People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Mohamed El Bachir El Ibrahimi University of Bordj Bou Arréridj

Thesis

Presented to the Faculty of Mathematics and Computer Science
Department of Computer Science
to obtain the degree of

Phd of Science

Specialty: Computer Science

Defended by
Sonia Benabid

THEME

A Biometric System Based on Deep Learning Techniques

Publicly defended on: 18/06/2026

Jury

Pr.	Nouioua Farid	University of Bordj Bou Arréridj	President
Dr.	Maza Sofiane	University of Bordj Bou Arréridj	Supervisor
Pr.	Attia Abdelouahab	University of Bordj Bou Arréridj	Co-Supervisor
Pr.	Khababa Abdallah	University of Sétif 1	Examiner
Pr.	Akrouf Samir	University of Msila	Examiner
Dr.	Belhadj Foudil	University of Bordj Bou Arréridj	Examiner

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed El Bachir El Ibrahimi de Bordj Bou Arréridj

Thèse

Présentée à la Faculté des Mathématiques et de l'Informatique
Département d'Informatique
en vue de l'obtention du diplôme de

Doctorat en Sciences

Spécialité : Informatique

Soutenue par
Sonia Benabid

THÈME

Un système biométrique basé sur des techniques d'apprentissage profond

---

Soutenue publiquement le : 18/06/2026

Jury

Pr.	Nouioua Farid	Université de Bordj Bou Arréridj	Président
Dr.	Maza Sofiane	Université de Bordj Bou Arréridj	Directeur de thèse
Pr.	Attia Abdelouahab	Université de Bordj Bou Arréridj	Co-Directeur de thèse
Pr.	Abdallah Khababa	Université de Sétif 1	Examinateur
Pr.	Samir Akrouf	Université de M'sila	Examinateur
Dr.	Belhadj Foudil	Université de Bordj Bou Arréridj	Examinateur

Acknowledgment

First and foremost, I give praise and gratitude to Allah-Alhamdulillah-for granting me the strength, patience, and guidance to complete this work and to reach this important milestone in my academic journey.

I am sincerely grateful to my supervisor, Dr. Sofiane Maza, for his steady guidance, valuable advice, and encouragement throughout the development of this thesis. His support and insightful feedback were instrumental in bringing this work to completion. I also want to thank my co-supervisor, Pr. Abdelouahab Attia, for his support and thoughtful suggestions that helped improve the quality of this research.

I would also like to express my sincere gratitude to Mr. Nadjib BENAOUDA, Mr. Salah MAACHE, Mrs. Zineb MAAREF, and Mrs. Nour El Houda CHALABI for their support, guidance, and cooperation throughout this project, which greatly contributed to its successful completion.

I am also grateful to all the teachers and researchers who contributed to my academic development and shared their knowledge and expertise throughout my studies.

My deepest thanks go to my family for their support and encouragement throughout this journey. Their belief in me has been a constant source of motivation and strength.

Finally, I would like to thank everyone who contributed, directly or indirectly, to the completion of this thesis. Their support and encouragement are sincerely appreciated.

Sonia

Table of Contents

List of Abbreviations

DL Deep Learning

ML Machine Learning

FKP Finger Knuckle Print

SQI Self-Quotient Image

xDNN Explainable Deep Neural Network

PCANet Principal Component Analysis Network

XAI Explainable Artificial Intelligence

FMR False Match Rate

FNMR False Non-Match Rate

ROC Receiver Operating Characteristic

EER Equal Error Rate

VR Verification Rate

FAR False Acceptance Rate

CMC Cumulative Match Characteristic

MSE Mean Squared Error

TPIR True Positive Identification Rate

FPIR False Positive Identification Rate

ANNs Artificial Neural Networks

CNNs Convolutional Neural Networks

RNNs Recurrent Neural Networks

GANs Generative Neural Networks

FNNs Feed-forward Neural Networks

AEs Autoencoders

CV Computer Vision

NLP Natural Language Processing

ECG Electrocardiograms

BCG ballistocardiograms

LSTM Long Short-Term Memory

GRUs Gated Recurrent Units

ROI Region of Interest

LIF Left Index Finger

LMF Left Middle Finger

RIF Right Index Finger

RMF Right Middle Finger

HP Hierarchical Prototype

HPL Hierarchical Prototype Layer

ASM Adaptive Similarity Metric

CFM Confidence Fusion Module

NN Nearest Neighbors

KNN K-Nearest Neighbors

LoG Laplacian of Gaussian

List of Figures

List of Tables

List of Algorithms

Chapter 1

General Introduction

1.1 Background and context

The way we prove who we are has changed profoundly in recent years. Not long ago, presenting an identity card, typing a password, or remembering a PIN code was enough to access a building, log into a computer, or authorize a financial transaction. These methods served us reasonably well for decades, but their limitations have become increasingly difficult to ignore. Passwords can be forgotten, stolen, or guessed. Cards can be lost, duplicated, or left at home. Perhaps most concerning, these credentials only verify that someone possesses the right information or object they cannot confirm that the person presenting them is actually who they claim to be.

This fundamental vulnerability has driven the search for authentication methods that are more directly tied to the individual. Biometric recognition the automated identification of people based on their physiological or behavioral characteristics offers a compelling alternative. By anchoring identity verification in the physical reality of the human body, biometric systems promise a level of security that traditional methods cannot match. Our fingerprints, irises, facial features, and even the way we walk are inherently part of who we are. They cannot be easily forgotten, and they require our physical presence at the moment of authentication.

Among the various biometric traits that researchers have explored, the Finger Knuckle Print has emerged as particularly promising. The skin patterns that surround our finger knuckles form complex and individually distinctive textures. These patterns remain relatively stable

throughout adult life and can be captured using simple, low-cost imaging devices. Unlike fingerprints, which we leave behind on virtually every surface we touch, knuckle patterns are less readily captured surreptitiously, offering enhanced privacy in certain applications. For these reasons, FKP has attracted growing attention from the biometric research community.

Yet despite these advantages, FKP recognition systems face two significant challenges that have limited their practical deployment. The first concerns image quality. When FKP images are captured under different lighting conditions varying between indoor and outdoor environments, different times of day, or simply changes in ambient illumination the appearance of the knuckle patterns can change dramatically. These illumination variations can mask the genuine biometric features and introduce false differences between images of the same person, leading to recognition errors.

The second challenge relates to how we understand and trust the decisions made by modern recognition systems. The deep learning (DL) approaches that have revolutionized biometric accuracy in recent years operate largely as "black boxes." They can tell us with impressive confidence that two images belong to the same person, but they cannot explain why. In security-critical applications border control, law enforcement, financial services this lack of transparency is deeply problematic. Users and operators need to understand the reasoning behind authentication decisions. Regulators demand accountability. When something goes wrong, we need to be able to trace how and why the error occurred.

This thesis addresses both challenges through a unified framework that combines robust feature extraction with transparent decision-making. At its core, the proposed approach integrates three key components. The Self-Quotient Image algorithm decomposes FKP images into illumination-invariant components, effectively separating the intrinsic knuckle patterns from the lighting conditions under which they were captured. A two-stage PCANet then extracts hierarchical features from these normalized images, capturing discriminative patterns efficiently without requiring massive amounts of training data. Finally, an explainable Deep Neural Network performs classification through prototype-based reasoning, generating decisions that can be traced back to specific learned patterns and understood by human users.

The work presented in this thesis has been validated through rigorous experimentation on the PolyU FKP database, a standard benchmark containing images from four finger types: left

index, left middle, right index, and right middle. Through systematic optimization of key parameters, the proposed framework achieves recognition rates between 91% and 96% across different modalities, outperforming existing methods while providing the interpretability that conventional approaches lack. Beyond the technical contributions, this work represents a step toward biometric systems that are not only accurate but also trustworthy systems that can explain themselves to the people who rely on them.

1.2 Problem statement

Despite significant advances in FKP recognition over the past decade, existing systems continue to struggle with two fundamental limitations that constrain their practical deployment in security-sensitive applications.

The first limitation concerns robustness to real-world imaging conditions. Most existing FKP recognition methods assume relatively controlled capture environments with consistent illumination. When deployed in practical settings where lighting varies naturally between indoor and outdoor environments, changes throughout the day, or differs between enrollment and recognition sessions these systems experience significant performance degradation. The biometric features extracted from images captured under different lighting conditions can appear more different than images from different individuals captured under the same conditions. This sensitivity to illumination undermines the reliability of FKP recognition in precisely the real-world scenarios where biometric authentication is most needed.

The second limitation relates to interpretability. The DL models that currently achieve the highest recognition accuracy operate through complex, non-linear transformations that obscure the reasoning behind their decisions. When such a system correctly identifies an individual, we cannot understand what features it used or why it was confident. When it makes an error, we cannot trace the source of the mistake. This opacity is unacceptable in security-critical applications, where decisions have real consequences and accountability is essential. Users must trust that the system is making decisions for the right reasons. Regulators require explanations for automated decisions that affect individuals. System developers need to understand failure modes to improve their designs.

These two problems, illumination sensitivity and lack of interpretability, have been ad-

dressed separately in the literature, but rarely together. Systems that achieve robustness to lighting variations typically remain black boxes. Interpretable systems often sacrifice accuracy or fail to handle real-world imaging conditions. What is needed is a unified approach that simultaneously addresses both challenges, providing accurate recognition under varying illumination while maintaining complete transparency in decision-making.

1.3 Objectives

This thesis aims to develop and validate a comprehensive framework for FKP recognition that achieves both high accuracy under challenging imaging conditions and full interpretability of classification decisions. The specific objectives that guide this work are as follows:

- **Objective 1: To develop an illumination-invariant preprocessing method for FKP images.** The first objective is to design and implement a preprocessing technique that effectively separates intrinsic knuckle patterns from the lighting conditions under which images are captured. This method should normalize FKP images so that subsequent feature extraction focuses on genuine biometric characteristics rather than artifacts introduced by illumination variations.

- **Objective 2: To implement a hierarchical feature extraction framework that captures discriminative knuckle patterns efficiently.** The second objective involves designing a feature extraction pipeline based on PCANet that learns hierarchical representations directly from normalized FKP images. This framework should capture both local texture details and global structural patterns while maintaining computational efficiency and avoiding the need for massive training datasets.

