0.14%**
Pre-trained models	88.73% ±1.45%	86.88% ±0.86%	1.23% ±1.02%	0.50% ±0.31%	62.68% ±2.87%	

Table 6.15: The accuracy rate of our system compared with different Deep CNN models on the CASIA-Iris-Thousand database based on the unprocessed iris region.

In this experiment, we applied Yolov4-tiny to detect the iris boundaries and extracted the iris region. The Xception and Inception-v3 models trained from scratch on the segmented CASIA-Iris-Thousand achieved a good result that surpassed 94%. Still, the obtained result is lower than our proposed CNN model followed by PCA and LinearSVC, which achieved 97.50%. The Data Augmentation improves the previous accuracies of the Xception, Inception-v3, and ResNet50, but our accuracy is higher than the obtained results. In addition, the ImagesNet pre-trained models were unable to outperform our accuracy. Table 6.16 illustrates the performance of our CNN on the segmented CASIA-Iris-Thousand database.

The Yolov4-tiny was used to detect the iris boundaries, and the Daugman rubber sheet model was applied to normalize the iris region. The Xception and the Inception-v3 models were trained from scratch on the normalized CASIA-Iris-Thousand, reaching 92.87% and 91.14% of accuracy, respectively. We applied Data Augmentation to enhance the obtained results. Still, the improved results are lower than our CNN (97.07% ±0.31%). In addition, the pre-trained model could not obtain an accuracy rate that competes with our results. Table 6.17 shows the results obtained on the normalized CASIA-Iris-Thousand database.

The previous experiments prove the robustness of our proposed CNN compared with the

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our system
Train from scratch	95.08% ±1.12%	94.48% ±1.73%	85.85% ±6.72%	0.05% ±0.00%	0.05% ±0.00%	
Train from scratch with DA	95.23% ±1.04%	95.01% ±0.88%	93.32% ±3.00%	0.05% ±0.00%	0.05% ±0.00%	**97.50% ±0.28%**
Pre-trained models	79.12% ±2.53%	74.02% ±3.16%	3.73% ±1.23%	0.30% ±0.18%	0.46% ±0.07%	

Table 6.16: The accuracy rate of our CNN compared with different Deep CNN models on the CASIA-Iris-Thousand database based on the segmented iris region.

	Xception	Inception-v3	ResNet50	VGG16	AlexNet-v2	Our system
Train from scratch	92.87% ±2.35%	91.14% ±3.09%	88.48% ±2.34%	0.05% ±0.00%	0.05% ±0.00%	
Train from scratch with DA	94.27% ±1.12%	93.40% ±0.77%	90.90% ±3.56%	0.05% ±0.00%	0.05% ±0.00%	**97.07% ±0.31%**
Pre-trained models	73.63% ±1.93%	71.77% ±2.78%	1.29% ±0.79%	0.19% ±0.20%	34.22% ±3.08%	

Table 6.17: The accuracy rate of our CNN compared with different Deep CNN models on the CASIA-Iris-Thousand database based on the normalized iris region.

different Deep CNN models. Figure 6.15 illustrates the superiority of the accuracy rate achieved by our proposed CNN based on the unprocessed iris region compared to the segmented and normalized irises, in which the unprocessed iris region achieved an accuracy rate of 99.19%. But, the normalized and the segmented irises could not surpass 97.50%.

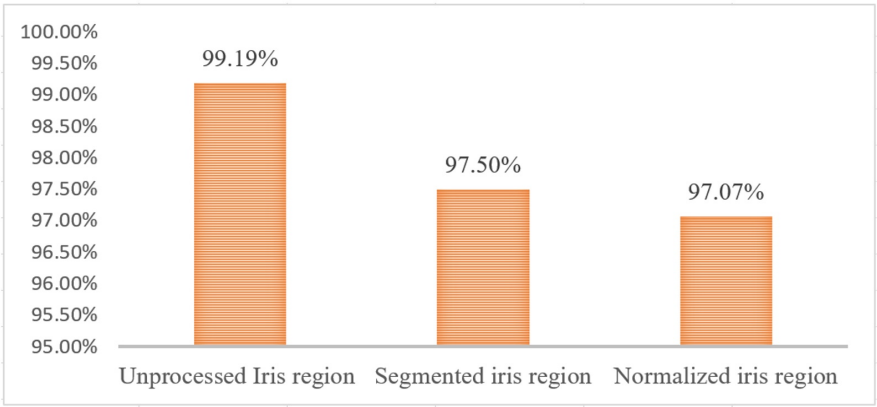


Figure 6.15: The accuracy rate of our proposed CNN+ PCA+ LinearSVC on the CASIA-Iris-Thousand database.

Table 6.18 shows the accuracy rate of our proposed system on the four databases used to evaluate its robustness. The obtained results prove the excellent performance of our system.

Datasets	Classes' number	Evaluation protocol	Accuracy	Standard Deviation
IITD	224	(Train: 50%, Test: 50%)	99.73%	/
	224	5-fold CV	99.91%	±0.20%
CASIA-Iris	108	(Train: 57%, Test: 43%)	100%	/
V1	108	5-fold CV	99.60%	±0.36%
CASIA-Iris	153	(Train: 86%, Test: 14%)	100%	/
Interval	153	(Train: 57%, Test: 43%)	100%	/
	153	5-fold CV	99.91%	± 0.21%
CASIA-Iris	2000	(Train: 70%,Test: 30%)	99.05%	/
Thousand	2000	5-fold CV	99.19%	± 0.14%

Table 6.18: The accuracy rate of our proposed iris recognition system on the IITD, CASIA-Iris-V1, CASIA-Iris-Interval, and CASIA-Iris-Thousand datasets.

6.2.3 Comparison with state-of-the-art methods

To demonstrate the effectiveness and robustness of our proposed system, we compare our results to state-of-the-art methods evaluated on the same databases used in our experiments.

On the IITD database, we obtained a high accuracy rate that surpassed all the methods mentioned in Table 6.19 , except Alaslani [153] and Alaslni et al. [154], which trained and tested on a small subset containing only 60 classes. On the other side, our system has been evaluated on the entire database (224 classes).

Datasets	State-of-the-art	No. classes	Eval protocol (Train, Test, Val)	Normal-ization	Accuracy
IITD	Winston and Hemanth [145]	/	(60%, 40%, -)	Yes	98.40%
	Alaslani [153]	**60**	**(80%, 20%, -)**	**No**	**100%**
	Alaslni et al [154]	**60**	**(80%, 20%, -)**	**No**	**100%**
	Arora and Bhatia [155]	224	(60%, 20%, 20%)	Yes	98%
	Yifeng Chen et al. [5]	224	(60%, 10%, 10%)	Yes	99.30%
	Sujana and Reddy [6]	224	(80%, 20%, -)	Yes	98.05%
	Shanto et al. [160]	25	(60%, 20%, 20%)	No	98%
	Our System	**224**	**(50%, 50%, -)**	**No**	**99.73%**
		224	**5-fold CV**	**No**	**99.91% ±0.20%**
CASIA Iris V1	Dua et al. [143]	108	/	Yes	97%
	Abdo et al. [147]	100	(57%, 43%, -)	Yes	98.15
	Alaslani [153]	60	(80%, 20%, -)	No	98%
	Alaslni et al. [154]	60	(80%, 20%, -)	No	98.30%
	Sujana and Reddy [6]	108	(80%, 20%, -)	Yes	95.40%
	Kranthi Kumar et al. [157]	108	(80%, 20%, -)	Yes	98%
	Hassan et al. [161]	108	(70%, 15%, 15%)	Yes	99.07
	Our System	**108**	**(57%, 43%, -)**	**No**	**100%**
		108	**5-fold CV**	**No**	**99.60% ±0.36%**
CASIA Iris Interval	Khotimah and Juniati [142]	10	5-fold CV	Yes	92.63%
	Abdo et al. [7]	100	(86%, 14%, -)	Yes	100%
		100	(71%, 29%, -)	Yes	98.50%
		100	(57%, 43%, -)	Yes	96.67%
	Abdalla et al. [144]	100	(86%, 14%, -)	Yes	100%
		100	(71%, 29%, -)	Yes	99%
		100	(57%, 43%, -)	Yes	97%
	Abdo et al. [147]	100	(57%, 43%, -)	Yes	99%
	Alaslani [153]	60	(80%, 20%, -)	No	89%
	Alaslni et al. [154]	60	(80%, 20%, -)	No	91.60%
	Our System	**153**	**(86%, 14%, -)**	**No**	**100%**
		153	**(57%, 43%, -)**	**No**	**100%**
		153	**5-fold CV**	**No**	**99.91% ±0.21%**
CASIA Iris Thousand	Alaslani [153]	60	(80%, 20%, -)	No	98%
	Alaslni et al. [154]	60	(80%, 20%, -)	No	95%
	Chakraborty et al.[156]	2000	(80%, 20%, -)	No	94.70%
	Yifeng Chen et al. [5]	2000	(75%, 25%, -)	No	99,14%
	Jayanthi et al. [159]	2000	(80%, 10%, 10%)	Yes	98.75%
	Our System	**2000**	(70%, 30%, -)	**No**	**99.05%**
		2000	**5-fold CV**	**No**	**99.19% ±0.14%**

Table 6.19: Our iris recognition rate compared to state-of-the-art approaches on the four databases, Where "Yes" or "No" was used to indicate the existence of the iris normalization process or not.

On the Casia-Iris-v1, our system reached 99.60% by applying five-fold cross-validation. In addition, it achieved an accuracy rate of 100% by using four images for each person for training and three images for testing. The obtained results surpassed state-of-the-art methods evaluated on the Casia-Iris-v1 database; Table 6.19 illustrates that.

Moreover, our system obtained a high recognition accuracy rate on the CASIA-Iris-Interval, in which we used 153 classes with seven samples. The proposed system achieved 99.91% by applying five-fold cross-validation and 100% by splitting the data into 57% for training and 43% for testing. Table 6.19 illustrates the superiority of our system over state-of-the-art methods, even though these methods use small subsets.

Furthermore, our proposed system achieved an accuracy rate that surpassed state-of-the-art methods on the CASIA-Iris-Thousand database, in which we obtained 99.05% by splitting the data into 70% for training and the rest for testing and 99.19% by applying five-fold cross-validation protocol. Yifeng Chen et al. [5] achieved excellent results in this database, reaching 99.14%. But, this method extracts features from the normalized iris region, which augment the system's complexity. In addition, it could not obtain a high result on a Small database like the IITD database (see Table 6.19).

Our recognition accuracy rate presented in Table 6.19 illustrates the robustness of our system against the uncontrolled conditions included in each database, such as illumination reflections, lighting variation, and occlusion.

6.3 Conclusion

In this Chapter, we proposed a robust iris recognition system based on the pre-trained Incaption-v3 model. The proposed system fine-tuned the Yolov4-tiny model and applied it to detect the iris region. Then, the system extracts features using a new Deep CNN model inspired by the ImageNet pre-trained inception-v3 model. In addition, the system applied Global Average Pooling (GAP) to reduce the dimensionality and PCA to keep relevant features. For the classification stage, our system used LinearSVC.

The proposed system used only five pre-trained inception modules without any fine-tuning; it also extracted features from the iris region without normalizing and segmenting them. All these characteristics help the system save computational resources and time.

The proposed system achieved high results on four benchmark iris datasets captured under non-cooperative conditions, such as lighting variation, specular reflection, and occlusions. By applying a five-fold cross-validation protocol, the proposed system achieved a new state-of-the-art accuracy rate reaching 99.91%, 99.60%, 99.91% and 99.19% on the IITD, CASIA-Iris-v1, CASIA-Iris-Interval, and CASIA-Iris-Thousand, respectively.

CHAPTER 7

FACE-IRIS MULTIMODAL RECOGNITION SYSTEM

"Multimodal biometric recognition is a promising personal identity authentication technology which can remedy the limitation of the traditional identity authentication and the unimodal biometrics."

Wang, Zhifang, et al. [224]

7.1 Introduction

In everyday situations, many applications identify users prior to providing requested services in order to ensure their privacy. Several applications have recently successfully used biometric technology to identify and authenticate users. However, due to the COVID-19 pandemic and the possibility of disease transmission through touch, people have become wary of using touch-based biometric systems. Wherefore, most people have become tend to use biometric systems that do not require physical contact with sensors, such as facial and iris recognition.

The face is widely used as a biometric trait for recognizing persons [44]. It has been employed in many applications such as mobile device authentication, border control, video surveillance, healthcare, law enforcement, and other applications where user identification is required.

Two main reasons for the extensive use of faces in biometrics recognition: the first one, face recognition is a non-cooperative user application where the system can be captured the face user's without any specific actions or poses. The second one, people's faces are publicly visible daily; collecting face images does not affect users' privacy. Despite the great success achieved by the

face biometric trait, many obstacles hindered the performance of the face recognition system, like facial expression, head poses, occlusion, illumination variation, and low light conditions [23].

In addition, iris recognition is one of the most unique and reliable biometric modalities used these days for personal identification. It has been used in many applications, such as mobile device unlocking, attendance tracking systems, and passport verification. The iris patterns are stable over time, can be captured without touching the sensor, and can use even the differences between twins [4]. Still, the performance of the iris recognition system is affected by certain eye diseases and by the images captured under non-cooperative conditions such as illumination variation, blur, and occlusion [225].

Therefore, relying solely on a single biometric trait such as the face or iris is inadequate for high-security requirements in military and government applications. Lately, multimodal systems have become a popular solution to address the limitations and problems affecting unimodal systems' performance and reliability, such as inter-class similarity, intra-class variation, non-universality issues, and vulnerability to spoofing [9, 10]. The effective use of the face and the iris modalities in person recognition indicates that the fusion of the two modalities may prove promising [66].