- **Objective 3: To integrate an explainable classifier that provides transparent decision-making.** The third objective is to incorporate the xDNN classifier into the recognition pipeline, enabling prototype-based reasoning that can be understood and verified by human users. This classifier should generate decisions that can be traced back to specific learned patterns, providing both local explanations for individual classifications and global insights into how the system represents different individuals.

- **Objective 4: To optimize system parameters through systematic ablation studies.**

The fourth objective involves conducting comprehensive experiments to identify optimal configurations for both the preprocessing algorithm and the feature extraction framework. These studies should determine how different parameter choices affect recognition performance and identify settings that provide the best trade-offs between accuracy, robustness, and efficiency.

- **Objective 5: To validate the proposed framework against state-of-the-art methods.** The fifth objective is to evaluate the complete system on standard benchmark databases, comparing its performance against existing FKP recognition approaches. This validation should demonstrate not only competitive or superior accuracy but also the unique advantages of interpretable decision-making.

- **Objective 6: To explore extensions of the proposed framework through complementary preprocessing and feature learning strategies.** This objective focuses on investigating the potential of incorporating edge-based enhancement techniques and diverse feature extraction approaches to improve the representation of FKP images. It also aims to assess the effectiveness of alternative classification strategies in enhancing recognition robustness and overall system performance.

1.4 Contributions

This thesis makes several original contributions to the field of biometric recognition, particularly in the area of finger knuckle print recognition with explainable DL. The main contributions can be summarized as follows:

- **Contribution 1: A unified framework for illumination-invariant and explainable FKP recognition.** The primary contribution of this work is the development of a complete recognition pipeline that simultaneously addresses the two major limitations of existing systems: sensitivity to illumination variations and lack of interpretability. By combining SQI preprocessing, PCANet feature extraction, and xDNN classification, the proposed framework achieves robust recognition under varying lighting conditions while maintaining full transparency in decision-making. This integration represents a novel approach that has not been previously explored in the FKP recognition literature.

- **Contribution 2: Systematic optimization of SQI and PCANet parameters for FKP recognition.** Through comprehensive ablation studies across four finger modalities, this work identifies optimal configurations for both the illumination normalization algorithm and the feature extraction framework. These parameter studies provide valuable insights into how different choices affect recognition performance and establish reference configurations that can guide future research. The results demonstrate that careful parameter tuning can yield significant improvements in both identification accuracy and verification performance.

- **Contribution 3: Demonstration of prototype-based explainability for biometric decisions.** This thesis provides the first detailed demonstration of how xDNN's prototype-based reasoning can be applied to FKP recognition, generating interpretable IF-THEN rules that link classification decisions to specific learned patterns. The visualization of these prototypes as image patches enables human experts to understand and verify the reasoning behind the system's decisions, addressing the critical need for transparency in security-sensitive applications.

- **Contribution 4: Comparative evaluation against state-of-the-art methods.** The proposed framework is rigorously evaluated against existing FKP recognition approaches, demonstrating competitive or superior performance across all four finger modalities. For middle finger modalities, where illumination variations are particularly pronounced, the system achieves recognition rates of 95.86% and 95.96%, outperforming all compared methods. These results validate the effectiveness of the proposed approach while establishing new benchmarks for explainable FKP recognition.

- **Contribution 5: Extension through edge-enhanced deep feature extraction and prototype based classification.** This contribution extends the basic framework by investigating edge detection techniques as complementary preprocessing methods to enhance feature discriminability. Operators such as Canny, Sobel, and Laplacian of Gaussian are studied to improve the representation of structural information and boost recognition performance. In addition, the effectiveness of hierarchical prototype-based classification is analyzed in comparison with conventional classifiers, highlighting its advantages for biometric recognition tasks.

- **Contribution 6: Published validation and extended experimental evaluation of the proposed framework.** The core ideas and experimental results presented in Chapter 4 have been validated through peer review and published in the journal Ingénierie des Systèmes d'Information (ISI), Volume 30, Number 9, September 2025. The article, titled "An Explainable Deep Neural Network Based PCANet Framework with Illumination-Invariant Features for Finger Knuckle Print Recognition System," confirms the novelty and significance of the proposed approach within the research community. Furthermore, an extended version of this work presented in Chapter 5 has been presented in an IEEE conference paper, introducing the EDHP-FKP framework. This extension incorporates edge-enhanced preprocessing using Canny, Sobel, and Laplacian of Gaussian operators, evaluates multiple deep feature extractors (VGG-16, AlexNet, ResNet-50) alongside Gabor descriptors, and integrates a hierarchical prototype (HP) classifier with comparative analysis against the Nearest Neighbor (NN) method.

Together, these contributions advance the state of the art in FKP recognition while addressing the critical need for interpretability in biometric systems. By demonstrating that high accuracy and full transparency are not mutually exclusive goals, this work lays the foundation for a new generation of biometric systems that can be both trusted and understood by the people who rely on them.

1.5 Thesis organization

This thesis is organized into six chapters, each building upon the foundations established in the preceding sections.

Chapter 2: Fundamentals of Biometrics provides the essential background for understanding biometric recognition systems. It introduces the criteria that distinguish useful biometric characteristics from those unsuitable for recognition applications, describes the standard architecture of biometric systems, and explains the performance metrics used to evaluate them. The chapter concludes with a survey of major biometric modalities, with particular attention to those most relevant to this work: finger knuckle patterns, face, and palmprint.

Chapter 3: Biometric Systems Based on Deep Learning presents the theoretical foundations of DL as applied to biometric recognition. It begins with a brief overview of artificial

neural networks and then progresses to deep architectures, including convolutional neural networks, recurrent neural networks, autoencoders, and generative adversarial networks. Special attention is given to lightweight and interpretable models PCANet and xDNN that form the basis of the proposed framework. The chapter concludes with a comprehensive review of recent work applying DL to biometric systems.

Chapter 4: Explainable xDNN-PCANet with Illumination-Invariant Features for FKP Recognition presents the first major contribution of this thesis. It describes in detail the proposed framework, including SQI-based illumination normalization, two-stage PCANet feature extraction, and xDNN classification. The chapter reports extensive experimental results on the PolyU FKP database, demonstrating the system's performance across four finger modalities and comparing it with state-of-the-art methods. The explainability of the xDNN classifier is illustrated through prototype-based decision rules.

Chapter 5: Hierarchical Prototype Learning with Edge-Enhanced Features for FKP extends the work of the previous chapter by investigating the integration of edge detection techniques with deep feature extraction. It explores how Canny, Sobel, and Laplacian of Gaussian operators can enhance discriminative patterns before feature extraction, and compares the Hierarchical Prototype classifier with conventional approaches. Experimental results demonstrate the effectiveness of this extended framework.

Chapter 6: General Conclusion summarizes the main contributions of this thesis, discusses their implications for the field of biometric recognition, and outlines promising directions for future research. It reflects on the broader significance of explainable biometric systems and their potential impact on security-critical applications.

1.6 List of publications

International Journal

1. Benabid, S., Maza, S., Attia, A., Chalabi, N.E., Chahir, Y. (2025). An explainable deep neural network based PCANet framework with illumination-invariant features for finger knuckle print recognition system. Ingénierie des Systèmes d'Information, Vol. 30, No. 9, pp. 2347-2364. https://doi.org/10.18280/isi.300912

International Conferences

1. Benabid, S., Maza, S., Maaref, Z., Attia, A., Ali, W. (2026). An Edge-Enhanced Deep Feature and Hierarchical Prototype Framework for Finger Knuckle Print Recognition. In: Proceedings of the 8th International Conference on Pattern Analysis and Intelligent Systems (PAIS), IEEE (to be published).

Chapter 2

Fundamentals of Biometrics

2.1 Introduction

This chapter establishes the conceptual framework and technical vocabulary necessary for understanding how biometric systems function, how their performance is measured, and what makes certain characteristics suitable for biometric recognition. It addresses biometric fundamentals, beginning with the criteria that distinguish useful biometric characteristics from those unsuitable for recognition applications. It proceeds to describe the architecture of biometric systems, tracing the flow of information from sensor capture through preprocessing, feature extraction, matching, and decision-making. A detailed discussion of performance metrics follows, explaining how equal error rates, rank-one recognition rates, and related measures enable objective comparison of different systems and modalities. The section concludes with a survey of major biometric modalities, with particular attention to those most relevant to this work: finger knuckle patterns, face, and palmprint in its two-dimensional, three-dimensional, and multispectral variants.

2.2 Overview

The question of personal identity has occupied human societies since their earliest formations. In small communities, identity was established through direct personal acquaintance; individuals knew their neighbors, their leaders, and their traders by sight and by reputation. As societies expanded and interactions became increasingly mediated by technology, this simple

model of identity verification proved inadequate. The rise of global travel, electronic commerce, and digital communication created an urgent need for reliable methods of establishing identity at a distance, without the benefit of personal familiarity.

Traditional approaches to identity verification fell into two broad categories: token-based methods, which relied on "something you have," and knowledge-based methods, which relied on "something you know" [1]. Physical keys, identification cards, and passports exemplify the first category; passwords, personal identification numbers, and answers to security questions exemplify the second. Both approaches have served society well for decades, yet both suffer from fundamental vulnerabilities that have become increasingly apparent in an era of sophisticated cyber threats. Tokens can be lost, stolen, or duplicated; knowledge can be forgotten, guessed, or surreptitiously observed. Perhaps most critically, both methods verify only the correctness of the credential, not the legitimacy of the individual presenting it [2].

These limitations have driven sustained interest in a third approach: biometric recognition, which bases identity verification on "something you are." Biometric systems operate by measuring and analyzing physiological or behavioral characteristics that are inherently tied to each individual. Unlike passwords or tokens, biometric traits cannot be easily forgotten, lost, or shared. They require the physical presence of the individual at the time of authentication, and they offer the possibility of binding identity claims directly to the person rather than to external credentials [3]. The formal definition of biometrics encompasses both a scientific discipline and a set of practical technologies. Biometrics, as a scientific subject, focuses on the measurement and application of an individual's qualities or features for unique identification. [4]. As a technology domain, biometrics encompasses the sensors, algorithms, and systems that implement this identification capability in practical applications ranging from mobile phone unlocking to international border control.

2.3 Biometric characteristics

Biometric characteristics constitute the foundation of identity recognition systems, relying on measurable physiological and behavioral traits to distinguish individuals. This section outlines the essential criteria that define a reliable biometric characteristic and presents the main categories of biometric modalities, highlighting the distinction between physiological and be-

havioral traits.

2.3.1 Criteria for biometric characteristics

Not every human characteristic is suitable for use in biometric recognition. For a trait to serve as a useful biometric modality, it must satisfy a set of criteria that determine its practical utility across different applications. These criteria, which have been refined through decades of research and practical experience, provide a framework for evaluating and comparing different biometric traits [1]. These criteria can be summarized as follows:

- **Universality:** means that a biometric trait should be present in all individuals. In reality, this is rarely fully achieved, as some people may lack or have altered traits due to conditions, injuries, or other factors. Therefore, biometric systems must account for such cases by providing alternative enrollment or authentication methods [5].

- **Uniqueness:** refers to the ability of a biometric trait to distinguish between individuals. For effective recognition, differences between individuals must be greater than variations within the same individual. Although no trait is perfectly unique, it should provide sufficient discriminatory power for the intended application [6].

- **Permanence:** refers to the stability of a biometric trait over time. An effective trait should remain relatively consistent, though some may change due to aging, health, or environmental factors. Systems must therefore account for possible variations, especially for traits that are more prone to change.

- **Collectability:** refers to how easily and reliably a biometric trait can be captured. Effective systems require data acquisition that is convenient, non-intrusive, and capable of producing consistent quality under varying conditions [7].

- **Performance:** refers to the accuracy, speed, and robustness of a biometric system. It must meet the requirements of its application while maintaining reliable operation under varying conditions [3].

- **Acceptability:** refers to how willing users are to use a biometric system. Even highly accurate systems may fail if they are perceived as intrusive or uncomfortable, so user convenience and comfort are essential for successful adoption.

- **Circumvention:** refers to how resistant a biometric system is to spoofing or fraudulent attacks. A robust system should make it difficult to use fake or manipulated biometric samples, often by incorporating mechanisms to detect genuine (live) inputs [7].

2.3.2 Categories of biometric characteristics

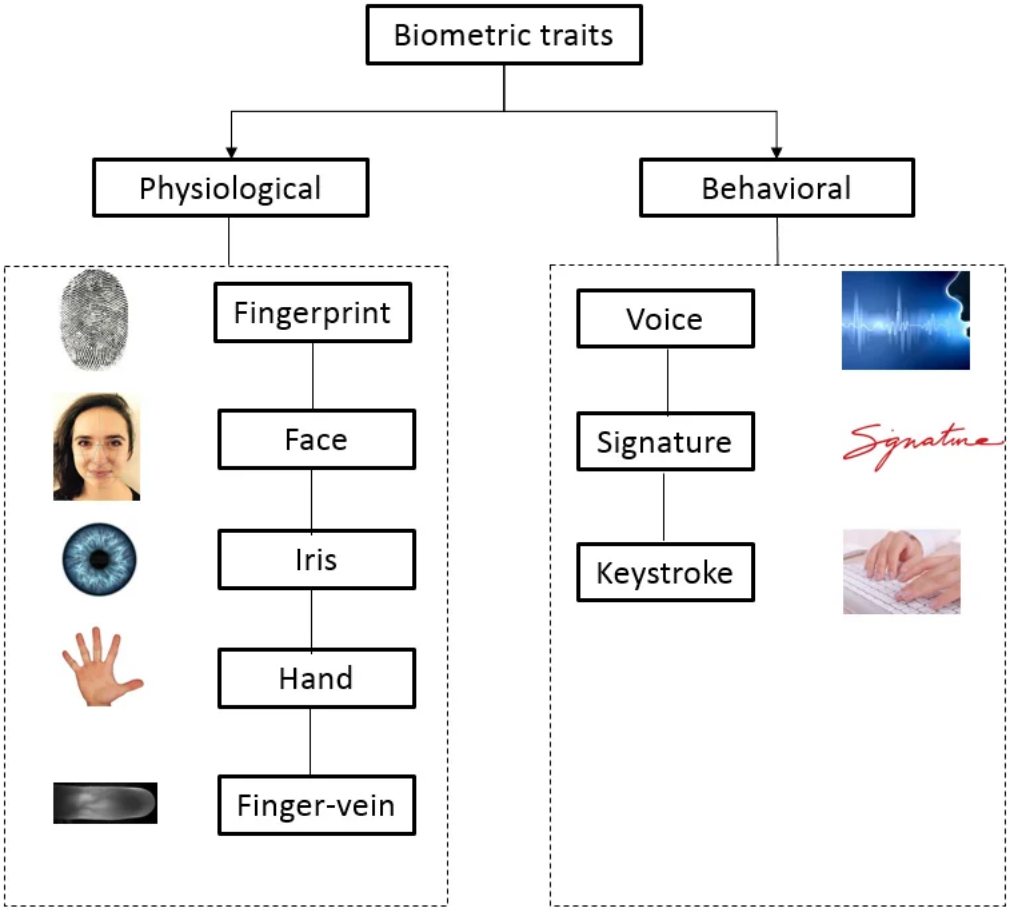


Figure 2.1: Categories of biometric characteristics [8].

Biometric characteristics are conventionally divided into two broad categories based on their origin as depicted in Figure 2.1:

- **Physiological biometrics:** derive from direct measurements of the human body's physical structures. These include fingerprints, facial features, iris patterns, retinal vasculature, palmprint, hand geometry, and DNA. Physiological traits tend to exhibit relatively high permanence and are generally less susceptible to short-term variation than behavioral traits. They typically require specialized sensors for capture but offer the advantage of being largely independent of user state or behavior [3].

- **Behavioral biometrics:** describe patterns in human activity that are characteristic of specific individuals. These include voice patterns, signature dynamics, keystroke rhythms, gait, and more subtle behavioral signatures derived from interaction with devices. Hafs et al. [9] demonstrated that online handwritten signature verification using empirical mode decomposition could capture distinctive behavioral patterns. Behavioral traits offer the advantage of being capturable without specialized hardware keystroke dynamics, for example, can be measured using standard keyboards, and gait can be observed using ordinary video cameras. However, behavioral traits generally exhibit greater intra-class variation than physiological traits, as they are influenced by factors such as fatigue, emotional state, and environmental context.

The distinction between physiological and behavioral biometrics is not always sharp. Voice recognition, for instance, depends on both the physiological structure of the vocal tract and the behavioral patterns of speech production. Similarly, signature recognition encompasses both the static image of the written name and the dynamic process of its creation. Modern biometric systems increasingly exploit this hybrid nature, combining multiple sources of information to achieve robust recognition.

2.4 Biometric system architecture

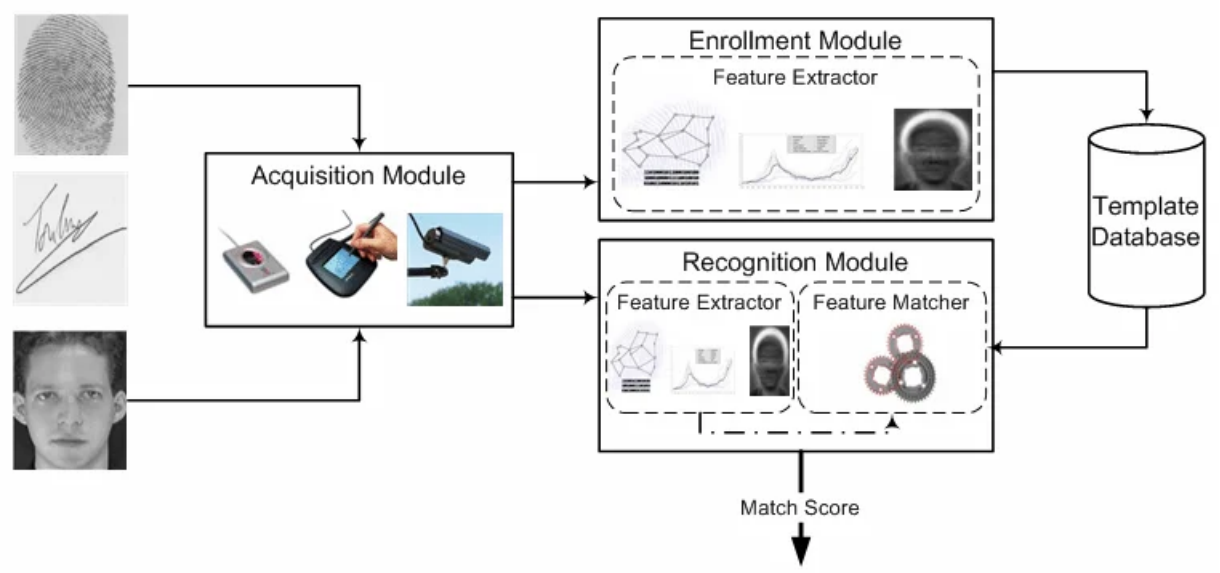


Figure 2.2: Basic architecture of a biometric system [10].

A biometric recognition system transforms raw biometric measurements into identity de-

cisions through a structured sequence of processing stages. Understanding this architecture is essential for comprehending how deep learning methods have been integrated into biometric systems and where they offer the greatest potential for performance improvement [5, 11].

Every biometric system operates in two distinct phases as depicted in Figure 2.2: enrollment and recognition. During **enrollment**, an individual's biometric data is captured, processed, and stored in the system's database as a reference template. This template serves as the canonical representation of that individual's identity against which future recognition attempts will be compared. Enrollment typically involves capturing multiple samples to account for intra-class variation and to ensure that the stored template adequately represents the range of possible presentations [1].

During **recognition**, a new biometric sample referred to as a probe is captured from an individual seeking authentication. This probe undergoes the same processing pipeline applied during enrollment, producing a query feature set that is compared against one or more stored templates. The outcome of this comparison informs a decision about the individual's identity.