Many papers published in the literature proposed multimodal biometric systems based on face and iris modalities. However, most of these works used the pre-processing stage before extracting the features, which is a delicate and complex task. Moreover, several face-iris multimodal systems have been evaluated on chimeric databases that combine two unimodal datasets [66]. These systems do not use real databases to evaluate the proposed system's performance. Furthermore, many proposed systems used handcraft methods to extract features from face images. Still, these systems' performance degraded under unconstrained conditions because the handcraft features are sensitive to environmental conditions, such as noise and illumination variations [226]. Also, some systems applied Deep CNN models containing tens to hundreds of millions of parameters, which consume time and computational resources [169]. Moreover, if one of the face or iris traits is missed or unavailable, the accuracy rate of the many proposed systems is dramatically degraded.

This Chapter presents the contributions of this PhD thesis in the multimodal biometrics recognition field.

We proposed three face-iris recognition systems; the first applied image-level fusion, the second used feature-level fusion, and the third was based on score-level fusion. The proposed systems used Yolov4-tiny to detect the face and the iris regions. The proposed system extracts features from the localized areas without any pre-processing, such as face image enhancement, iris segmentation or normalization. Therefore, the proposed systems save computational resources and time. Our systems extract features based on a new deep CNN built by only eight modules from the ImageNet pre-trained Xception. The proposed model uses the off-the-shelf weights without fine-tuning them on the new databases. Therefore, the proposed CNN model saves training time and computational resources, also, does not require massive data. In addition, they

used the LinearSVC in the classification stage and the PCA to keep relevant iris features. The results of these proposed systems have been published in [227].

In this Chapter, we present the main steps of our proposed iris-face recognition systems, followed by detailed architecture of each proposed system. Then, we conducted multiple experiments on many benchmark datasets to prove the robustness of our propositions. We evaluated the performance of our systems on unimodal databases; to demonstrate our systems' effectiveness even if one of the biometric traits is missed or unavailable. In addition, we used a real and a chimeric multimodal database in our evaluation. Moreover, this Chapter illustrated the superiority of our results by comparing them to state-of-the-art methods.

7.2 Contribution V: Effective Face-Iris multimodal biometric systems based on the Xception model

This Phd thesis proposed three face-iris recognition systems:

I. The first one fused the detected face with the left and the right iris regions, then used our proposed CNN to extract features and PCA to reduce the dimensionality. The linearSVC is used in the classification step. To our knowledge, this is the first proposed system that used image-level fusion to fuse the face and both iris traits, then applied deep CNN to extract features from the fused image.

II. The second one detects the face and both iris regions and extracts deep features from the three detected areas. Then the proposed system concatenates the features and applies PCA to keep relevant components. In the classification stage, LinearSVC is used.

III. After detecting the regions of interest and extracting features, the third system applied PCA to reduce the dimensionality and LinearSVC decision function to calculate the confidence scores. Then, it fused the obtained scores to identify the user.

In this section, we present the detail of the proposed face-iris systems. Then, we evaluate their performance on different benchmark datasets. At the end of this section, we compare our accuracy rate with state-of-the-art methods to prove the effectiveness of our proposition.

7.2.1 The proposed Face-Iris multimodal biometric systems

In this Chapter, we proposed three face-iris recognition systems based on three main steps. This section presents these three stages and then shows the detailed architecture of each system.

7.2.1.1 The proposed systems' main steps

The proposed systems share three primary stages, which we will illustrate in the following subsections:

I. Region of interest detection: The Region of Interest (ROI) detection step uses Yolov4-tiny [21], a faster version of the popular Yolov4 object detection model. Yolov4-tiny has fewer parameters and a simpler structure than the original Yolov4 [192], resulting in lower memory/computing requirements and faster detection speed (average of <3 ms per frame using 1080Ti GPU) [196]. Our system uses a pre-trained MS COCO YOLOv4-tiny model to detect faces and iris regions in images, fine-tuned with only 400 images (70% train, 20% test, 10% validation). Figure 7.1 shows examples of localized ROI.

Figure 7.1: Region of interest localization based on Yolov4-tiny.

II. Features extraction: Feature extraction is the most critical stage in pattern recognition. Deep CNNs have seen great success in this area in the last decade [213]. The Xception model [110] is a deep CNN model that replaced inception modules of the Inception-v3 [187] with depthwise separable convolution layers, improving the recognition accuracy rate. The result achieved by the Xception model surpassed AlexNet [107], VGG16 [17], RESNET [109], Inception-v3 [187] and other prominent deep CNNs in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC).

The main disadvantage of the Xception model is its requirement for huge datasets to train its parameters [110]. The ImageNet dataset [198] overcame this issue, in which papers frequently used the ImageNet pre-trained Xception model to classify images in various tasks [213, 228].

Our study inspired a deep CNN for biometric recognition by the pre-trained Xception model. The Xception network consists of 36 convolutional layers arranged in 14 residual blocks and has over 22 million parameters. To build a reduced face-iris biometric feature extraction model based on the pre-trained Xception, we conducted experiments using the SDUMLA-HTM [229] face and SDUMLA-HTM [229] iris datasets.

The pre-trained Xception architecture showed lower accuracy (max 97% on SDUMLA-HTM iris, 98% on SDUMLA-HTM face) when fewer than eight residual blocks were used. However, accuracy was high when eight residual blocks were used (over 99% on both datasets). But it decreased or remained the same with more than eight residual blocks. In our experiment, we followed the output of the last residual block by a Global Average Pooling (GAP) to reduce the dimensionality, PCA to keep relevant features, and SVM to calculate the accuracy rate.

For this reason, we introduced a new feature extraction model based on the Xception pre-trained model, which consists of 20 convolutional layers divided into eight residual blocks, building by only around 7 million parameters. Figure 7.2 shows our proposed model below and the original Xception model above.

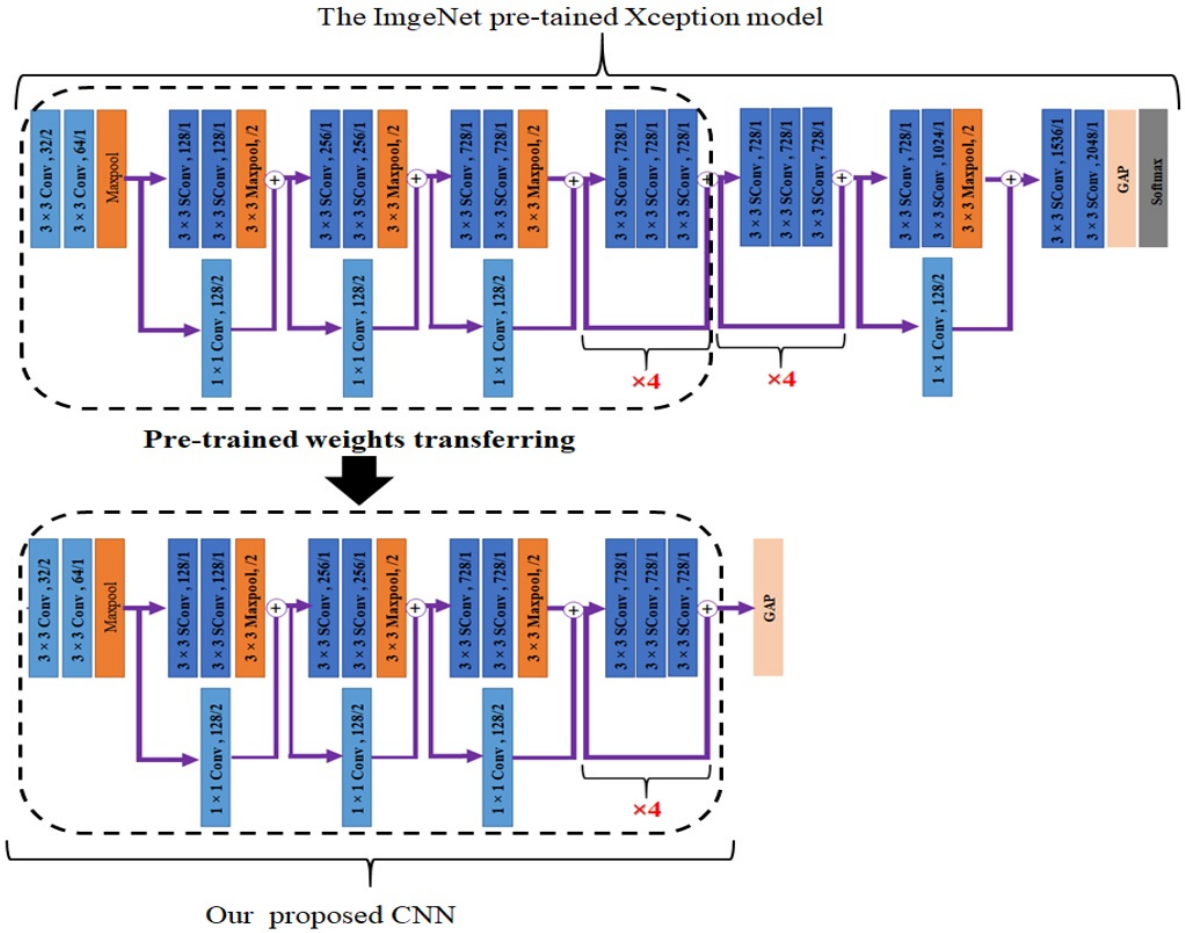


Figure 7.2: The architecture of our proposed CNN.

III. Classification: Generally, the last two layers in a Deep CNN model are a Fully Connected (FC) layer followed by a softmax layer. However, overfitting can impact the softmax classifier's accuracy due to the lack of training data and the model's complexity [230]. To combat this curse, we proposed utilizing an efficient and popular machine-learning technique called Support Vector Machine (SVM) classifier.

In our study, we used a fast SVM algorithm called LinearSVC that applied a One-Versus-All approach; this approach achieved high success in many biometric systems [138, 144, 153]. We applied PCA before training the LinearSVC to decrease the complexity, obtain a higher accuracy rate, and save computational time and resources.

7.2.1.2 The proposed face-iris recognition system based on image-level fusion

The first face-iris proposed system is based on image-level fusion; it applied Yolo-tiny to localized face, left iris, and right iris regions, then converted the detected regions into grayscale images, each containing 299×299 pixels. The system concatenates the three images to generate a new image with 3×299×299 pixels, which is the default Xception model input size. In addition, the system applied our proposed Deep CNN model to extract deep features and used PCA for dimensionality reduction. Finally, the system applied LinearSVC to determine the user's identity. To the author's knowledge, this proposed system is the first multimodal biometric system that fused images and used Deep CNN to extract features from the fused image. Figure 7.3 shows the detailed architecture of our multimodal image-level fusion system.

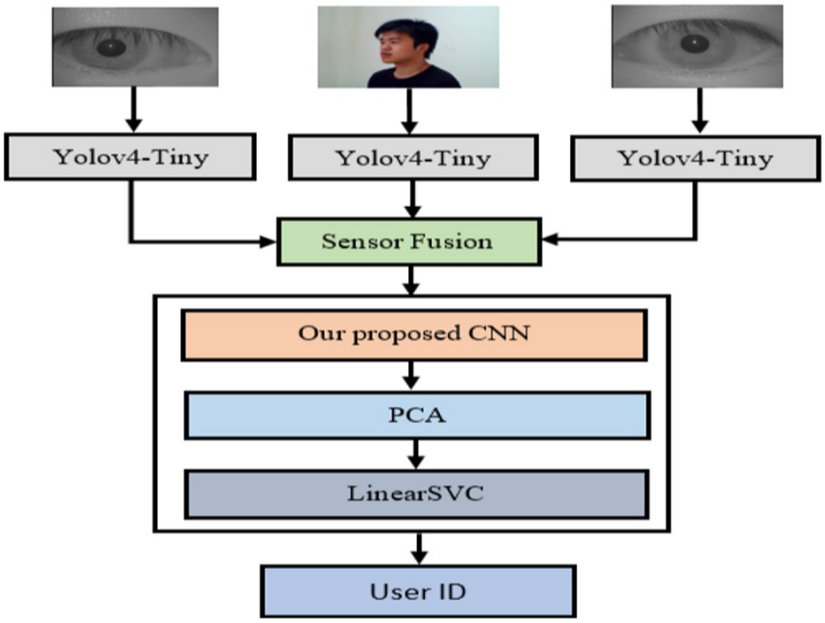


Figure 7.3: The diagram of our proposed face-iris recognition system based on image-level fusion.

7.2.1.3 The proposed face-iris recognition system based on feature-level fusion

We used Yolov4-tiny to localize the ROIs from the images, then we used the proposed CNN to extract features from the detected face and both irises regions. We build a new feature vector (X) by concatenating the three feature vectors extracted from the localized regions. The feature vector (X) is created based on the following equation.

$$X = x_l \left| x_f \right| x_r \tag{7.1}$$

Where x_l represents the feature vector obtained from the left iris region, x_f is acquired from the face image, and x_r is obtained from the right iris region.

After building the vector (X) containing the fused features, we determined relevant features by applying PCA, and then we used the LinearSVC to classify user identity. Figure 7.4 shows the detailed architecture of our multimodal system that applied the fusion at the feature level.