The recognition phase itself encompasses two distinct operational modes that serve different application requirements. **Verification**, also called authentication, involves confirming that an individual is who they claim to be. In verification mode, the user provides an identity claim, typically through a username, card, or PIN, and the system performs a one-to-one comparison between the captured probe and the specific template corresponding to that claimed identity. Verification answers the question, "Am I who I claim to be?" and is the appropriate mode for applications such as unlocking a personal device or authorizing a financial transaction [3].

Identification, by contrast, involves determining an individual's identity without any prior claim. The system performs a one-to-many comparison between the captured probe and all templates stored in the database, returning either the identity of the closest match or an indication that the individual is not enrolled. Identification answers the question, "Who am I?" and is used in applications such as law enforcement, where a latent fingerprint from a crime scene must be matched against a database of known individuals, or in duplicate enrollment detection, where the system must ensure that the same person has not registered multiple times under different identities [3].

The architectural components that implement these functions follow a standardized pipeline

[11]: The **sensor module** acquires raw biometric data from the individual. The specific sensor depends on the modality: optical or capacitive sensors for fingerprints, cameras for face and iris, microphones for voice, and specialized imaging devices for modalities such as palmprint or finger vein. The sensor module must capture data with sufficient quality to support reliable recognition while maintaining user convenience and acceptability. Practical considerations include resolution, capture speed, robustness to environmental conditions, and resistance to spoofing attempts [7].

The **preprocessing module** transforms raw sensor data into a form suitable for feature extraction. Preprocessing typically includes multiple operations tailored to the specific modality and sensor characteristics. Image-based modalities undergo operations such as segmentation to isolate the region of interest, normalization to correct for variations in position and orientation, enhancement to improve contrast and reduce noise, and quality assessment to reject samples that do not meet minimum standards. For fingerprint images, preprocessing might include ridge orientation estimation and frequency analysis; for face images, it might include alignment based on eye coordinates and illumination normalization. [7] demonstrated the importance of preprocessing for multispectral biometric systems, where alignment across spectral bands is critical for effective fusion.

The **feature extraction module** converts preprocessed biometric data into a compact mathematical representation that preserves discriminative information while discarding irrelevant variations. This representation, typically referred to as a template or feature vector is the core of the biometric system's recognition capability. Feature extraction must achieve two potentially conflicting objectives: the template must be sufficiently distinctive to enable accurate discrimination between different individuals, yet sufficiently invariant to accommodate the natural variations in how the same individual presents their biometric trait. Traditional feature extraction approaches relied on handcrafted algorithms designed to capture specific characteristics of each modality's minutiae points for fingerprints, textural patterns for iris, and geometric relationships for face. Bachay and Abdulameer [12] demonstrated that hybrid deep learning models combining autoencoders and Convolutional Neural Networks (CNNs) could achieve superior performance for palmprint authentication by learning robust feature representations directly from data.

The **template database** stores the enrolled templates along with associated identity in-

formation. The security of this database is critical, as compromised templates cannot be revoked and reissued like passwords—an individual has only ten fingerprints, two irises, and one face. Modern systems employ various protection mechanisms, including encryption, secure hardware storage, and cancelable biometric techniques that transform templates through non-invertible functions, enabling revocation if compromise occurs [3]. Dwivedi and Dey [13] proposed a novel hybrid score-level and decision-level fusion scheme for cancelable multi-biometric verification, demonstrating that template protection could be integrated with multi-modal fusion without sacrificing recognition accuracy.

The **matching module** evaluates a probe feature set against one or several stored templates to get similarity scores. For verification, the probe is compared with the singular template associated with the asserted identity; for identification, it is compared with all templates in the database. The matching algorithm produces a score indicating the degree of similarity between the probe and each template, with higher scores indicating greater similarity. The design of matching algorithms must account for the characteristics of the feature representation—some features naturally support distance-based matching, while others require more complex comparison strategies [11].

The **decision module** interprets the matching scores to produce an identity decision. For verification, the decision module compares the similarity score against a threshold: if the score exceeds the threshold, the system accepts the claimed identity; otherwise, it rejects the claim. For identification, the decision module typically returns the identity corresponding to the highest similarity score, optionally with a confidence measure or a rejection option if the highest score falls below a minimum threshold. The choice of threshold involves trading off between different types of errors—accepting impostors versus rejecting genuine users—and must be calibrated based on application requirements [3].

2.5 Performance metrics

The evaluation of biometric systems requires standardized performance metrics that quantify the trade-off between security and convenience. These metrics provide an objective basis for comparing different systems, modalities, and operating conditions, and are essential for selecting appropriate configurations for specific applications [6, 11].

In a biometric verification system, a decision is made by comparing a probe sample against a stored template. This process may lead to two types of errors. A **false match** (also referred to as false acceptance) occurs when the system incorrectly accepts a probe from an impostor as a genuine match. A **false non-match** (also referred to as false rejection) occurs when the system incorrectly rejects a genuine user.

These errors are commonly quantified using the **False Acceptance Rate (FAR)** and the **False Rejection Rate (FRR)**, defined as:

$$\text{FAR} = \frac{\text{Number of false accepts}}{\text{Number of impostor attempts}}$$

$$\text{FRR} = \frac{\text{Number of false rejects}}{\text{Number of genuine attempts}}$$

In verification literature, the terms *False Match Rate* (FMR) and FAR are often used interchangeably, as are *False Non-Match Rate* (FNMR) and FRR, depending on the context of the evaluation protocol. These two types of errors are inherently linked: any reduction in FAR generally leads to an increase in FRR, and vice versa, depending on the decision threshold [3].

To further characterize system behavior, the **True Acceptance Rate (TAR)** and **True Rejection Rate (TRR)** are defined as:

$$\text{TAR} = 1 - \text{FRR}$$

$$\text{TRR} = 1 - \text{FAR}$$

The relationship between FAR and FRR across different decision thresholds is typically illustrated using the **Receiver Operating Characteristic (ROC)** curve, which plots TAR against FAR. The ROC curve provides a comprehensive view of system performance across all operating points. An alternative representation is the **Detection Error Trade-off (DET)** curve, which displays error rates on normal deviate scales, emphasizing performance differences in low-error operating regions, which are critical for high-security applications [14].

A widely used scalar summary of biometric performance is the **Equal Error Rate (EER)**, defined as the operating point at which FAR equals FRR.

The EER provides a single-value indicator of system accuracy. Lower EER values indicate better overall performance. However, EER reflects performance at only one operating threshold and may not represent system behavior under application-specific constraints. For instance, high-security systems typically operate at very low FAR values, accepting higher FRR to reduce the risk of unauthorized access [6]. Figure 2.3 illustrates the FAR, FRR, and their intersection at the EER.

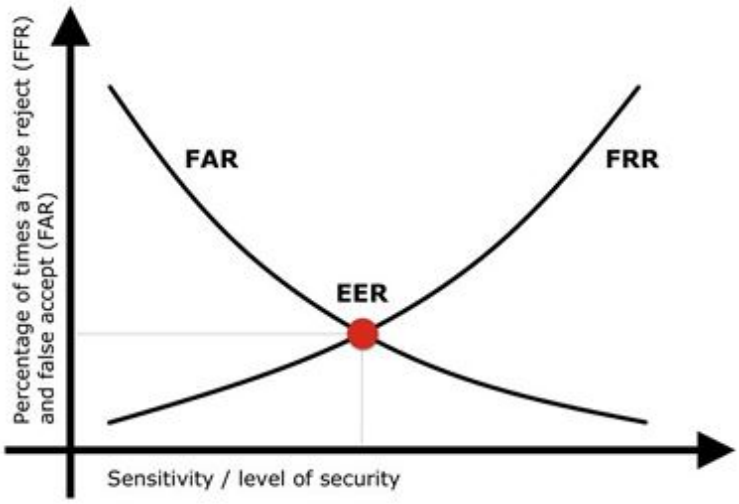


Figure 2.3: Illustration of FAR, FRR, and their intersection at the EER [15].

In identification scenarios (closed-set recognition), performance is commonly evaluated using the **Cumulative Match Characteristic (CMC)** curve, which represents the probability that the correct identity appears within the *top-k* ranked matches. The **rank-1 recognition rate** is the most commonly reported metric and indicates the percentage of cases where the correct identity is ranked first [14].

In open-set identification scenarios, where the probe may not belong to any enrolled identity, additional metrics are required. The **True Positive Identification Rate (TPIR)** measures the probability that an enrolled identity is correctly identified, while the **False Positive Identification Rate (FPIR)** measures the probability that an unenrolled identity is incorrectly accepted as belonging to the database [14].

The computation of these metrics relies on carefully designed evaluation protocols. Typically, datasets are divided into *gallery sets* for enrollment and *probe sets* for testing. Genuine comparisons are performed between samples of the same individual, while impostor comparisons involve samples from different individuals. The resulting score distributions form the

basis for computing performance curves and error rates [6].

It is important to note that aggregate performance metrics do not fully capture individual variability. The concept of the **"Biometric Menagerie"** introduced by Hicklin and Ulery [16] categorizes users based on their matching behavior, such as users who are consistently easy to recognize ("sheep") or difficult to match ("goats"). These individual differences can significantly affect system performance and must be considered in system design.

International standards such as **ISO/IEC 19795** define standardized methodologies for biometric performance evaluation, ensuring consistency, reproducibility, and comparability across studies. These standards provide formal definitions for key metrics and guide the design of fair and reliable biometric systems [6].

2.6 Biometric modalities

The diversity of human characteristics has given rise to an equally diverse set of biometric modalities, each offering distinct advantages and limitations for different applications. Understanding this landscape is essential for selecting appropriate modalities and for designing systems that combine multiple traits to achieve enhanced performance.

2.6.1 Fingerprint recognition

Fingerprint recognition remains the most widely deployed biometric technology, benefiting from over a century of forensic use and decades of automated system development. Fingerprint patterns are formed by ridges and valleys on the fingertip surface, with distinctive features including ridge endings (minutiae), ridge bifurcations, and broader pattern classes such as loops, whorls, and arches. Modern fingerprint systems capture images using optical, capacitive, or ultrasonic sensors, extracting minutiae locations and orientations as compact templates. The technology achieves excellent accuracy, with false match rates below 0.1% achievable in controlled conditions [3]. Fingerprint sensors have become sufficiently small and inexpensive to integrate into consumer devices, driving widespread adoption for mobile authentication. However, fingerprint systems face challenges including sensitivity to skin condition (dryness, moisture, damage), vulnerability to spoofing using artificial replicas, and cultural or hygienic concerns about touching shared sensors.

2.6.2 Face recognition

Face recognition has attracted intensive research attention due to its non-intrusive nature and compatibility with ubiquitous camera infrastructure. Face images can be captured at a distance without requiring user cooperation, enabling applications from surveillance to automatic photo tagging. Conventional face recognition algorithms depended on manually produced features, including eigenfaces, local binary patterns, or Gabor filters, but deep learning has revolutionized the field, with modern systems achieving accuracy exceeding human performance under controlled conditions. Zhang et al. [7] proposed a joint face detection and alignment framework using multitask cascaded convolutional networks that significantly improved face recognition performance in unconstrained environments. Face recognition faces challenges including sensitivity to illumination, pose, expression, occlusion, and aging. The technology has also raised significant privacy and ethical concerns, particularly regarding surveillance applications and potential demographic biases in system performance [3].

2.6.3 Iris recognition

Iris recognition is considered one of the most precise biometric modalities, leveraging the complex and stable of the iris the colored ring surrounding the pupil. The iris develops in utero and remains stable throughout life, with patterns that are highly distinctive even between identical twins.

2.6.4 Palmprint recognition

Palmprint recognition leverages the rich and diverse features of the human palm—such as principal lines, wrinkles, ridges, and minutiae over a larger surface area than fingerprints, resulting in enhanced discriminative capability and a less intrusive acquisition process. Advances in deep learning, including scattering convolutional networks and unsupervised CNNs combined with SVMs, have significantly improved recognition performance. Moreover, 3D palmprint systems incorporate depth information to enhance robustness against illumination and skin color variations, while multispectral imaging captures both surface textures and subsurface vascular patterns at different wavelengths. The fusion of these complementary features leads to a highly robust and secure biometric representation [17, 18].

2.6.5 Finger knuckle patterns (FKP)

Finger knuckle patterns (FKP) represent an emerging hand-based modality that analyzes the skin surrounding the finger knuckles. The dorsal surface of the hand, where these patterns appear, is typically exposed during ordinary hand movements, enabling non-intrusive acquisition. Knuckle patterns form complex and individually specific textures that remain relatively stable throughout adult life. Unlike fingerprints, which leave latent impressions on touched surfaces, knuckle patterns are less readily captured surreptitiously, offering enhanced privacy in certain applications. Herbadji et al. [19] demonstrated that multimodal biometric verification combining iris and major finger knuckles achieved a correct recognition rate of 95.54%, showing the potential of FKP as a complementary modality. Daas et al. [20] further showed that fusion of finger vein and finger knuckle print using deep learning could achieve robust recognition performance. Attia et al.[21] demonstrated that multiresolution analysis of knuckle, combined with appropriate classification frameworks, can achieve recognition performance competitive with established modalities. More recently, Attia et al. [22] developed deep rule-based classifiers for FKP recognition that offered the interpretability of rule-based systems alongside the representational power of deep learning architectures.

2.6.6 Other modalities

Other modalities continue to be explored for specialized applications. Gait recognition identifies individuals by their walking patterns, enabling recognition at a distance using video surveillance. Keystroke dynamics authenticates users based on typing rhythm, offering continuous authentication without additional sensors. Signature verification remains important for forensic and legal applications; Chattopadhyay et al. [23] proposed a self-supervised attention-guided reconstruction framework with dual triplet loss that achieved promising results for writer-independent offline signature verification. Emerging modalities include ear shape, finger vein patterns, retinal vasculature, and even cardiac signals from electrocardiogram (ECG) or photoplethysmography (PPG) sensors [24].

2.6.7 Multimodal biometric systems

Multimodal biometric systems combine information from multiple traits to overcome the limitations of individual modalities. By integrating complementary sources of biometric in-

formation, multimodal systems achieve recognition performance that exceeds any constituent unimodal system while simultaneously providing robustness against modality-specific failures. Qin et al. [25] provided a comprehensive survey of identity recognition via data fusion and feature learning, cataloging the various approaches to multimodal integration. Fusion can occur at multiple levels: Sensor-level fusion mixes raw data prior to processing; feature-level fusion combines feature vectors from each modality; score-level fusion combines matching scores from modality-specific classifiers; and decision-level fusion integrates final acceptance/rejection decisions. Rasool [26] conducted a systematic comparison of feature-level versus score-level fusion, demonstrating that the optimal approach depends on the specific modalities and the correlation between their representations. Score-level fusion is particularly common due to its flexibility and the relative ease of combining scores from different modalities. Ammour et al. [27] demonstrated the effectiveness of face-iris multimodal identification using score-level fusion, while Wang et al. [28] showed that CNN-based approaches with fusion of face and finger vein features could achieve robust performance. Alay and Al-Baity [19] developed deep learning approaches for multimodal biometric recognition fusing iris, face, and finger vein traits, achieving significant performance improvements over unimodal systems. Deep learning has enabled more sophisticated fusion strategies that learn optimal combinations directly from data, discovering cross-modal patterns that no single modality could provide [29].

2.7 Conclusion

This chapter has introduced the fundamental concepts of biometric recognition, highlighting the principles, system architecture, and evaluation criteria that underpin reliable identity verification. The seven key criteria universality, uniqueness, permanence, collectability, performance, acceptability, and resistance to circumvention provide a framework for assessing different biometric modalities, each with distinct advantages and trade-offs. Standardized system architectures include enrollment and recognition phases, while performance measures like false match rate, false non-match rate, and equal error rate facilitate objective assessment and enhancement. Understanding these fundamentals establishes the necessary foundation for exploring how deep learning techniques can enhance biometric systems, which will be the focus of the next chapter.

Chapter 3

Biometric Systems Based on Deep Learning

3.1 Introduction

The rebirth of artificial neural networks through deep learning (DL) has radically altered the domain of pattern recognition, encompassing biometric systems. DL denotes a category of machine learning techniques founded on artificial neural networks characterized by numerous layers that learn hierarchical representations. These approaches have attained significant success across various domains by autonomously identifying the representations required for detection, classification, or prediction straight from raw data [30].

This chapter provides the theoretical background on DL and its applications in biometric systems. First, artificial neural networks (ANNs) are presented. The chapter then focuses on key aspects of DL by highlighting its advantages and describing commonly used architectures in biometric applications, namely: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Autoencoders (AEs), and Generative Adversarial Networks (GANs). Furthermore, particular attention is given to lightweight and interpretable deep CNN models, which aim to address the limitations of conventional deep CNN architectures. In this context, the main drawbacks of traditional CNNs are first outlined, followed by a presentation of two representative approaches, PCANet and xDNN, which provide efficient and more interpretable alternatives. Finally, the chapter concludes with a dedicated section reviewing recent related

work on the application of DL techniques in biometric systems.

3.2 Artificial neural network

ANNs in artificial intelligence are a computational architecture inspired by the mental processes of the human brain. They comprise numerous highly interconnected processing units, termed artificial neurons, which function collaboratively to address complex problems [31].

They draw on the structure and function of the biological nervous system, adopting principles of neuronal communication and information processing. The human brain comprises between 86 to 100 billion neurons that convey and process messages throughout the body. Neurons are involved in sensory perception, motor control, and the regulation of involuntary physiological functions, while interneurons facilitate the integration and processing of information within the nervous system [32].

To model these mechanisms computationally, simplified mathematical representations of biological neurons were introduced. Perceptrons were developed as one such representation and were introduced by Frank Rosenblatt in 1957. Also referred to as formal neurons, they represent one of the earliest supervised methods for binary classification, separating linearly separable data into two distinct classes [32]. For example, a perceptron can be used to decide whether a given image represents the digit zero or not.

Figure 3.1 illustrates the structure of the perceptron, demonstrating its function of mapping a set of input signals to a singular output. It specifically gets an input vector:

$$\mathbf{X} = \{x_1, x_2, \ldots, x_l\}$$

and a weight vector:

$$\mathbf{W} = \{w_1, w_2, \ldots, w_l\}.$$

The i^{th} element of $\mathbf{X}$, possibly provided by adjacent neurons, is associated with the weight w_i. This weight may be either positive or negative, indicating the significance of the associated input.

The neuron computes the weighted sum of the inputs and includes a bias factor b, yielding

an internal activation value z, defined by the subsequent mathematical expression:

$$z = \sum_{i=1}^{l} w_i x_i + b. \tag{3.1}$$

This value is subsequently processed by an activation function $f(\cdot)$, which produces the neuron output y. Traditionally, the perceptron uses the following binary step function:

$$f(z) = \begin{cases} 1, & \text{if } z > 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

This function enables the perceptron to perform binary classification. Thus, an output value of 1 indicates that $\mathbf{X}$ belongs to the target class (i.e., the image represents the digit zero), while 0 indicates that it does not.

The bias term is essential to the operation of the perceptron. It allows the model to shift the decision boundary and better fit the data. In practice, the bias is often implemented as an additional neuron with a constant input of 1 and its own associated weight.

Besides the binary step function, many other activation functions can be used, such as the logistic, hyperbolic tangent, ReLU, and linear functions [32, 33].

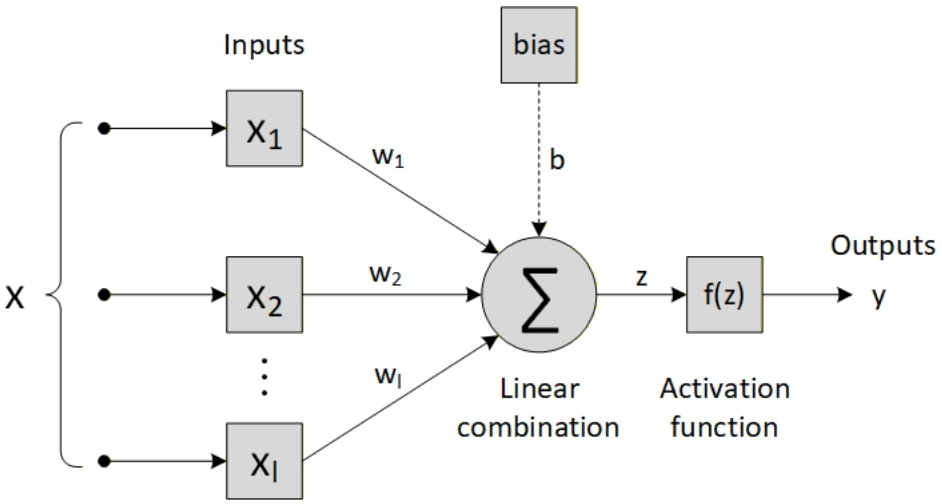


Figure 3.1: Perceptron structure [34].