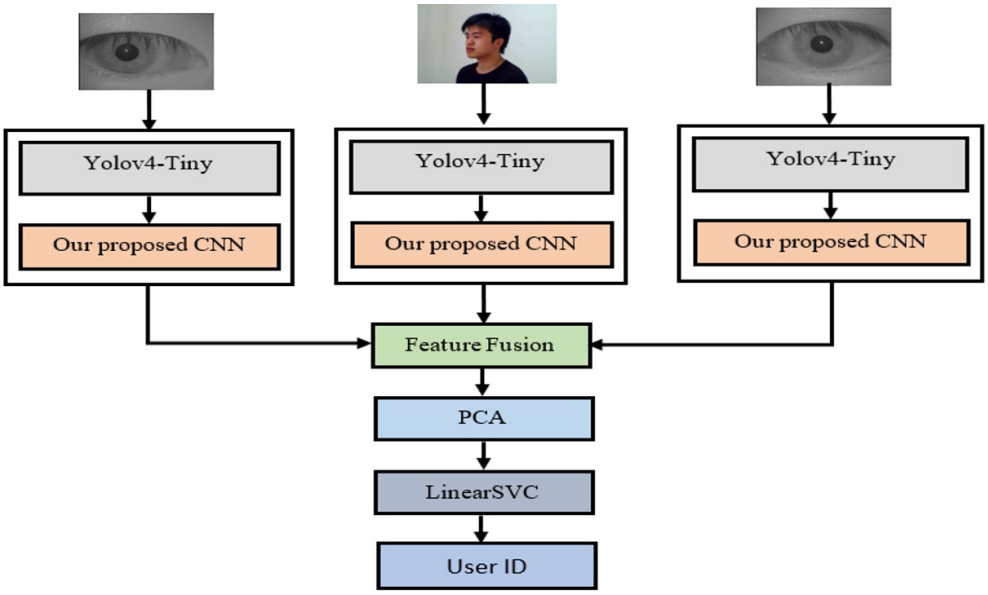


Figure 7.4: The diagram of our proposed face-iris recognition system based on feature-level fusion.

7.2.1.4 The proposed face-iris recognition system based on score-level fusion

This proposed system localized the ROIs by applying the Yolov4-tiny and extracted features from the detected regions based on the proposed Deep CNN model. In addition, it used PCA to select relevant features. The system applied the LinearSVC decision function to compute the confidence scores. The system normalized the obtained scores before combining them based on an effective normalization technique called tanh-normalization [73]. The following equation illustrates the formula of the score normalization technique used by our system:

$$S_t = \frac{1}{2}\left(\tanh\left(0.01\left(\frac{s_t - \mu_{GH}}{\sigma_{GH}}\right)\right) + 1\right) \tag{7.2}$$

Where s_t represents the score vector of the trait t computed by the LinearSVC library, σ_{GH} is the standard deviation, and μ_{GH} is the mean of the score vector (s_t). S_t represents the normalized score vector of the trait t.

Finally, our proposed system computed the final score by fusing the normalized scores based on the sum or the product rule fusion techniques. Then it applied the argmax function to pick the maximum score for defining the user's identity.

The sum rule computed the final score by fusing the normalized scores based on the following equation:

$$S_s = \sum_{t=1}^{N} S_t \tag{7.3}$$

In which S_t represents the normalized score vector of the trait t, and N is the number of traits ($N = 3$ in our case).

In addition, the product rule computed the final score by fusing the normalized scores based on the following equation:

$$S_p = \prod_{t=1}^{N} S_t \tag{7.4}$$

In which S_t represents the normalized score vector of the trait t, and N is the number of traits ($N = 3$ in our case).

Figure 7.5 shows the detailed architecture of our proposed multimodal system that applied the fusion at the score level.

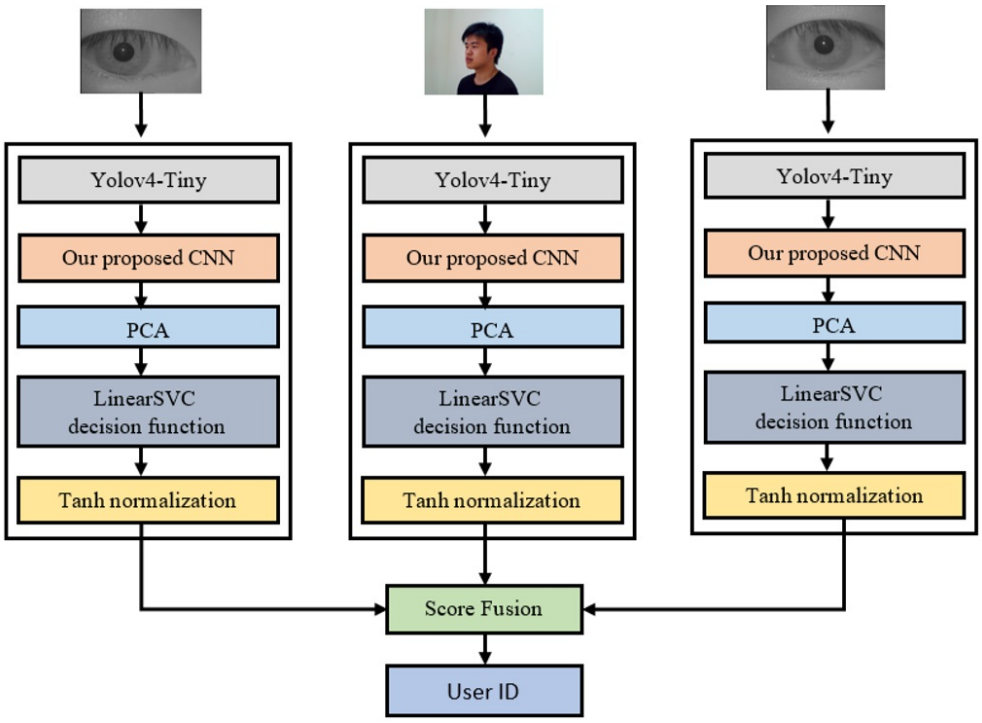


Figure 7.5: The diagram of our proposed face-iris recognition system based on score-level fusion.

7.2.2 Experiments

To prove the robustness of our proposed systems, we conducted many experiments. We demonstrate the effectiveness of our proposal if one of the biometric traits is missed or unavailable,

where we conducted many experiments on eight face and iris unimodal datasets captured under unconstrained conditions. We used four face databases: ORL, FERET, Georgia Tech, and SDUMLA-HTM face databases. In addition, we used four iris databases: IITD, UBIRIS-v1, CASIA-Iris-Interval, and SDUMLA-HTM iris databases.

Moreover, we demonstrate the performance of our proposed systems by conducting experiments on two face-iris multimodal databases captured under an uncontrolled environment. The first is a real database called SDUMLA-HTM containing face and iris images captured from the same persons. The second is a chimeric database called CASIA-ORL multimodal database built by combining unimodal iris and face datasets. The next sub-section introduced the databases used in our experiments.

7.2.2.1 Datasets

Our experiments used multiple datasets: three face databases, three iris databases, and two face-iris multimodal databases.

I. The Georgia Tech face database: The Center for Signal and Image Processing of the Georgia Institute of Technology collected the Georgia Tech database in 1999. The database has been collected under non-cooperative conditions, such as facial expression, scale, and illumination variation. A cluttered background has been used to capture this database's images. This database comprises 750 images collected from 50 persons, each with 15 face images. Some Georgia Tech face database samples are shown in Figure 7.6.

Figure 7.6: Samples from the Georgia Tech database.

II. The FERET database: Is a face dataset collected under an unconstrained environment; in our experiments, we used 1400 face images captured from 200 people, each has seven face images. More detail about this database is presented in Chapter 5.

III. The ORL database: Is a benchmark face database captured under non-controlled conditions from 40 persons, in which each person have ten different images. More detail and simples from this database were presented in Chapter 5.

IV. The UBIRIS-v1 database [231]: Comprises 1877 iris images of 241 persons, captured in two separate sessions in September 2004. The first session collected the iris images in a controlled environment to minimize noise factors such as brightness, contrast, and reflections. During the second session, images were captured under non-cooperative conditions, affecting the iris images by illumination reflections, focus, and contrast. The UBIRIS-v1 database is commonly used to evaluate the performance of iris recognition systems due to its images containing various noise factors. Figure 7.7 showcases samples from the UBIRIS-v1 database.

Figure 7.7: Samples from the UBIRIS-v1 database.

V. Casia-Iris-Interval database: Is an iris database containing 2939 images captured under multiple challenges from 249 persons. Our experiments explore 1071 iris images collected from 153 eyes; each has seven images. Samples and more detail about the Casia-Iris-Interval database were presented in the previous Chapter (See 6.2.2.1).

VI. The IITD database: Is an iris images database captured from 224 students and staff under non-cooperative conditions at the IIT Delhi. The database contains 2240 images, where each person has ten iris images. Samples and more detail about the IITD database were presented in the previous Chapter (See 6.2.2.1).

VII. CASIA-ORL Database: is a multimodal database comprising 40 subjects, each has seven face images selected randomly from the ORL database. Also, this chimeric database contains 14 iris images for each subject randomly chosen from the Casia-Iris-Interval database, seven left and seven right iris images, which these iris images collected from the same person.

VIII. The SDUMLA-HMT multimodal database [229]: during the summer of 2010 at Shandong University, Jinan, China, the SDUMLA-HMT multimodal database was collected. The database comprises 8,904 face images captured from 106 persons; each has 84 face images. These images are collected from seven different angles and with variations in facial expressions (closed eyes, smile, frown, and surprise), poses (look forward, downward, and upward), accessories (hat and glasses), and lighting conditions. The face image size of this database is 640x480 pixels. Samples from the SDUMLA-HMT face dataset are shown in Figure 7.8.

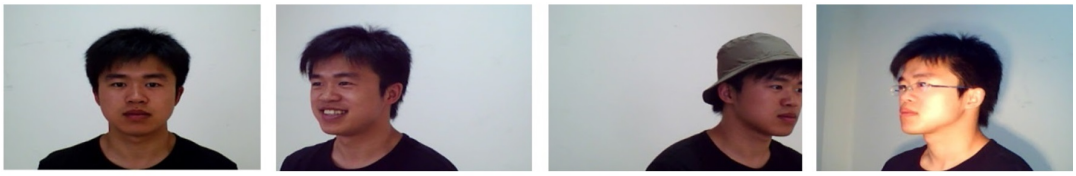


Figure 7.8: Face Samples from the SDUMLA-HTM database.

The SDUMLA-HMT database contains 1060 images captured with an intelligent iris sensor under near-infrared illumination. The iris images were collected from 106 persons; each had ten images, five captured from the left and five from the right eye. Samples from the SDUMLA-HMT iris dataset are shown in Figure 7.9.

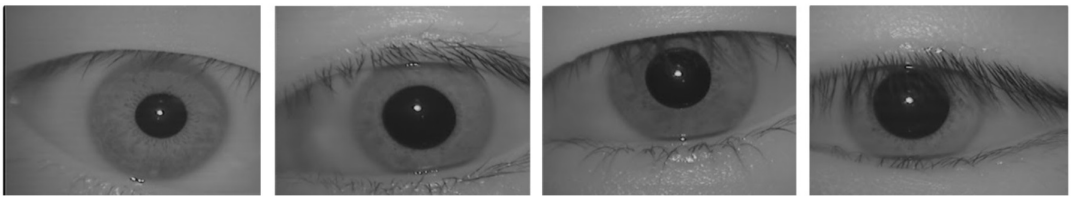


Figure 7.9: Iris Samples from the SDUMLA-HTM database.

All the datasets used to evaluate our proposed system have been summarized in Table 7.1.

Dataset		Number of samples in each class	Number of classes
The Georgia Tech face dataset		15	50
The FERET face dataset		7	200
The ORL face dataset		10	40
The UBIRIS-v1 iris dataset		least 5	241
The Casia-Iris-Interval dataset		7	153
The IITD iris dataset		10	224
The CASIA-ORL multimodal dataset	face Right iris Left iris	7 7 7	40 40 40
The SDUMLA-HMT multimodal dataset	face Right iris Left iris	84 5 5	106 106 106

Table 7.1: Summary of datasets used in our experiments

7.2.2.2 Experiments on face datasets

The biometric multimodal system accuracy is directly affected by the unimodal systems' performance that makes it up. To assess the robustness of our proposed face-iris multimodal system, we evaluated it by entering only the face images, in which we used four face databases (Georgia Tech, FERET, ORL, and SDUMLA-HMT). In this case, our proposed multimodal systems become a unimodal face recognition system, using YOLOv4-tiny to locate the region of interest and our proposed Deep CNN to extract features. The dimensionality is then reduced with PCA, and a LinearSVC classifier is used for the final classification.

The proposed multimodal system showed acceptable accuracy if the iris trait is missed or unavailable, achieving 98.08 % on the FARET database and 91.39% on the ORL database by using just one face image for training. Improved accuracy of 99.90%, 95.63%, and 93.90% was obtained on FERET, ORL, and Georgia Tech databases, respectively, when using two images for training. Despite using only two images for training and 82 images for testing, 63.20% accuracy was obtained on the SDUMLA-HTM face database. Experiments with cross-validation protocol achieved high accuracy exceeding 98.99% on the four databases (as shown in Table 7.2).

	Georgia Tech database	FERET database	ORL database	SDUMLA-HTM face database
Training with one sample, testing with the rest	81.75%	98.08%	91.39%	38.62%
Training with two samples, testing with the rest	93.08%	99.90%	95.63%	63.20%
Two-fold Cross-Validation	99.47%	99.86%	99.75%	98.99%
Five-fold Cross-Validation	100%	100%	100%	99.72%

Table 7.2: The accuracy rate of our proposed system on the different face datasets.