Training a perceptron consists of adjusting its weights and bias so that it correctly classifies the training samples. The learning process is supervised, signifying that every input vector

corresponds to a designated output.

Given a training example $(\mathbf{x}, t)$, where t is the desired output, the perceptron produces a prediction y. The error is defined as:

$$e = t - y. \tag{3.3}$$

In the event of an erroneous prediction, the weights and bias are adjusted in accordance with the perceptron learning rule:

$$w_i \leftarrow w_i + \eta\, e\, x_i, \tag{3.4}$$

$$b \leftarrow b + \eta\, e, \tag{3.5}$$

where η is the learning rate.

This update rule increases the influence of inputs that contribute to correct classification and decreases the influence of those leading to errors. The perceptron learning algorithm is guaranteed to converge in a finite number of steps if the training data are linearly separable [33].

The perceptron model discussed previously represents the fundamental building block of any artificial neural network. Because of its limited structure, the perceptron is not well suited for modeling complex machine learning problems. A single neuron can solely establish a decision boundary that separates between two classes. Consequently, it is unsuitable for classifying more than two distinct classes [35].

A well-known example of this limitation is the XOR problem, which cannot be solved by a single perceptron regardless of the choice of weights and bias. Furthermore, the perceptron lacks the representational capacity to model complex, non-linear relationships commonly found in real-world data [32].

These limitations motivate the use of multilayer Feed-forward Neural Networks, commonly abbreviated as FNNs, which extend the perceptron model by incorporating multiple layers of neurons and non-linear activation functions, enabling the learning of more complex patterns.

A FNN has a layered architecture, as illustrated in Figure 3.2. Each layer is composed of a set of artificial neurons that receive inputs from the preceding layer, perform the required computations, and forward their outputs to the subsequent layer. Three main types of layers are

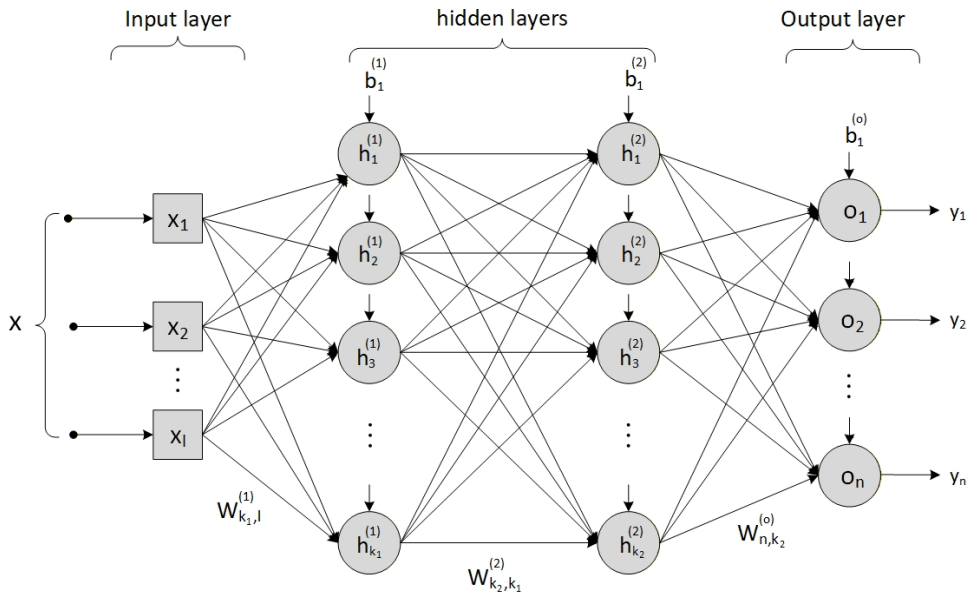


Figure 3.2: Architecture of a multilayer feed-forward neural network comprising two hidden layers.

distinguished.

Each training sample is fed into the network through the input layer, which the vector $\mathbf{X} \in \mathbb{R}^l$ representing the training features. The units forming this layer merely forward the input values to the next layer without performing any processing or transformation.

The network may consist of one or several hidden layers, each containing a collection of neurons that accept inputs from the preceding layer. In the first hidden layer, the input vector $\mathbf{X}$ is multiplied by the weight matrix $\mathbf{W}^{(1)} \in \mathbb{R}^{k_1 \times l}$ and combined with the bias vector $\mathbf{b}^{(1)} \in \mathbb{R}^{k_1}$. The linear combination is further processed by a non-linear activation function $f(\cdot)$, yielding the output of the hidden layer:

$$\mathbf{H}^{(1)} = f\left(\mathbf{W}^{(1)}\mathbf{X} + \mathbf{b}^{(1)}\right). \tag{3.6}$$

The output of this hidden layer is subsequently transmitted to any following hidden layers, enabling the network to acquire increasingly intricate feature representations. The output of the i-th hidden layer is generally expressed as:

$$\mathbf{H}^{(i)} = f\left(\mathbf{W}^{(i)}\mathbf{H}^{(i-1)} + \mathbf{b}^{(i)}\right), \quad i = 2, \ldots, L, \tag{3.7}$$

where $\mathbf{W}^{(i)} \in \mathbb{R}^{k_i \times k_{i-1}}$ and $\mathbf{b}^{(i)} \in \mathbb{R}^{k_i}$.

The final hidden layer, with k_L neurons, conveys its results to the output layer, which comprises of n neurons, thereby allowing the network to provide its definitive predictions. The network's final output is obtained by applying a weight matrix $\mathbf{W}^{(o)} \in \mathbb{R}^{n \times k_L}$ and a bias vector $\mathbf{b}^{(o)} \in \mathbb{R}^n$ to the output of the last hidden layer as follows:

$$\mathbf{Y} = g\left(\mathbf{W}^{(o)}\mathbf{H}^{(L)} + \mathbf{b}^{(o)}\right), \tag{3.8}$$

where $g(\cdot)$ represents the activation function applied at the output layer, and L specifies the total number of hidden layers within the network.

The learning process in a FNN consists of updating the network's weights and biases through an iterative training procedure [36, 37]. The network produces an output prediction by a forward pass, in which the input data is sent through the network n successive layers and processed using the aforementioned activation functions.

A loss function measures the difference between the model's projected output and the actual target output, acting as a quantitative performance metric during training. It functions as a benchmark for the optimization process, demonstrating the model's learning efficacy and assisting in the adjustment of its parameters to minimize errors. Prevalent loss functions encompass cross-entropy loss, predominantly utilized in classification jobs, and mean squared error (MSE), extensively employed for regression difficulties.

The backpropagation algorithm adjusts the weights and biases by calculating the gradients of the loss function concerning each network parameter, enabling the model to iteratively minimize prediction errors. This technique guarantees that the network learns from its errors by incrementally adjusting parameters to reduce the loss. Furthermore, backpropagation effectively transmits the mistake from the output layer to all preceding levels, allowing each layer to enhance the model's predictions. For layer l, the gradients of the loss $\mathscr{L}$ concerning the weight matrix $\mathbf{W}^{(l)}$ and the bias vector $\mathbf{b}^{(l)}$ are expressed as follows:

$$\frac{\partial \mathscr{L}}{\partial \mathbf{W}^{(l)}} \quad \text{and} \quad \frac{\partial \mathscr{L}}{\partial \mathbf{b}^{(l)}}$$

The gradients are calculated recursively using the chain rule, commencing with the output layer and progressing backward through the network. The parameters are subsequently modified utilizing gradient descent as follows:

$$\mathbf{W}^{(l)} \leftarrow \mathbf{W}^{(l)} - \eta \frac{\partial \mathscr{L}}{\partial \mathbf{W}^{(l)}} \tag{3.9}$$

$$\mathbf{b}^{(l)} \leftarrow \mathbf{b}^{(l)} - \eta \frac{\partial \mathscr{L}}{\partial \mathbf{b}^{(l)}} \tag{3.10}$$

where η denotes the learning rate. This process is executed iteratively until the loss reaches a minimum convergence.

3.3 Deep learning

DL is a subfield of machine learning that extends ANNs by introducing multiple interconnected hidden layers, giving rise to the notion of depth in neural architectures. Unlike shallow ANNs, deep neural networks leverage this depth to progressively learn hierarchical and increasingly abstract representations of data [32, 38]. The specific characteristics that distinguish DL from conventional ANNs are explored in this section by examining the transition from shallow to deep networks, the advantages brought by depth, and the main architectures commonly used in DL.

3.3.1 From shallow to deep networks: advantages of depth

The first ANNs were shallow, meaning that they consisted of only one or two hidden layer [35]. This limited their ability to learn complex functions, rendering them unsuitable for addressing numerous real-world challenges, including image recognition and speech processing. Furthermore, shallow networks often required extensive human feature engineering, wherein experts had to discern the most pertinent features to input into the model. This process is time-consuming, requires expert knowledge, and limits both the scalability and performance of the resulting models [37].

In contrast, deep neural networks comprise numerous hidden layers, facilitating hierarchical representation learning. Thus, each layer in a deep network extracts features at different

levels of abstraction. In image recognition tasks, the primary layers may capture fundamental qualities like as edges and textures, which are then relayed to subsequent layers that can identify more complex elements like forms and objects. This layered processing allows the network to discern intricate patterns and relationships within the data. This technique reduces the need for laborious feature engineering, as the deep neural network can independently extract significant features from raw data [39, 40].

The depth characteristic endows deep neural networks with the capability to learn intricate data exhibiting non-linear correlations, a hallmark of most real-world challenges. Utilizing more layers enables the network to approximate more complex functions, hence capturing nuanced patterns that shallow networks are incapable of representing.