7.2.2.3 Experiments on iris datasets

Several face-iris multimodal systems' accuracy rates are dramatically degraded if one of the face or iris traits is missed or unavailable. To prove the effectiveness of our proposed face-iris multimodal system, we evaluated it by entering only the iris images, in which we used four iris databases (UBIRIS-v1, CASIA-Iris-Interval, IITD, and SDUMLA-HMT). In this case, our proposed systems become a unimodal iris recognition system, using YOLOv4-tiny to locate the iris region. Then, our proposed Deep CNN is used to extract features from the iris regions without any pre-processing. The dimensionality is reduced then with PCA, and the system used LinearSVC for the classification task.

The proposed multimodal systems achieved a remarkable accuracy rate if the face trait is missed or unavailable, reaching 99% on the Casia interval and IITD databases by applying a

cross-validation protocol. In addition, our system reached 100% and 97.87% based on the five-fold cross-validation protocol on the SDUMLA-HTM and the UBIRIS-v1 databases, respectively. The accuracy rates obtained by our proposed multimodal systems on the different iris databases are shown in Table 7.3.

	UBIRIS-v1 database	Casia-Iris-Interval database	IITD database	SDUMLA-HTM database	
				Left irises	Right irises
Training with one sample, testing with the rest	84.68%	84.64%	75%	89.13%	88.90%
Training with two samples, testing with the rest	94.67%	94.64%	86.38%	98.73%	98.11%
Two-fold Cross-Validation	96.11%	99.35%	99.46%	98.68%	99.62%
Five-fold Cross-Validation	97.87%	99.63%	99.82%	99.81%	100 %

Table 7.3: The accuracy rate of our proposed systems on the different iris datasets.

7.2.2.4 Experiments on face-iris multimodal datasets

We proposed three face-iris multimodal systems; in this section, we evaluated their performance on the SDUMLA-HTM and CASIA-ORL multimodal databases. We used several scenarios (fusion of right iris and face, left iris and face, both irises, or face and both irises) and multiple evaluation protocols. We started with the system that applied the fusion at the image level, and then we moved to the system that fused the features. Finally, we prove the performance of our system that applied the scores level fusion.

I. Experiments on our face-iris multimodal system that applied the fusion at the image level: The face region and both irises were detected utilizing Yolov4-tiny, and then the system combined the three detected regions to generate a fused image. Our Deep CNN model was applied to obtain features from the fused image. Dimension reduction was carried out with PCA, and the LinearSVC was used for classification.

Our system achieved a high accuracy rate despite it extract features from only one image, the fused image. Although we used only one image for training on the CASIA-ORL multimodal database, the system reached 99.58% when we fused both irises and face region and more than 96.25 in the other scenarios. In addition, if we used two images for training or applied a two-fold cross-validation protocol, the system achieved an accuracy rate of 100% on all the scenarios (see Table 7.4).

	Right iris and face	Left iris and face	Both irises	Both irises and face
Training with one sample, testing with the rest	97.92 %	96.67 %	96.25 %	99.58 %
Training with two samples, testing with the rest	100 %	100 %	100 %	100 %
Two-fold Cross-Validation	100 %	100 %	100 %	100 %

Table 7.4: The accuracy rate of our face-iris multimodal system that applied the fusion at the image level on the CASIA-ORL multimodal database.

On the SDUMLA-HTM multimodal database, by using only one image for training, our system that combines both irises achieved an accuracy rate of 97.90% (see Table 7.5). In addition, by applying a two-fold cross-validation protocol, our system reached 100% by fusing both irises or combining the face with both irises.

	Right iris and face	Left iris and face	Both irises	Both irises and face
Training with one sample, testing with the rest	89.60%	88.19%	97.90%	92.59%
Training with two samples, testing with the rest	90.20%	88.38%	98.41%	93.97%
Two-fold Cross-Validation	99.71%	99.84%	100 %	100%

Table 7.5: The accuracy rate of our face-iris multimodal system that applied the fusion at the image level on the SDUMLA-HTM multimodal dataset.

II. Experiments on our face-iris multimodal system that applied the fusion at the feature level: The face region and both irises were detected utilizing Yolov4-tiny. Then, the system applied our Deep CNN model inspired by the pre-trained Xception model to extract features from the detected region. Our system concatenated the obtained features and applied PCA for dimensionality reduction, and the LinearSVC was used for classification. On the CASIA-ORL database, our proposed system showed improved accuracy by using two samples in the training set.

The accuracy was 98.50% for both irises' fused features, 99.50% for the left iris and face, and 100% for other scenarios. In addition, it reached 100% by applying a two-fold cross-validation protocol in all scenarios (see Table 7.6).

On the SDUMLA-HTM, our proposed system was able to attain 100% when utilizing both irises with only two samples in the training set. In addition, by applying the same evaluation

	Right iris and face	Left iris and face	Both irises	Both irises and face
Training with one sample, testing with the rest	96.67 %	97.08 %	90.83 %	97.50 %
Training with two samples, testing with the rest	100 %	99.50 %	98.50 %	100 %
Two-fold Cross-Validation	100 %	100 %	100 %	100 %

Table 7.6: The accuracy rate of our face-iris multimodal system that applied the fusion at the feature level on the CASIA-ORL multimodal database.

protocol, our system reached 99.92% by fusing the face and both irises (see Table 7.7). Furthermore, it achieved 100% for most scenarios by applying a two-fold cross-validation protocol (see Table 7.7).

	Right iris and face	Left iris and face	Both irises	Both irises and face
Training with one sample, testing with the rest	84.52 %	83.97 %	95.06 %	93.51 %
Training with two samples, testing with the rest	98.40 %	98.62 %	100 %	99.92 %
Two-fold Cross-Validation	100 %	99.88 %	100 %	100 %

Table 7.7: The accuracy rate of our face-iris multimodal system that applied the fusion at the feature level on the SDUMLA-HTM multimodal database.

III. Experiments on our face-iris multimodal system that applied the fusion at the score level: Our proposed system, which is based on score-level fusion, first localized the ROIs based on YOLOv4-tiny and extracted features relying on our proposed Deep CNN model. Then, it applied PCA to reduce the dimensionality and calculated confidence scores using the LinearSVC decision function. The scores were normalized with Tanh normalization, and finally, the Product Rule (PR) or the Sum Rule (SR) was applied to combine the scores. The proposed system, applying the fusion at the score level, achieved outstanding accuracy surpassing all other proposed systems.

On the CASIA-ORL database, the proposed multimodal system, fusing both irises and face region at score level fusion, achieved 100% with only one sample image in the training set. In addition, it achieved 100% in most fusion scenarios when it used two samples for training. Furthermore, by applying the two-fold cross-validation, the proposed system reached 100% for all scenarios. (see Table 7.8).

	Right iris and face		Left iris and face		Both irises		Both irises and face	
	SR	PR	SR	PR	SR	PR	SR	PR
Training with one sample, testing with the rest	97.08%	97.08%	97.92%	97.92%	95.42%	95.42%	100%	100%
Training with two samples, testing with the rest	100%	99.50%	100%	100%	100%	100%	100%	100%
Two-fold Cross-Validation	100%	100%	100%	100 %	100 %	100 %	100%	100%

Table 7.8: The accuracy rate of our face-iris multimodal system that applied the fusion at the score level on the CASIA-ORL multimodal database

	Right iris and face		Left iris and face		Both irises		Both irises and face	
	SR	PR	SR	PR	SR	PR	SR	PR
Training with one sample, testing with the rest	93.40%	93.33%	91.79%	91.78%	97.65%	97.65%	99.05%	99.03%
Training with two samples, testing with the rest	99.45%	99.42%	99.18%	99.17%	100%	100%	100%	100%
Two-fold Cross-Validation	100%	100%	100%	100%	100%	100%	100%	100%

Table 7.9: The accuracy rate of our face-iris multimodal system that applied the fusion at the score level on the SDUMLA-HTM multimodal database.

On the SDUMLA-HTM multimodal database, by combining both irises and face region at the score level, our system obtained more than 99% with only one image in the training set. In addition, it reached 100% when we used two samples for the training. Moreover, the fusion of both irises only achieved an accuracy of 100% with two samples in the training set. By applying a two-fold cross-validation protocol, the accuracy rate of all the scenarios reached 100% (see Table 7.9).

Table 7.10 demonstrates the superiority of the proposed multimodal system that applied the scores fusion compared to other propositions, in which it achieved an accuracy rate of 100% in various scenarios.

Datasets	Level fusion	Scenario	Accuracy
CASIA-ORL	image-level fusion	Right iris + face	**100%**
		Left iris + face	**100%**
		Left iris + Right iris	**100%**
		Left iris+Face+Right iris	**100%**
	Feature-level fusion	Right iris + face	**100%**
		Left iris + face	**100%**
		Left iris + Right iris	**100%**
		Left iris+Face+Right iris	**100%**
	Score-level fusion (Sum Rule and product Rule)	Right iris + face	**100%**
		Left iris + face	**100%**
		Left iris + Right iris	**100%**
		Left iris+Face+Right iris	**100%**
SDUMLA-HMT	image-level fusion	Right iris + face	99.71%
		Left iris + face	99.84%
		Left iris + Right iris	**100%**
		Left iris+Face+Right iris	**100%**
	Feature-level fusion	Right iris + face	**100%**
		Left iris + face	99.88%
		Left iris + Right iris	**100%**
		Left iris+Face+Right iris	**100%**
	Score-level fusion (Sum Rule and Product Rule)	Right iris + face	**100%**
		Left iris + face	**100%**
		Left iris + Right iris	**100%**
		Left iris+Face+Right iris	**100%**

Table 7.10: The performance of the proposed multimodal systems on the CASIA-ORL and SDUMLA-HMT databases by applying the two-fold cross-validation protocol.

7.2.3 Comparison with state-of-the-art methods

In this section, a comparison was made between our approach and state-of-the-art methods. The comparison began with the state-of-the-art face recognition methods, followed by the iris recognition state-of-the-art methods. Finally, the results of our proposed multimodal systems are compared with current methods evaluated on SDUMLA-HTM and CASIA-ORL multimodal databases.

7.2.3.1 Comparison with face recognition state-of-the-art methods

Our system obtained an accuracy rate of 99.75% on the ORL database based on the two-fold cross-validation protocol and 100% relying on the five-fold cross-validation. On the FERET database, we obtained an accuracy rate of 100% with two and five-fold cross-validation protocols. The results achieved on the ORL and FERET databases outperformed the methods listed in Table 7.11.

By applying two-fold cross-validation, our proposed system surpassed state-of-the-art face recognition methods on the Georgia Tech database, achieving an accuracy rate of 100%. On the SDUMLA-HMT face database, our system obtained an accuracy rate of 98.99% with the two-fold cross-validation evaluation protocol, surpassing all the other methods listed in Table 7.11.

7.2.3.2 Comparison with iris recognition state-of-the-art methods

Our method outperformed all state-of-the-art methods listed in Table 7.12 on the IITD database, except the methods proposed by Alaslani [153] and Alaslni et al. [154], which were evaluated on a subset with only 60 classes. Our system proves its effectiveness on a subset containing 224 classes. It achieved an accuracy rate of 99.82% and 99.46%, applying five-fold cross-validation and two-fold cross-validation, respectively.

On the CASIA-Iris-Interval database, our system also obtained excellent accuracy rates surpassing state-of-the-art methods with a rate of 99.35% by applying a two-fold cross-validation protocol. Additionally, the UBIRIS-v1 database saw an accuracy rate of 97.87% with our system, outperforming all other state-of-the-art methods listed in Table 7.12.

Finally, on the SDUMLA-HMT iris dataset, the proposed system obtained a noteworthy accuracy rate of 100% by applying a five-fold cross-validation protocol. The method proposed by Al-Waisy et al. [168] also had a high accuracy rate. Still, it used complex pre-processing steps for segmenting and normalizing the iris region. Moreover, it used a deep CNN model called IrisConvNet, leading to high computing demands and lengthy training times due to the vast number of parameters.

Dataset	Method	Protocol	Accuracy
Georgia Tech	Aldhahab et al. [126]	five-fold cross-validation	98.40%
	See and Noor [127]	five-fold cross-validation	95.10%
	Zeghina et al [137]	Train: 80%, Test: 20%	97.41%
	Almabdy and Elrefaei [138]	Train: 80%, Test: 20%	98.31%
	Our system	**Two-fold cross-validation**	**99.33%**
		Five-fold cross-validation	**100%**
FERET	Ouslimani et al. [115]	Train: 71%, Test: 29%	93.46%
	Qin et al. [117]	Train: 71%, Test: 29%	76.50%
	Qin et al. [117]	Train: 71%, Test: 29%	78.00%
	Min et al. [130]	Train: 14%, Test: 86%	93.04%
	Zeng el al. [136]	Train: 14%, Test: 86%	93.90%
	Our system	**Two-fold cross-validation**	**100%**
		Five-fold cross-validation	**100%**
ORL	Ouyang et al. [114]	Train: 80%, Test: 20%	97.22%
	Ouslimani et al. [115]	Train: 80%, Test: 20%	98.61%
	Sapijaszko and Mikhael [116]	Train: 80%, Test: 20%	98.80%
	Elaggoune et al. [129]	70 %, 30 %	99.50%
	Elaggoune et al. [129]	70 %, 30 %	99.17%
	Min et al. [130]	Train: 10%, Test: 90%	97.77%
	Hosgurmath et al. [131]	80 %, 20 %	93.10%
	Tamilselvi and Karthikeyan [3]	/	97%
	Yallamandaiah and Purnachand [133]	70 %, 30 %	98.48%
	Our system	**Two-fold cross-validation**	**99.50%**
		Five-fold cross-validation	**100%**
SDUMLA HMT face	Al-Waisy et al. [11]	Train: 60%, Test: 40%	85.34%
	Alay and Al-Baity [170]	10-fold cross-validation	94.07%
	Our system	**Two-fold cross-validation**	**98.99%**
		Five-fold cross-validation	**99.72%**

Table 7.11: The average accuracy of our proposed multimodal system compared to state-of-the-art face recognition methods.