Training deep networks requires solving several challenges that limited earlier approaches. The backpropagation algorithm computes the gradients of the loss function concerning network parameters, facilitating effective learning. Nonetheless, it may encounter issues of vanishing or inflating gradients in exceedingly deep networks. Architectural innovations such as skip connections, normalization layers, and specialized activation functions have addressed these issues, enabling reliable training of networks with hundreds of layers. However, increases in computational capacity and the availability of large datasets, along with the development of advanced activation functions, optimization techniques, and improvements in related techniques, have contributed to the feasibility and widespread adoption of deep neural networks [41].

3.3.2 Deep learning architectures

This section provides a concise overview of some prominent DL models. The primary architectures examined include CNNs, RNNs, GANs, and AEs. These models are highlighted because of their notable applications in biometric systems [42]. CNNs excel in image-based biometrics such as face, fingerprint, and iris recognition [43]. RNNs are effective for sequential biometric data, including voice, mouse movement, keystroke dynamics, handwritten passwords, and gait [44]. Autoencoders provide feature extraction and dimensionality reduction for biometric templates [45]. Finally, GANs can generate synthetic biometric samples that help improve the development, testing, and robustness evaluation of biometric systems [46].

3.3.2.1 Convolutional neural networks

A CNN is a specific category of deep learning models particularly engineered to identify patterns in images, audio, and various signal kinds. CNNs are extensively employed for applications like image classification, audio processing, object detection, and synthetic data creation [37]. Convolutional layers enable CNNs to independently learn spatial structures of features from incoming data, progressively identifying patterns from basic edges and textures to more intricate and abstract structures [47].

Figure 3.3 illustrates that a conventional CNN design consists of a sequence of convolutional layers, pooling layers, and fully connected layers, collectively generating the network's final prediction [47]. The CNN architecture can be conceptually divided into two main parts:

- **Feature Learning:** this element is tasked with recognizing and extracting significant features from the supplied image. The convolutional and pooling layers collaborate to identify patterns, textures, and other pertinent features for further analysis.
- **Prediction:** at this stage, the classification of the input image is ascertained based on the characteristics acquired during the feature learning phase. This technique is generally executed by several completely linked layers that generate the final output corresponding to the designated class.

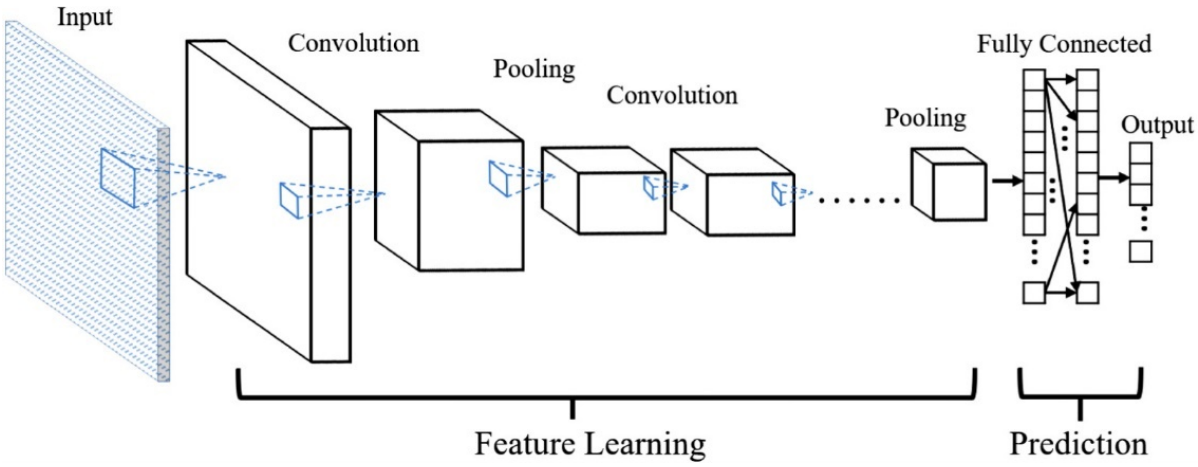


Figure 3.3: Typical Convolutional Neural Network Architecture [47].

CNNs process input data via a sequence of specialized layers that autonomously acquire hierarchical features. The convolutional layers utilize trainable filters (or kernels) to generate feature maps that identify specific patterns, edges, textures, and other prominent attributes.

Subsequently, pooling layers, often executed as max-pooling, diminish the spatial dimensions of the feature maps. This approach reduces computing complexity and alleviates overfitting while retaining essential information by selecting the maximum (or average) value from each region of the feature map. By layering many convolutional and pooling layers, the network may incrementally extract more abstract and intricate features from the input data. Following this procedure, one or more fully connected layers are employed to consolidate the retrieved features, integrate information across several regions, and execute the ultimate task, such as classification or regression, by producing the network's output [32, 47]. This architecture enables CNNs to learn directly from unprocessed data, minimizing the necessity for manual feature engineering. Moreover, it facilitates strong performance in tasks such as image and video recognition by identifying both local and global patterns.

3.3.2.2 Recurrent neural networks

RNNs are a class of DL architectures designed for the processing of sequential input. Unlike FNNs, RNNs have the potential to retain information from previous inputs through its internal state, thus creating a sort of memory. This capability enables the modeling of temporal correlations, making RNNs particularly appropriate for tasks like NLP and speech recognition. Figure 3.4 illustrates the essential architecture of an RNN, made up of an input layer, a hidden layer, and an output layer. The hidden layer has recurrent connections, allowing the network to capture temporal dependencies in the data [47, 48].

While conventional RNNs are theoretically suited to learn long-term dependencies in sequential data, they frequently face practical obstacles, such as disappearing and bursting gradients, which can significantly impede their capacity to model prolonged sequences successfully. To address these intrinsic limitations, Long Short-Term Memory (LSTM) networks, developed by Hochreiter and Schmidhuber [49], utilize specialized gating mechanisms, namely input, forget, and output gates, that facilitate the selective retention, alteration, and transmission of pertinent information across extended time periods. Building upon this notion, Gated Recurrent Units (GRUs), introduced by Cho et al. [50], provide a streamlined and computationally efficient framework that attains performance akin to LSTMs, enabling expedited training and rendering them highly suitable for various sequential modeling tasks.

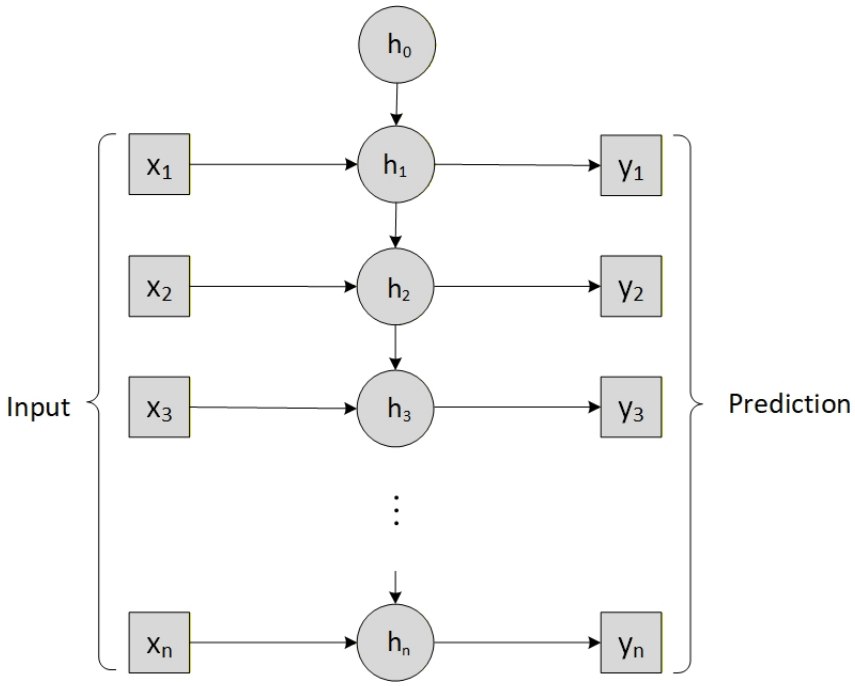


Figure 3.4: Core RNN Architecture [48].

3.3.2.3 Autoencoders

An AE is an unsupervised learning model capable of autonomously extracting features from input data across numerous samples. It is often used for dimensionality reduction and feature extraction, allowing the model to represent the essential information of the data [51].

An AE consists of two main components, each serving a distinct role in learning and reconstructing the input data [52]:

- **Encoder:** compresses the input data into a more succinct and informative representation (hidden layer), effectively extracting critical features and converting the input into an alternative feature space.
- **Decoder:** Reconstructs what was originally entered using the encoded representation, striving to reduce the divergence between the rebuilt output and the initial data, thus maintaining essential information.

As depicted in Figure 3.5, the typical structure of an autoencoder includes an input layer, hidden layer, plus an output layer. The input and output layers possess an equal number of neurons, however the hidden layer typically comprises less neurons ($t < n$) to facilitate data compression. This architecture necessitates that the network identifies and encodes the most

salient elements, resulting in a succinct representation of the input data. Thus, the autoencoder effectively learns significant embeddings that retain critical information, enabling precise reconstruction of the original input or facilitation of subsequent tasks.

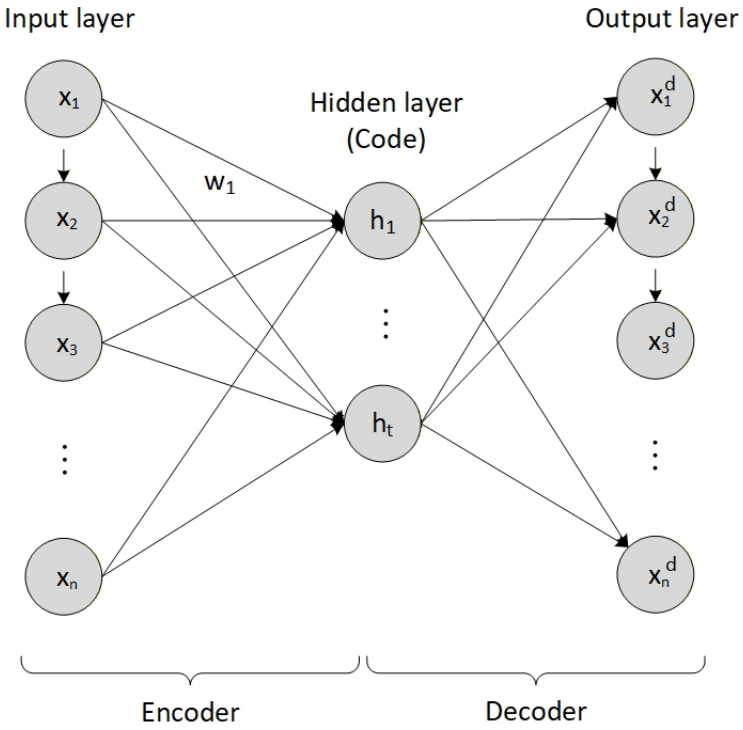


Figure 3.5: Simplified illustration of the autoencoder architecture [52].