7.2.3.3 Comparison with face-iris multimodal recognition state-of-the-art methods

Our proposed multimodal system that applied the fusion at the score level has been evaluated with a two-fold cross-validation protocol using two multimodal biometric databases.

On the CASIA-ORL, our accuracy surpassed all the current methods illustrated in Table 7.13. The system proposed by Alsubari and Ramteke [166] rached 100% on the CASIA-ORL multimodal databse, but it was evaluated based on a subset with only 24 classes. Our multimodal system proves its robustness on this database by reaching 100% on a subset containing 40 classes.

On the SDUMLA-HMT database, our system obtained 100% accuracy in all fusion scenarios,

Datasets	State-of-the-art	No. classes	Evaluation protocol (Train, Test, Val)	Normal ization	Accuracy
UBIRIS v1	Taha and Ahmed [146]	100	(86%, 14%, -)	Yes	86.00%
	Taha and Ahmed [149]	100	(86%, 14%, -)	Yes	83.00%
	Patil and Vasanth [164]	241	/	No	95.03%
	Our System	**241**	**Two-fold CV**	**No**	**96.11%**
		241	**Five-fold CV**	**No**	**97.87%**
CASIA-Iris-Interval	Khotimah and Juniati [142]	10	Two-fold CV	Yes	92.63%
	Abdo et al. [7]	100	(71%, 29%, -)	Yes	98.50%
		100	(57%, 43%, -)	Yes	96.67%
	Abdalla et al. [144]	100	(71%, 29%, -)	Yes	99%
		100	(57%, 43%, -)	Yes	97%
	Abdo et al. [147]	100	(57%, 43%, -)	Yes	99%
	Alaslani [153]	60	(80%, 20%, -)	No	89%
	Alaslni et al. [154]	60	(80%, 20%, -)	No	91.60%
	Our System	**153**	**Two-fold CV**	**No**	**99.35%**
		153	**Five-fold CV**	**No**	**99.63%**
IITD	Winston and Hemanth [145]	/	(60%, 40%, -)	Yes	98.40%
	Alaslani [153]	**60**	**(80%, 20%, -)**	**No**	**100%**
	Alaslni et al [154]	**60**	**(80%, 20%, -)**	**No**	**100%**
	Arora and Bhatia [155]	224	(60%, 20%, 20%)	Yes	98%
	Yifeng Chen et al. [5]	224	(60%, 10%, 10%)	Yes	99.30%
	Sujana and Reddy [6]	224	(80%, 20%, -)	Yes	98.05%
	Shanto et al. [160]	25	(60%, 20%, 20%)	No	98%
	Our System	**224**	**Two-fold CV**	**No**	**99.46%**
		224	**Five-fold CV**	**No**	**99.82%**
SDUM-LA-HTM iris	Gad et al. [150]	100 Left	/	Yes	99.43%
		100 Right	/	Yes	99.49%
	Ak et al. [151]	106	Leave-one-out CV	Yes	93.60%
	Al-Waisy et al. [168]	106 Left	(60%, 40%, -)	Yes	99.62%
		106 Right	**(60%, 40%, -)**	**Yes**	**100%**
	Alay and Al-Baity [170]	106	(60%, 20%, 20%)	No	98.58%
	Our system	**106 Left**	**Two-fold CV**	**No**	**98.68%**
		106 Left	**Five-fold CV**	**No**	**99.81%**
		106 Right	**Two-fold CV**	**No**	**99.62%**
		106 Right	**Five-fold CV**	**No**	**100%**

Table 7.12: The recognition rate of our multimodal system compared to state-of-the-art iris recognition methods on the four databases, Where “Yes” or “No” was used to indicate the existence of the iris normalization process or not.

Datasets	Method	Level fusion	Scenario	Accuracy
CASIA-ORL	Ammour et al. [12]	Score-level fusion (SR)	Face + iris	99.16%
	Alsubari and Ramteke [166]	**Feature-level fusion**	**Face + iris**	**100%**
	Mansoura et al. [167]	Score-level fusion	Face + iris	98.33%
	Rasool [13]	Score-level fusion	Face + iris	97.53%
	Our system	**Score-level fusion (SR, PR)**	**Right iris + face**	**100%**
		Score-level fusion (SR, PR)	**Left iris + face**	**100%**
		Score-level fusion (SR, PR)	**Left iris + Right iris**	**100%**
		Score-level fusion (SR, PR)	**Left iris+Face+Right iris**	**100%**
SDUMLA-HMT	Al-Waisy et al. [11]	Score-level fusion (PR)	Face+Left iris	99.91%
		Score-level fusion (SR)	Face+Left iris	99.95%
		Score-level fusion (SR, PR)	**Right iris+Left iris**	**100%**
		Score-level fusion (SR, PR)	**Face+right iris**	**100%**
		Score-level fusion (SR, PR)	**Face+Right iris+Left iris**	**100%**
	Alay and Al-Baity [169]	Feature-level fusion	Face + iris+ finger vein	99.39%
		Score-level fusion (SR, PR)	**Face + iris+ finger vein**	**100%**
	Alay and Al-Baity [170]	Feature-level fusion	Face + iris	99.22%
		Score-level fusion (SR)	**Face + iris**	**100%**
	Soleymani et al. [171]	Feature-level fusion	Face + iris+ fingerprint	99.34%
	Soleymani et al. [172]	Feature-level fusion	Face + iris+ fingerprint	99.30%
	Our system	**Score-level fusion (SR, PR)**	**Right iris + face**	**100%**
		Score-level fusion (SR, PR)	**Left iris + face**	**100%**
		Score-level fusion (SR, PR)	**Left iris + Right iris**	**100%**
		Score-level fusion (SR, PR)	**Left iris+Face+Right iris**	**100%**

Table 7.13: The recognition rate of our multimodal systems compared to state-of-the-art methods on the two multimodal databases.

face with right iris, face with left iris, both irises, and face image with both irises. The method proposed by Al-Waisy et al. [11] also achieved good accuracy. However, its accuracy rate is dramatically degraded if the iris trait is missed or unavailable, where it obtained only 85.34% on the SDUMLA-HMT face dataset. Moreover, our proposed system outperformed Al-Waisy et al. [11] when we used the face and left iris (see Table 7.13).

7.3 Conclusion

In this Chapter, we proposed three face-iris multimodal biometric systems; the first applied the fusion at the image level, the second fused at the feature level, and the third was based on the score level fusion. The proposed systems applied Yolov4-tiny to detect the face and both iris regions and the proposed Deep CNN models to extract features. In addition, the systems used PCA to keep relevant features, and the identity is predicted with the help of the LinearSVC library. We used ten benchmark datasets to prove the effectiveness of our proposed system. Firstly, we evaluate the performance of our proposed systems on four face datasets to demonstrate their robustness in the case when the iris trait is missed or unavailable. Secondly, we evaluate the performance of our proposed systems on four iris datasets to prove their robustness if the face trait is missed or unavailable. Thirdly, we used two multimodal datasets to demonstrate the effectiveness and robustness of our proposed systems.

Our proposed systems reached an accuracy rate of 100% on the Georgia Tech, FERET, and ORL face datasets through the implementation of a five-fold cross-validation procedure. Furthermore, we obtained an accuracy rate of 99.72% on the SDUMLA-HMT face dataset. Additionally, the iris systems' recognition accuracy reached 97.87%, 99.63%, 99.82%, and 100% on the UBIRIS-v1, CASIA-Iris-Interval, IITD, and SDUMLA-HMT iris datasets, respectively. Our proposed multimodal systems achieved an accuracy rate of 100% on the CASIA-ORL and SDUMLA-HMT multimodal databases by applying the two-fold cross-validation protocol, despite the system that used image-level fusion, it extracted the features from only a single image (the fused image).

The proposed face-iris multimodal system that applied the fusion at the score level achieved good results; it reached 100% and more than 99% on the CASIA-ORL and SDUMLA-HMT multimodal databases, respectively, by using only one sample for training. The experiments proved the robustness and effectiveness of the proposed face-iris multimodal system, which can be used in different applications requiring a high level of security.

CHAPTER 8

CONCLUSION AND FUTURE WORK

"Little by little, a little becomes a lot."

Tanzanian proverb

Biometrics is one of the most important applications in computer vision. It outperformed traditional authentication methods, achieving high performance in identifying persons. Face and iris recognition has recently been integrated into several applications and succeeded wildly, especially since these biometric traits can be captured without touching the sensors, limiting the potential spread of the COVID-19 pandemic and other diseases by touch. However, the face recognition system's performance dramatically degraded under unconstrained conditions, such as illumination variation, occlusion, and poses. Similarly, the performance of iris recognition systems can be affected by non-cooperative conditions, such as reflection, blur, partial closure of eyes, and pupil dilation/constriction due to light conditions, which can reduce the system's performance.

In order to overcome the challenges that affect face recognition accuracy, this thesis proposed a robust face recognition system that identifies the significant facial regions using SIFT and extracts features from them using ALTP. The proposed handcraft system has been evaluated on two face datasets (ORL and Feret dataset), which contain images captured under uncontrolled conditions such as illumination variations, facial expressions, occlusion, etc. The obtained accuracy rate demonstrated the robustness of our proposed, reaching 99.75% on the ORL dataset and 95.12% on the FERET dataset. In addition, a Deep Learning approach was used to build a second face recognition system to improve accuracy. The proposed system built relying on the pre-trained AlexNet-v2 model achieved a high accuracy rate on the two previous datasets, reaching 100% on the first and 99.89% on the second dataset.

Furthermore, this study proposed a new effective system, overcoming the curse of illumination variation that affects the accuracy of most face recognition systems. The proposed robust-illumination system used a new Deep CNN model inspired by the pre-trained VGG16 model. The system has been evaluated on two benchmark datasets captured under lighting variation and obtained high results compared to state-of-the-art methods, reaching a 99.32% accuracy on the Extended Yale B dataset and a 99.79% accuracy on the AR dataset.

This thesis presents novel contributions to address the issues affecting iris recognition systems' performance. Specifically, it proposes an effective iris recognition system based on Deep Learning. The proposed system fine-tuned the pre-trained Yolov4-tiny model and used them to localized iris regions. In addition, our system used a novel Deep CNN inspired by the pre-trained Inception-v3 model to extract features from the iris regions. The system used only five inception modules from the ImageNet pre-trained Inception-v3 model without fine-tuning them on the iris database; it also extracted features without normalizing and segmenting the iris region. All these characteristics help the system save computational resources and time. The proposed iris recognition system achieves exceptional performance on four benchmark iris datasets captured under non-ideal conditions, including illumination variations, specular reflection, pupil dilation/constriction, and occlusions. The system achieves a new state-of-the-art accuracy rate reaching 99.91%, 99.60%, 99.91% and 99.19% on the IITD, CASIA-Iris-v1, CASIA-Iris-Interval, and CASIA-Iris-Thousand, respectively.

The proposed face and iris recognition systems achieved high accuracy compared to state-of-the-art methods. However, relying solely on a unimodal biometric trait is inadequate for high-security needs such as those required in military and government applications.

This dissertation proposed three face-iris multimodal biometric systems designed for applications requiring high-level security. The first system is based on image-level fusion, the second performs fusion at the feature level, and the third is based on the score-level-fusion. The proposed systems used Yolov4-tiny to detect the face and both iris regions, and a new deep CNN model inspired by the pre-trained Xception model is used to extract features. The proposed systems achieved excellent accuracy rates by applying the two-fold cross-validation protocol, reaching 100% on the CASIA-ORL and SDUMLA-HMT benchmark multimodal databases. By comparing the three proposed multimodal systems, the face-iris multimodal system that applied the fusion at the score level achieved the best results. By using only one sample for training, the system achieved accuracy rates of 100% and more than 99% on the CASIA-ORL and SDUMLA-HMT multimodal databases, respectively.

The dissertation's future work can be summarized as follows:

- We will focus on developing robust techniques against spoofing attacks. These techniques test the biometric sample to ensure that it is a real sample, not a fake one. Subsequently, we will seek to integrate these proposed techniques into the proposed systems.

- We aim to study multi-level fusion; we will seek to propose a new multimodal system that applies fusion at different levels to improve user identification.
- A new multimodal system will be developed for smartphone devices by fusing physical and behavioral biometric traits, such as face patterns, touch stroke, and phone-movement.
- We plan to evaluate the performance of our proposed systems in other physical or behavioral traits such as fingerprint, retina, signature, and voice.

BIBLIOGRAPHY

[1] L. Yang, B. Yang, and X. Gu, "Adversarial reconstruction CNN for illumination-robust frontal face image recovery and recognition," *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, vol. 15, no. 2, pp. 18–33, 2021.

[2] R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, "Illumination-robust face recognition based on deep convolutional neural networks architectures," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 2, pp. 1015–1027, 2020.

[3] M. Tamilselvi and S. Karthikeyan, "An ingenious face recognition system based on HRPSM_CNN under unrestrained environmental condition," *Alexandria Engineering Journal*, vol. 61, no. 6, pp. 4307–4321, 2022.

[4] G. Liu, W. Zhou, L. Tian, W. Liu, Y. Liu, and H. Xu, "An Efficient and Accurate Iris Recognition Algorithm Based on a Novel Condensed 2-ch Deep Convolutional Neural Network," *Sensors*, vol. 21, no. 11, p. 3721, 2021.

[5] Y. Chen, C. Wu, and Y. Wang, "T-center: a novel feature extraction approach towards large-scale iris recognition," *IEEE Access*, vol. 8, pp. 32365–32375, 2020.

[6] S. Sujana and V. S. K. Reddy, "An Effective CNN based Feature Extraction Approach for Iris Recognition System," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 6, pp. 4595–4604, 2021.

[7] A. A. Abdo, A. Lawgali, and A. K. Zohdy, "Iris recognition based on histogram equalization and discrete cosine transform," in *Proceedings of the 6th International Conference on Engineering & MIS 2020*, pp. 1–5, 2020.

[8] M. Trokielewicz, A. Czajka, and P. Maciejewicz, "Iris recognition in cases of eye pathology," *Biometrics under Biomedical Considerations*, pp. 41–69, 2019.

[9] M. O. Oloyede and G. P. Hancke, "Unimodal and multimodal biometric sensing systems: a review," *IEEE access*, vol. 4, pp. 7532–7555, 2016.

[10] M. M. Selim, R. O. Mahmoud, and O. A. Muhi, "A Feature level Fusion of Multimodal Biometric Authentication System," *Journal of Convergence Information Technology*, vol. 13, no. 1, pp. 1–11, 2018.

[11] 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)*, pp. 163–168, IEEE, 2017.

[12] B. Ammour, L. Boubchir, T. Bouden, and M. Ramdani, "Face–iris multimodal biometric identification system," *Electronics*, vol. 9, no. 1, p. 85, 2020.

[13] R. A. Rasool, "Feature-Level vs. Score-Level Fusion in the Human Identification System," *Applied Computational Intelligence and Soft Computing*, vol. 2021, pp. 1–10, 2021.

[14] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.

[15] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[18] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2, pp. 1150–1157, Ieee, 1999.

[19] W. Yang, Z. Wang, and B. Zhang, "Face recognition using adaptive local ternary patterns method," *Neurocomputing*, vol. 213, pp. 183–190, 2016.

[20] A. Krizhevsky, "One weird trick for parallelizing convolutional neural networks," *arXiv preprint arXiv:1404.5997*, 2014.

[21] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "Scaled-YOLOv4: Scaling Cross Stage Partial Network," pp. 13024–13033, 2021.

[22] M. Ramya, V. Krishnaveni, and K. S. Sridharan, "Certain investigation on iris image recognition using hybrid approach of Fourier transform and Bernstein polynomials," *Pattern Recognition Letters*, vol. 94, pp. 154–162, 2017.

[23] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on circuits and systems for video technology*, vol. 14, no. 1, pp. 4–20, 2004.

[24] S. Angle, R. Bhagtani, and H. Chheda, "Biometrics: A further echelon of security," in *UAE International Conference on Biological and Medical Physics*, Citeseer, 2005.

[25] H. Cummins, "Ancient finger prints in clay," *Journal of Criminal Law and Criminology (1931-1951)*, vol. 32, no. 4, pp. 468–481, 1941.

[26] S. D. Raut and V. T. Humbe, "Biometric palm prints feature matching for person identification," *International Journal of Modern Education and Computer Science*, vol. 4, no. 11, p. 61, 2012.

[27] Y. L. Nigeria, "Analysis, design and implementation of human fingerprint patterns system "towards age & gender determination, ridge thickness to valley thickness ratio (RTVTR) & ridge count on gender detection," *International Journal of Advanced Research in Artificial Intelligence*, vol. 1, no. 2, 2012.

[28] R. Belguechi, V. Alimi, E. Cherrier, P. Lacharme, C. Rosenberger, and Others, "An overview on privacy preserving biometrics," *Recent Application in Biometrics*, vol. 65, p. 84, 2011.

[29] H. Faulds, "Poroscopy: the scrutiny of sweat-pores for identification," *Nature*, vol. 91, no. 2286, pp. 635–636, 1913.

[30] Z. R. Hagins, "Fashioning the 'born criminal'on the beat: Juridical photography and the police municipale in Fin-de-Siècle Paris," *Modern & Contemporary France*, vol. 21, no. 3, pp. 281–296, 2013.

[31] J. E. Hoover, "International exchange of fingerprints," *Am. Inst. Crim. L. & Criminology*, vol. 24, p. 664, 1933.

[32] K. P. Li, J. E. Dammann, and W. D. Chapman, "Experimental studies in speaker verification, using an adaptive system," *The Journal of the Acoustical Society of America*, vol. 40, no. 5, pp. 966–978, 1966.

[33] P. Komarinski, *Automated fingerprint identification systems (AFIS)*. Elsevier, 2005.

[34] L. D. Harmon, M. K. Khan, R. Lasch, and P. F. Ramig, "Machine identification of human faces," *Pattern Recognition*, vol. 13, no. 2, pp. 97–110, 1981.

[35] H. D. Crane and J. S. Ostrem, "Automatic signature verification using a three-axis force-sensitive pen," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 3, pp. 329–337, 1983.

[36] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE transactions on pattern analysis and machine intelligence*, vol. 15, no. 11, pp. 1148–1161, 1993.

[37] Z. Korotkaya, "Biometric person authentication: Odor," *Department of Information Technology, Laboratory of Applied Mathematics, Lappeenranta University of Technology*, p. 1, 2003.

[38] C.-K. Chen, C.-L. Lin, C.-T. Chiang, and S.-L. Lin, "Personalized information encryption using ECG signals with chaotic functions," *Information Sciences*, vol. 193, pp. 125–140, 2012.

[39] C. D. Holland and O. V. Komogortsev, "Complex eye movement pattern biometrics: Analyzing fixations and saccades," in *2013 International conference on biometrics (ICB)*, pp. 1–8, IEEE, 2013.

[40] M. K. Bhowmik, K. Saha, S. Majumder, G. Majumder, A. Saha, A. N. Sarma, D. Bhattacharjee, D. K. Basu, and M. Nasipuri, "Thermal infrared face recognition—a biometric identification technique for robust security system," *Reviews, refinements and new ideas in face recognition*, vol. 7, pp. 113–138, 2011.

[41] D. Day, *Biometric Applications, Overview*, pp. 169–174. Boston, MA: Springer US, 2015.

[42] J. Sohankar, K. Sadeghi, A. Banerjee, and S. K. S. Gupta, "E-bias: A pervasive eeg-based identification and authentication system," in *Proceedings of the 11th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, pp. 165–172, 2015.

[43] J. Priesnitz, R. Huesmann, C. Rathgeb, N. Buchmann, and C. Busch, "Mobile contactless fingerprint recognition: implementation, performance and usability aspects," *Sensors*, vol. 22, no. 3, p. 792, 2022.

[44] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, present, and future of face recognition: A review," *Electronics*, vol. 9, no. 8, p. 1188, 2020.

[45] A. Kamboj, R. Rani, and A. Nigam, "A comprehensive survey and deep learning-based approach for human recognition using ear biometric," *The Visual Computer*, vol. 38, no. 7, pp. 2383–2416, 2022.

[46] A. Jain, R. Bolle, and S. Pankanti, *Biometrics: personal identification in networked society*, vol. 479. Springer Science & Business Media, 1999.

[47] A. Bertillon, *La photographie judiciaire: avec un appendice sur la classification et l'identification anthropométriques*. París: Gauthier-Villars, 1890.

[48] Ž. Emeršič, V. Štruc, and P. Peer, "Ear recognition: More than a survey," *Neurocomputing*, vol. 255, pp. 26–39, 2017.

[49] Z. Mu, L. Yuan, Z. Xu, D. Xi, and S. Qi, "Shape and structural feature based ear recognition," in *Chinese Conference on Biometric Recognition*, pp. 663–670, Springer, 2004.

[50] F. Sadikoglu and S. Uzelaltinbulat, "Biometric retina identification based on neural network," *Procedia Computer Science*, vol. 102, pp. 26–33, 2016.

[51] R. CHLAOUA, *Combination of Multiple Biometrics for Recognition of Persons*. Phd thesis, UNIVERSITY of KASDI MERBAH OUARGLA, Algeria, 2019.

[52] M. U. Akram, A. Tariq, and S. A. Khan, "Retinal recognition: Personal identification using blood vessels," in *2011 International Conference for Internet Technology and Secured Transactions*, pp. 180–184, IEEE, 2011.

[53] L. Zoubida, *Reconnaissance des Formes appliquée à la Biométrie*. PhD thesis, University Djillali liabes of Sidi Bel Abbes, Algeria, 2018.

[54] P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki, "An introduction evaluating biometric systems," *Computer*, vol. 33, no. 2, pp. 56–63, 2000.

[55] W. K. Kong and D. Zhang, "Palmprint texture analysis based on low-resolution images for personal authentication," in *2002 International Conference on Pattern Recognition*, vol. 3, pp. 807–810, IEEE, 2002.

[56] F. Monrose and A. Rubin, "Authentication via keystroke dynamics," in *Proceedings of the 4th ACM Conference on Computer and Communications Security*, pp. 48–56, 1997.

[57] M. S. Nixon, J. N. Carter, D. Cunado, P. S. Huang, and S. V. Stevenage, "Automatic gait recognition," in *Biometrics*, pp. 231–249, Springer, 1996.

[58] N. Karimian, D. L. Woodard, and D. Forte, "On the vulnerability of ECG verification to online presentation attacks," in *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 143–151, IEEE, 2017.

[59] M. S. Islam and N. Alajlan, "Biometric template extraction from a heartbeat signal captured from fingers," *Multimedia Tools and Applications*, vol. 76, no. 10, pp. 12709–12733, 2017.

[60] C. Steinberg, F. Philippon, M. Sanchez, P. Fortier-Poisson, G. O'Hara, F. Molin, J.-F. Sarrazin, I. Nault, L. Blier, K. Roy, and Others, "A novel wearable device for continuous ambulatory ECG recording: proof of concept and assessment of signal quality," *Biosensors*, vol. 9, no. 1, p. 17, 2019.

[61] S. Wahabi, S. Pouryayevali, S. Hari, and D. Hatzinakos, "On evaluating ECG biometric systems: Session-dependence and body posture," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 11, pp. 2002–2013, 2014.

[62] I. Odinaka, P.-H. Lai, A. D. Kaplan, J. A. O'Sullivan, E. J. Sirevaag, and J. W. Rohrbaugh, "ECG biometric recognition: A comparative analysis," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1812–1824, 2012.

[63] N. Belgacem, F. Bereksi-Reguig, A. Nait-Ali, and R. Fournier, "Person identification system based on electrocardiogram signal using LabVIEW," *International Journal on Computer Science and Engineering*, vol. 4, no. 6, p. 974, 2012.

[64] Nist, "Summary of NIST Standards for Biometric Accuracy, Tamper Resistance, and Interoperability," 2002.

[65] Y. Chen, S. C. Dass, and A. K. Jain, "Fingerprint quality indices for predicting authentication performance," in *International conference on audio-and video-based biometric person authentication*, pp. 160–170, Springer, 2005.

[66] A. S. N. Nassar, *A Hybrid Multibiometric System for Personal Identification Based on Face and Iris Traits. The Development of an automated computer system for the identification of humans by integrating facial and iris features using Localization, Feature Extraction, Hand*. PhD thesis, 2018.

[67] A. Herbadji, *Amélioration de la performance des Systèmes d'identification et authentification biométriques par des techniques multimodales avancées*. PhD thesis, Université de M'sila, Algeria, 2021.

[68] V. Dhir, A. Singh, R. Kumar, and G. Singh, "Biometric recognition: A modern era for security," *International Journal of Engineering Science and Technology*, vol. 2, no. 8, pp. 3364–3380, 2010.

[69] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction*, pp. 1–49. Boston, MA: Springer US, 2011.

[70] A. Ross and K. Nandakumar, "J. anil k,"handbook of multibiometrics,"," *J. Chem. Inf. Model*, vol. 53, no. 9, pp. 1689–1699, 2006.

[71] R. Brunelli and D. Falavigna, "Person identification using multiple cues," *IEEE transactions on pattern analysis and machine intelligence*, vol. 17, no. 10, pp. 955–966, 1995.

[72] E. S. Bigün, J. Bigün, B. Duc, and S. Fischer, "Expert conciliation for multi modal person authentication systems by Bayesian statistics," in *International Conference on Audio-and Video-Based Biometric Person Authentication*, pp. 291–300, Springer, 1997.

[73] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[74] A. Ross and A. K. Jain, "Multimodal biometrics: An overview," in *2004 12th European signal processing conference*, pp. 1221–1224, IEEE, 2004.

[75] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[76] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[77] L. Fei, B. Zhang, C. Tian, S. Teng, and J. Wen, "Jointly learning multi-instance hand-based biometric descriptor," *Information Sciences*, vol. 562, pp. 1–12, 2021.

[78] A. Rattani, D. R. Kisku, M. Bicego, and M. Tistarelli, "Feature level fusion of face and fingerprint biometrics," in *2007 First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, pp. 1–6, IEEE, 2007.

[79] P. Wild, P. Radu, L. Chen, and J. Ferryman, "Robust multimodal face and fingerprint fusion in the presence of spoofing attacks," *Pattern Recognition*, vol. 50, pp. 17–25, 2016.

[80] D. L. Woodard, S. Pundlik, P. Miller, R. Jillela, and A. Ross, "On the fusion of periocular and iris biometrics in non-ideal imagery," in *2010 20th International Conference on Pattern Recognition*, pp. 201–204, IEEE, 2010.

[81] Q. Zhang, H. Li, Z. Sun, and T. Tan, "Deep feature fusion for iris and periocular biometrics on mobile devices," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2897–2912, 2018.

[82] B. A. El-Rahiem, F. E. A. El-Samie, and M. Amin, "Multimodal biometric authentication based on deep fusion of electrocardiogram (ECG) and finger vein," *Multimedia Systems*, vol. 28, no. 4, pp. 1325–1337, 2022.

[83] P. P. Sarangi, D. R. Nayak, M. Panda, and B. Majhi, "A feature-level fusion based improved multimodal biometric recognition system using ear and profile face," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 4, pp. 1867–1898, 2022.

[84] P. S. Sanjekar and J. B. Patil, "Multimodal biometrics with serial, parallel and hierarchical mode at decision level fusion," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 3, pp. 1303–1310, 2019.

[85] D. R. Kisku, J. K. Sing, M. Tistarelli, and P. Gupta, "Multisensor biometric evidence fusion for person authentication using wavelet decomposition and monotonic-decreasing graph," in *2009 seventh international conference on advances in pattern recognition*, pp. 205–208, IEEE, 2009.

[86] S. Noushath, M. Imran, K. Jetly, A. Rao, and G. H. Kumar, "Multimodal biometric fusion of face and palmprint at various levels," in *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1793–1798, IEEE, 2013.

[87] S. Z. Li, *Encyclopedia of Biometrics: I-Z.*, vol. 2. Springer Science & Business Media, 2009.

[88] M. F. Nadheen and S. Poornima, "Feature level fusion in multimodal biometric authentication system," *International Journal of Computer Applications*, vol. 69, no. 18, 2013.

[89] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of biometrics*. Springer Science & Business Media, 2007.

[90] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE transactions on pattern analysis and machine intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[91] L. I. Kuncheva, "That elusive diversity in classifier ensembles," in *Iberian conference on pattern recognition and image analysis*, pp. 1126–1138, Springer, 2003.

[92] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE transactions on systems, man, and cybernetics*, vol. 22, no. 3, pp. 418–435, 1992.

[93] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *bmvc*, vol. 1, p. 6, British Machine Vision Association, 2015.

[94] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "Elasticface: Elastic margin loss for deep face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1578–1587, 2022.

[95] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[96] M. Kas, *Development of handcrafted and deep based methods for face and facial expression recognition*. PhD thesis, Bourgogne Franche-Comté, 2021.

[97] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[98] W. Chen, H. Huang, S. Peng, C. Zhou, and C. Zhang, "YOLO-face: a real-time face detector," *The Visual Computer*, vol. 37, no. 4, pp. 805–813, 2021.

[99] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[100] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE transactions on image processing*, vol. 19, no. 6, pp. 1635–1650, 2010.

[101] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*, pp. 404–417, Springer, 2006.

[102] D. Harwood, T. Ojala, M. Pietikäinen, S. Kelman, and L. Davis, "Texture classification by center-symmetric auto-correlation, using Kullback discrimination of distributions," *Pattern Recognition Letters*, vol. 16, no. 1, pp. 1–10, 1995.

[103] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *European conference on computer vision*, pp. 469–481, Springer, 2004.

[104] X. Wang, T. X. Han, and S. Yan, "An HOG-LBP human detector with partial occlusion handling," in *2009 IEEE 12th international conference on computer vision*, pp. 32–39, IEEE, 2009.

[105] S. K. Naji and M. H. Hamad, "HUMAN IDENTIFICATION BASED ON FACE RECOGNITION SYSTEM," *Journal of Engineering and Sustainable Development*, vol. 25, no. 1, pp. 80–91, 2021.

[106] S. Jain, B. L. S. Kumar, and R. Shettigar, "Comparative study on SIFT and SURF face feature descriptors," in *2017 International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pp. 200–205, 2017.

[107] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[108] 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*, pp. 1–9, 2015.

[109] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[110] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.

[111] P. Kalaiarasi and P. Esther Rani, "A Comparative Analysis of AlexNet and GoogLeNet with a Simple DCNN for Face Recognition," in *Advances in Smart System Technologies*, pp. 655–668, Springer, 2021.

[112] J. F. G. Tremoço, *Improving deep learning face recognition for ID and travel document applications with quality assessment*. PhD thesis, Universidade de Coimbra, 2021.

[113] V. Vapnik, "Pattern recognition using generalized portrait method," *Automation and remote control*, vol. 24, pp. 774–780, 1963.

[114] A. Ouyang, Y. Liu, S. Pei, X. Peng, M. He, and Q. Wang, "A hybrid improved kernel LDA and PNN algorithm for efficient face recognition," *Neurocomputing*, vol. 393, pp. 214–222, 2020.

[115] F. Ouslimani, A. Ouslimani, and Z. Ameur, "Rotation-invariant features based on directional coding for texture classification," *Neural Computing and Applications*, vol. 31, no. 10, pp. 6393–6400, 2019.

[116] G. M. Sapijaszko and W. B. Mikhael, "Facial Recognition System Using Mixed Transform and Multilayer Sigmoid Neural Network Classifier," *Circuits, Systems, and Signal Processing*, vol. 39, pp. 6142–6161, 2020.

[117] Y. Qin, L. Sun, and Y. Xu, "Exploring of alternative representations of facial images for face recognition," *International Journal of Machine Learning and Cybernetics*, pp. 1–7, 2020.

[118] J. Yadav, N. Rajpal, and R. Mehta, "A new illumination normalization framework via homomorphic filtering and reflectance ratio in DWT domain for face recognition," *Journal of Intelligent & Fuzzy Systems*, vol. 35, no. 5, pp. 5265–5277, 2018.

[119] J. Yadav, N. Rajpal, and R. Mehta, "An improved hybrid illumination normalisation and feature extraction model for face recognition," *International Journal of Applied Pattern Recognition*, vol. 5, no. 2, pp. 149–170, 2018.

[120] J. Yadav, R. Mehta, and Others, "An improved illumination normalization and robust feature extraction technique for face recognition under varying illuminations," *Arabian Journal for Science and Engineering*, vol. 44, no. 11, pp. 9067–9086, 2019.

[121] J. Wen, X. Fang, J. Cui, L. Fei, K. Yan, Y. Chen, and Y. Xu, "Robust sparse linear discriminant analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 2, pp. 390–403, 2018.

[122] M.-D. Yuan, D.-Z. Feng, Y. Shi, and W.-J. Liu, "Dimensionality reduction by collaborative preserving Fisher discriminant analysis," *Neurocomputing*, vol. 356, pp. 228–243, 2019.

[123] L. Hu, W. Zhang, and Z. Dai, "Joint Sparse Locality-Aware Regression for Robust Discriminative Learning," *IEEE Transactions on Cybernetics*, 2021.

[124] T. Ayyavoo and J. J. Suseela, "Illumination pre-processing method for face recognition using 2D DWT and CLAHE," *Iet Biometrics*, vol. 7, no. 4, pp. 380–390, 2018.

[125] A. Dahmouni, N. Aharrane, K. El Moutaouakil, and K. Satori, "A face recognition based biometric solution in education," *Pattern Recognition and Image Analysis*, vol. 28, no. 4, pp. 758–770, 2018.

[126] A. Aldhahab, T. Alobaidi, A. Q. Althahab, and W. B. Mikhael, "Applying multiresolution analysis to vector quantization features for face recognition," in *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*, pp. 598–601, IEEE, 2019.

[127] Y. C. See and N. M. Noor, "Integrating complete gabor filter to the random forest classification algorithm for face recognition," *Journal of Engineering Science and Technology*, vol. 14, no. 2, pp. 859–874, 2019.

[128] V. Struc, B. Vesnicer, and N. Pavesic, "The phase-based gabor fisher classifier and its application to face recognition under varying illumination conditions," in *2008 2nd International Conference on Signal Processing and Communication Systems*, pp. 1–6, IEEE, 2008.

[129] H. Elaggoune, M. Belahcene, and S. Bourennane, "Hybrid descriptor and optimized CNN with transfer learning for face recognition," *Multimedia Tools and Applications*, vol. 81, no. 7, pp. 9403–9427, 2022.

[130] R. Min, S. Xu, and Z. Cui, "Single-sample face recognition based on feature expansion," *IEEE Access*, vol. 7, pp. 45219–45229, 2019.

[131] S. Hosgurmath, V. V. Mallappa, N. B. Patil, and V. Petli, "A face recognition system using convolutional feature extraction with linear collaborative discriminant regression classification," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 2, p. 1468, 2022.

[132] X. Qu, S. Kim, R. Cui, and H. J. Kim, "Linear collaborative discriminant regression classification for face recognition," *Journal of Visual Communication and Image Representation*, vol. 31, pp. 312–319, 2015.

[133] S. Yallamandaiah and N. Purnachand, "A novel face recognition technique using Convolutional Neural Network, HOG, and histogram of LBP features," in *2022 2nd International Conference on Artificial Intelligence and Signal Processing (AISP)*, pp. 1–5, IEEE, 2022.

[134] L. Yang, B. Yang, and X. Gu, "A deep reconstruction CNN for illumination-robust face image recovery and recognition," in *2018 IEEE 17th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*, pp. 417–422, IEEE, 2018.

[135] N.-S. Vu and A. Caplier, "Face recognition with patterns of oriented edge magnitudes," in *European conference on computer vision*, pp. 313–326, Springer, 2010.

[136] J. Zeng, X. Zhao, J. Gan, C. Mai, Y. Zhai, and F. Wang, "Deep convolutional neural network used in single sample per person face recognition," *Computational intelligence and neuroscience*, vol. 2018, 2018.

[137] A. O. Zeghina, O. Zoubia, and A. Behloul, "Face Recognition Based on Harris Detector and Convolutional Neural Networks," in *International Symposium on Modelling and Implementation of Complex Systems*, pp. 163–171, Springer, 2020.

[138] S. Almabdy and L. Elrefaei, "Feature extraction and fusion for face recognition systems using pre-trained convolutional neural networks," *International Journal of Computing and Digital Systems*, vol. 9, pp. 1–7, 2021.

[139] J. Illingworth and J. Kittler, "The adaptive hough transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 5, pp. 690–698, 1987.

[140] J. Daugman, "How iris recognition works," in *The essential guide to image processing*, pp. 715–739, Elsevier, 2009.

[141] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[142] C. Khotimah and D. Juniati, "Iris recognition using feature extraction of box counting fractal dimension," in *Journal of Physics: Conference Series*, vol. 947, p. 12004, IOP Publishing, 2018.

[143] M. Dua, R. Gupta, M. Khari, and R. G. Crespo, "Biometric iris recognition using radial basis function neural network," *Soft Computing*, vol. 23, no. 22, pp. 11801–11815, 2019.

[144] M. A. E. Abdalla, A. A. Abdo, and A. O. Lawgali, "Utilizing Discrete Wavelet Transform and Discrete Cosine Transform for Iris Recognition," in *2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, pp. 283–286, IEEE, 2020.

[145] J. J. Winston and D. J. Hemanth, "Performance-enhanced modified self-organising map for iris data classification," *Expert Systems*, vol. 38, no. 1, p. e12467, 2021.

[146] M. A. Taha and H. M. Ahmed, "Iris Features Extraction and Recognition based on the Local Binary Pattern Technique," in *2021 International Conference on Advanced Computer Applications (ACA)*, pp. 16–21, IEEE, 2021.

[147] A. Abdo, W. El-Tarhouni, W. Younus, and A. Abraheem, "Iris recognition system based on fuzzy local binary pattern histogram and multiple classifiers," in *2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*, pp. 452–457, IEEE, 2022.

[148] S. Katsigiannis, E. Keramidas, and D. Maroulis, "Flbp: Fuzzy local binary patterns," in *Local Binary Patterns: New Variants and Applications*, pp. 149–175, Springer, 2014.

[149] M. A. Taha and H. M. Ahmed, "Speeded Up Robust Features Descriptor for Iris Recognition Systems," *Journal of University of Babylon for Pure and Applied Sciences*, vol. 29, no. 2, pp. 244–257, 2021.

[150] R. Gad, M. Talha, A. A. Abd El-Latif, M. Zorkany, E.-S. Ayman, E.-F. Nawal, and G. Muhammad, "Iris recognition using multi-algorithmic approaches for cognitive internet of things (ciot) framework," *Future Generation Computer Systems*, vol. 89, pp. 178–191, 2018.

[151] T. A. Ak, A. Steluta, *et al.*, "An iris recognition system using a new method of iris localization," *International Journal of Open Information Technologies*, vol. 9, no. 7, pp. 67–76, 2021.

[152] A. Tahir and A. I. Bindian, "Localizarea Irisului Pentru Sistemul Biometric De Identificare A Ersoanelor," in *The XVIII International Conference on Multidisciplinary,'Professor Dorin Paul-Romanian hydropower founder', Sebes, Romania, June-2016, in the Romanian Journal of Sceince and Engineering, Revista" ȘTIINȚĂ ȘI INGINERIE"*, vol. 30, p. 215, 2016.

[153] M. G. Alaslani, "Convolutional neural network based feature extraction for iris recognition," *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 10, 2018.

[154] G. Alaslni, L. A. Elrefaei, and Others, "Transfer Learning with Convolutional Neural Networks for IRIS Recognition," *Int. J. Artif. Intell. Appl*, vol. 10, no. 5, pp. 47–64, 2019.

[155] S. Arora and M. P. S. Bhatia, "A computer vision system for iris recognition based on deep learning," in *2018 IEEE 8th International Advance Computing Conference (IACC)*, pp. 157–161, IEEE, 2018.

[156] M. Chakraborty, M. Roy, P. K. Biswas, and P. Mitra, "Unsupervised Pre-Trained, Texture Aware and Lightweight Model for Deep Learning Based Iris Recognition Under Limited Annotated Data," in *2020 IEEE International Conference on Image Processing (ICIP)*, pp. 1351–1355, IEEE, 2020.

[157] K. Kranthi Kumar, R. Bharadwaj, S. Ch, and S. Sujana, "Effective Deep Learning approach based on VGG-Mini Architecture for Iris Recognition," *Annals of the Romanian Society for Cell Biology*, pp. 4718–4726, 2021.

[158] T. R. Shanbagavalli and Others, "EMiCoAReNet: An Effective Iris Recognition Using Emerging Mixed Convolutional and Adaptive Residual Network Approach," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 7, pp. 2242–2255, 2021.

[159] J. Jayanthi, E. L. Lydia, N. Krishnaraj, T. Jayasankar, R. L. Babu, and R. A. Suji, "An effective deep learning features based integrated framework for iris detection and recognition," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 3271–3281, 2021.

[160] S. H. Shanto, M. N. Ali, and S. M. M. Ahsan, "An advanced cnn based iris recognition and segmentation for visible spectrum images," in *2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, pp. 1–5, IEEE, 2022.

[161] I. A. Hassan, S. A. Ali, and H. K. Obayes, "Iris recognition system based on efficient model for cnn features extraction and svm classifier," *Journal of Positive School Psychology*, pp. 3930–3939, 2022.

[162] J. E. Zambrano, D. P. Benalcazar, C. A. Perez, and K. W. Bowyer, "Iris recognition using low-level cnn layers without training and single matching," *IEEE Access*, vol. 10, pp. 41276–41286, 2022.

[163] L. Jia, X. Shi, Q. Sun, X. Tang, and P. Li, "Second-order convolutional networks for iris recognition," *Applied Intelligence*, pp. 1–15, 2022.

[164] P. Patil and K. Manikrao, "Safe and optimized iris recognition system using deep learning and gabor transform based approach," *J. Gr. Eng.*, vol. 10, pp. 8627–8642, 2020.

[165] W. Zhang, X. Lu, Y. Gu, Y. Liu, X. Meng, and J. Li, "A robust iris segmentation scheme based on improved U-net," *IEEE Access*, vol. 7, pp. 85082–85089, 2019.

[166] A. Alsubari and R. J. Ramteke, "MULTIMODAL OF FACE AND IRIS BASED ON LOCAL BINARY PATTERN AND GABOR-ZERNIKE MOMENTS.," *International Journal of Advanced Research in Computer Science*, vol. 9, no. 1, 2018.

[167] L. Mansoura, A. Noureddine, O. Assas, and A. Yassine, "Biometric recognition by multimodal face and iris using FFT and SVD methods With Adaptive Score Normalization," in *2019 4th World Conference on Complex Systems (WCCS)*, pp. 1–5, IEEE, 2019.

[168] A. S. Al-Waisy, R. Qahwaji, S. Ipson, S. Al-Fahdawi, and T. A. M. Nagem, "A multi-biometric iris recognition system based on a deep learning approach," *Pattern Analysis and Applications*, vol. 21, no. 3, pp. 783–802, 2018.

[169] 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*, vol. 20, no. 19, p. 5523, 2020.

[170] N. Alay and H. H. Al-Baity, "A multimodal biometric system for personal verification based on different level fusion of iris and face traits," *Biosci. Biotechnol. Res. Commun*, vol. 12, pp. 565–576, 2019.

[171] S. Soleymani, A. Dabouei, H. Kazemi, J. Dawson, and N. M. Nasrabadi, "Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification," in *2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 3469–3476, IEEE, 2018.

[172] 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)*, pp. 763–767, IEEE, 2018.

[173] K. Xiao, Y. Tian, Y. Lu, Y. Lai, and X. Wang, "Quality assessment-based iris and face fusion recognition with dynamic weight," *The Visual Computer*, vol. 38, no. 5, pp. 1631–1643, 2022.

[174] S. S. Harakannanavar and Others, "Performance Evaluation of Feature Level Fusion for Multimodal Biometric Systems," *Mathematical Statistician and Engineering Applications*, vol. 71, no. 4, pp. 2775–2792, 2022.

[175] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[176] S. A. Kalogirou, "Applications of artificial neural networks in energy systems," *Energy Conversion and Management*, vol. 40, no. 10, pp. 1073–1087, 1999.

[177] D. Zhang, *Artificial Neural Network*, pp. 229–270. Cham: Springer International Publishing, 2021.

[178] F. Bre, J. M. Gimenez, and V. D. Fachinotti, "Prediction of wind pressure coefficients on building surfaces using artificial neural networks," *Energy and Buildings*, vol. 158, pp. 1429–1441, 2018.

[179] H. Dao, "Image classification using convolutional neural networks," bachelor's thesis, Oulu University of Applied Sciences, 2020.

[180] J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei-Fei, "Imagenet large scale visual recognition competition 2012 (ILSVRC2012)," *See net. org/challenges/LSVRC*, vol. 41, 2012.

[181] "Introduction to Deep Learning: What Are Convolutional Neural Networks? Video - MATLAB." https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html.

[182] "Student Notes: Convolutional Neural Networks (CNN) Introduction – Belajar Pembelajaran Mesin Indonesia." https://indoml.com/2018/03/07/student-notes-convolutional-neural-networks-cnn-introduction/.

[183] I. Kandel and M. Castelli, "Transfer learning with convolutional neural networks for diabetic retinopathy image classification. A review," *Applied Sciences*, vol. 10, no. 6, p. 2021, 2020.

[184] S. Ioffe, "Batch renormalization: Towards reducing minibatch dependence in batch-normalized models," *Advances in neural information processing systems*, vol. 30, 2017.

[185] F. Yang, W. Li, B. Liang, S. Han, and X. Zhu, "Multi-stage attention network for video-based person re-identification," *IET Computer Vision*, vol. 16, no. 5, pp. 445–455, 2022.

[186] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *32nd International Conference on Machine Learning, ICML 2015*, vol. 1, pp. 448–456, PMLR, 2015.

[187] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

[188] L. Sifre, *Rigid-Motion Scattering For ImageClassification*. PhD thesis, Ecole Polytechnique, CMAP, 2014.

[189] S. Saponara, A. Elhanashi, and Q. Zheng, "Developing a real-time social distancing detection system based on YOLOv4-tiny and bird-eye view for COVID-19," *Journal of Real-Time Image Processing*, vol. 19, no. 3, pp. 551–563, 2022.

[190] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263–7271, 2017.

[191] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[192] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[193] L. R. Heinsius, "Real-Time YOLOv4 FPGA Design with Catapult High-Level Synthesis," Master's thesis, University of Twente, 2021.

[194] W. Zeng, J. Guo, L. Hao, J. Liu, and C. Wang, "Evaluation of Physical Electrical Experiment Operation Process Based on YOLOv5 and ResNeXt Cascade Networks," *Neural Processing Letters*, pp. 1–21, 2022.

[195] J. Yao, D. Cai, X. Fan, and B. Li, "Improving YOLOv4-Tiny's Construction Machinery and Material Identification Method by Incorporating Attention Mechanism," *Mathematics*, vol. 10, no. 9, p. 1453, 2022.

[196] Z. Jiang, L. Zhao, S. Li, and Y. Jia, "Real-time object detection method based on improved YOLOv4-tiny," *arXiv preprint arXiv:2011.04244*, 2020.

[197] X. Wang, K. Wang, and S. Lian, "A survey on face data augmentation for the training of deep neural networks," *Neural computing and applications*, vol. 32, no. 19, pp. 15503–15531, 2020.

[198] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[199] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pp. 740–755, Springer, 2014.

[200] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 806–813, 2014.

[201] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," *Advances in neural information processing systems*, vol. 27, 2014.

[202] A. Hattab and A. Behloul, "A Robust Face Recognition Method Based on ALTP and SIFT," in *Advances in Communication Technology, Computing and Engineering*, pp. 155–169, RGN Publications, 2021.

[203] A. Hattab and A. Behloul, "New approaches for automatic face recognition based on deep learning models and local handcrafted altp," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 9, no. 34, pp. e11–e11, 2022.

[204] A. Hattab and A. Behloul, "An illumination-robust face recognition approach based on convolutional neural network," in *Modelling and Implementation of Complex Systems: Proceedings of the 7th International Symposium, MISC 2022, Mostaganem, Algeria, October 30-31, 2022*, pp. 135–149, Springer, 2022.

[205] A. Khurshid, S. C. Tamayo, E. Fernandes, M. R. Gadelha, and M. Teofilo, "A robust and real-time face anti-spoofing method based on texture feature analysis," in *HCI International 2019–Late Breaking Papers: 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21*, pp. 484–496, Springer, 2019.

[206] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and vision computing*, vol. 16, no. 5, pp. 295–306, 1998.

[207] Y. LeCun, Y. Bengio, G. Hinton, and Others, "Deep learning. nature 521 (7553), 436-444," *Google Scholar Google Scholar Cross Ref Cross Ref*, 2015.

[208] S. Almabdy and L. Elrefaei, "Deep convolutional neural network-based approaches for face recognition," *Applied Sciences*, vol. 9, no. 20, p. 4397, 2019.

[209] K. Yang, J. Yau, L. Fei-Fei, J. Deng, and O. Russakovsky, "A Study of Face Obfuscation in ImageNet," *arXiv preprint arXiv:2103.06191*, 2021.

[210] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[211] A. Martinez and R. Benavente, "The AR face database: CVC technical report, 24," 1998.

[212] F. Alonso-Fernandez, P. Tome-Gonzalez, V. Ruiz-Albacete, and J. Ortega-Garcia, "Iris recognition based on sift features," in *2009 First IEEE International Conference on Biometrics, Identity and Security (BIdS)*, pp. 1–8, IEEE, 2009.

[213] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan, "Iris recognition with off-the-shelf CNN features: A deep learning perspective," *IEEE Access*, vol. 6, pp. 18848–18855, 2017.

[214] K. Y. Shin, G. P. Nam, D. S. Jeong, B. J. Kang, K. R. Park, J. Kim, *et al.*, "New iris recognition method for noisy iris images," *Pattern Recognition Letters*, vol. 33, no. 8, pp. 991–999, 2012.

[215] S. S. Arora, M. Vatsa, R. Singh, and A. Jain, "On iris camera interoperability," in *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 346–352, IEEE, 2012.

[216] A. HATTAB and A. BEHLOUL, "A robust iris recognition approach based on transfer learning," *International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 1065–1080, 2023.

[217] A. Shubha Rao and K. Mahantesh, "Image classification based on inception-v3 and a mixture of handcrafted features," in *Distributed Computing and Optimization Techniques: Select Proceedings of ICDCOT 2021*, pp. 527–537, Springer, 2022.

[218] Z. Zhou, X. Yang, J. Ji, Y. Wang, and Z. Zhu, "Classifying fabric defects with evolving inception v3 by improved l2, 1-norm regularized extreme learning machine," *Textile Research Journal*, p. 00405175221114633, 2022.

[219] H. M. Ahmed, M. Javed Awan, N. S. Khan, A. Yasin, and H. M. Faisal Shehzad, "Sentiment analysis of online food reviews using big data analytics," *Hafiz Muhammad Ahmed, Mazhar Javed Awan, Nabeel Sabir Khan, Awais Yasin, Hafiz Muhammad Faisal Shehzad (2021) Sentiment Analysis of Online Food Reviews using Big Data Analytics. Elementary Education Online*, vol. 20, no. 2, pp. 827–836, 2021.

[220] M. Qorib, T. Oladunni, M. Denis, E. Ososanya, and P. Cotae, "Covid-19 vaccine hesitancy: Text mining, sentiment analysis and machine learning on covid-19 vaccination twitter dataset," *Expert Systems with Applications*, vol. 212, p. 118715, 2023.

[221] B. Talafha, W. Farhan, A. Altakrouri, and H. Al-Natsheh, "Mawdoo3 ai at madar shared task: Arabic tweet dialect identification," in *Proceedings of the Fourth Arabic Natural Language Processing Workshop*, pp. 239–243, 2019.

[222] F. Chollet and Others, "Keras: The python deep learning library," *Astrophysics source code library*, pp. ascl—-1806, 2018.

[223] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," *NIPS Autodiff workshop*, 2017.

[224] Z. Wang, J. Dong, F. Zhu, and J. Song, "Generalized locality preserving projection for multimodal biometric recognition.," *J. Inf. Hiding Multim. Signal Process.*, vol. 9, no. 4, pp. 853–863, 2018.

[225] H. F. Liau and D. Isa, "Feature selection for support vector machine-based face-iris multimodal biometric system," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11105–11111, 2011.

[226] X. Tu, Z. Ma, J. Zhao, G. Du, M. Xie, and J. Feng, "Learning generalizable and identity-discriminative representations for face anti-spoofing," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 11, no. 5, pp. 1–19, 2020.

[227] A. Hattab and A. Behloul, "Face-iris multimodal biometric recognition system based on deep learning," *Multimedia Tools and Applications*, pp. 1–28, 2023.

[228] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep learning earth observation classification using ImageNet pretrained networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 105–109, 2015.

[229] Y. Yin, L. Liu, and X. Sun, "SDUMLA-HMT: a multimodal biometric database," in *Chinese Conference on Biometric Recognition*, pp. 260–268, Springer, 2011.

[230] S. Deepak and P. M. Ameer, "Automated categorization of brain tumor from mri using cnn features and svm," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 8, pp. 8357–8369, 2021.

[231] H. Proença and L. A. Alexandre, "UBIRIS: A noisy iris image database," in *Proceed. of ICIAP 2005 - Intern. Confer. on Image Analysis and Processing*, vol. 1, pp. 970–977, 2005.