AEs are powerful in DL because they can automatically extract features, reduce dimensionality, and help avoid overfitting. However, they also have limitations, such as longer training times for deep models and sometimes poor interpretability of the extracted features [52].

3.3.2.4 Generative adversarial networks

GANs represent an impressive category of generative models in machine learning that have attracted much attention in recent years, especially within the domain of computer vision. GANs utilize an competitive adversarial learning framework to generate highly realistic generated data, which can be used to enhance training datasets, improve data variety, or facilitate numerous imaginative and analytical applications. A standard GAN has two neural networks: a generator that creates real-world samples from random noise, and a discriminator that assesses these samples by trying to differentiate them from genuine data (see Figure 3.6). The networks are concurrently trained in a two-player minimax game, in which the generator attempts to deceive the discriminator, while the discriminator seeks to accurately distinguish

between genuine and created data. In this iterative adversarial procedure, both networks enhance their capabilities: the generator grows proficient at creating more convincing outputs, while the discriminator improves its ability to identify nuanced distinctions. Consequently, the trained generator frequently produces synthetic data that is nearly indistinguishable from authentic samples [53, 54].

GANs have been applied across various areas, including image production, image-to-image translation, face attribute manipulation, NLP, time-series synthesis, and semantic segmentation [54]. In contrast to other generative models, GANs do not necessitate an explicit likelihood function. They are trained inside a game-theoretic framework, enabling them to provide precise and realistic outputs, frequently exceeding the performance of approaches like variational autoencoders or limited Boltzmann machines [53].

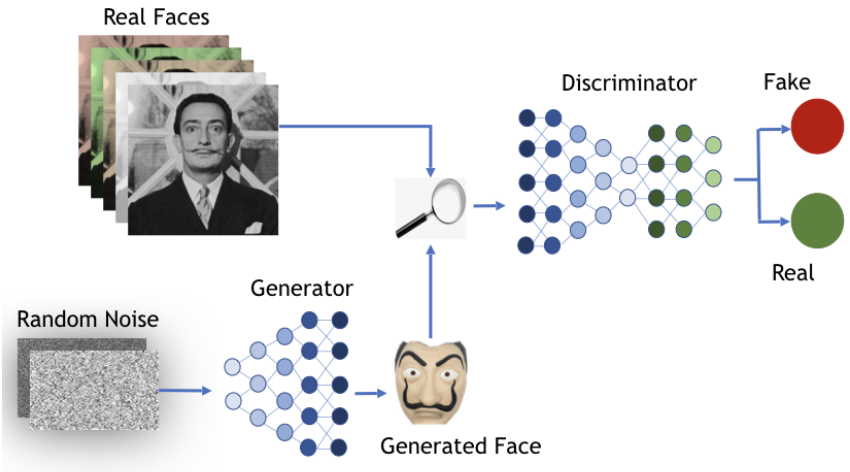


Figure 3.6: Architecture of a Generative Adversarial Network [55].

3.4 Lightweight and explainable CNN models

Conventional deep CNNs achieve high accuracy but are computationally expensive and poorly interpretable. To address this, lightweight methods such as PCANet and explainable models such as xDNN have been proposed. The following subsections briefly present these approaches and their advantages in efficiency, feature extraction, and explainability.

3.4.1 Limitations of conventional deep CNNs

CNNs have achieved remarkable success, but they present several limitations. First, their large number of parameters makes them prone to overfitting, often requiring techniques such

as dropout [56], data augmentation [57], and pretraining [58] to improve generalization. Moreover, CNNs demand large amounts of labeled training data to learn effectively, which can be costly and time-consuming to collect, especially for fine-grained tasks. They are also sensitive to data quality: noisy or biased datasets can lead the model to learn artifacts instead of meaningful patterns. Another challenge is the computational cost of CNNs. Training and inference for deep architectures require significant memory and processing power, though advances in GPUs and methods like pruning, quantization, and model compression can alleviate this issue [59].

On the other side, CNNs are often considered black-box models with low interpretability. Their decision-making process is opaque, making it difficult for humans to understand or trust predictions. This lack of transparency is critical in sensitive applications, such as autonomous vehicles, where users and regulators need explanations for decisions and model behavior, including the detection of biases related to gender, race, or age [60].

These limitations highlight the need for alternative deep learning approaches that are computationally efficient and/or explainable.

3.4.2 PCANet: A Deep learning feature extraction approach

PCANet, a fundamental DL network baseline introduced in [61], is extensively utilized in the domain of image categorization within the burgeoning field of DL. In contrast to other DL architectures, such as CNNs, which require extensive labeled training data and intricate understanding, PCANet facilitates a more straightforward training process. PCANet is founded on three fundamental processing components: (1) cascaded Principal Component Analysis (PCA) for high-level feature extraction, (2) binary hashing, and (3) histograms. The PCANet approach is summarized in the subsequent sections [61, 62].

3.4.2.1 PCA filter bank

The PCA filter bank comprises two convolutional stages. The initial phase involves estimating the filter banks through the application of PCA on a collection of filter vectors. Each vector is derived from a local window of dimensions $k_1 \times k_2$ centered on each pixel of the FKP image.

First, the mean of each vector is computed, and mean-centering is performed by subtracting the mean value from each entry of the corresponding vector. PCA is then applied to the resulting centered vectors in order to obtain the principal components. We retain the leading eigenvectors denoted by W, with dimensions $k_1 \times k_2 \times LS_1$, where LS_1 represents the number of principal components (leading eigenvectors).

Each principal component W is then regarded as a convolutional filter and reshaped into a kernel of size $k_1 \times k_2$. The resulting filters are subsequently applied to the input image through convolution, as follows:

$$T_l(x,y) = h_l(x,y) * I(x,y) \tag{3.11}$$

where $i \in [1,\ldots,LS_1]$. The symbol $*$ denotes the standard discrete convolution operator applied between the filter and the input image. The resulting image corresponds to the filtered output produced by applying the h_i filter to the input data. By applying the LS_1 learned filter columns of W to a given input FKP image I, a set of LS_1 distinct output feature maps is obtained, each capturing complementary structural information.

The second stage is subsequently performed by iteratively repeating the same convolutional procedure over all feature maps generated in the first stage of filter bank operations. For each resulting output image I', a local vector representation is constructed by considering a small spatial neighborhood of pixels around each location, thereby capturing local contextual information. The mean of each such vector is then computed to characterize its local distribution. A mean-centering operation is applied by subtracting the corresponding mean value from each element of the vector. The centered vectors extracted from all output images are then concatenated to form a global representation, which is used to estimate a second PCA filter bank composed of LS_2 filters. Finally, each learned filter is convolved with the input feature maps, producing a new and more discriminative set of output images for subsequent processing stages.

$$I_{l,m}(x,y) = h_m(x,y) * I_l(x,y) \quad \text{where } i \in [1..l_{s2}] \tag{3.12}$$

Therefore, by repeatedly applying the convolution process using both filter banks (LS_1 and LS_2), the output images of the first stage are further processed to generate the final set of output

images..

3.4.2.2 Binary hashing

During this phase, the output pictures L_{S1} and L_{S2} from the preceding stage are transformed into a binary format utilizing the Heaviside step function, which assigns a value of 1 for positive inputs and 0 for non-positive inputs.

$$I_{l,m}^{B}(i,j) = \begin{cases} 1, & \text{if } I_{l,m}(i,j) \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.13}$$

where $I_{l,m}^{B}$ denotes the binary image. In addition, for each pixel, the vector formed by the L_{S2} binary responses is interpreted as a binary code and converted into its corresponding decimal value. Consequently, the L_{S2} output feature maps are compacted into a single integer-valued image.

$$I_{l}^{D} = \sum_{m=1}^{L_{s2}} 2^{m-1} I_{l,m}^{B}(i,j) \tag{3.14}$$

where I_{l}^{D} represents the hashed image, whose pixel values are integer values lying in the range $[0, 2^{L_{S2}} - 1]$.

3.4.2.3 Histogram composition

At this stage, the hashed image I_{l}^{D} is subdivided into NB smaller regions, referred to as blocks B_i, where $i = 1, \ldots, NB$. For every individual block B_i, a corresponding histogram $H(B_i)$ is generated to capture its distribution characteristics. According to the specific requirements and design choices of the application, these blocks can be organized either as overlapping regions or as fully separate (non-overlapping) segments. Finally, the feature representation derived from $I_{l}^{D'}$ is obtained by concatenating the histograms computed over all blocks leading to the following formulation:

$$v_{l}^{hist} = [\mathbf{B}_{1}^{hist}, \mathbf{B}_{2}^{hist}, \ldots, \ldots, \mathbf{B}_{N_B}^{hist}] \tag{3.15}$$

Subsequent to the encoding phase, the feature vectors derived from the input picture I are concatenated as follows:

$$v_I^{hist} = [\mathbf{V}_1^{hist}, \mathbf{V}_2^{hist}, \ldots, \ldots, \mathbf{V}_{L_{S1}}^{hist}] \tag{3.16}$$

To summarize, the parameters of PCANet primarily include the filter size (k_1, k_2), the number of filters employed at each stage (L_{S_i}), the overall number of stages (N_s), and the block size (B), which is specifically used for computing local histograms in the output layer.

3.4.2.4 PCANet Parameters

The performance evaluation of the proposed PCANet-based identification system requires a careful and systematic adjustment of several important parameters, namely the number of stages, the number of filters at each stage, the filter dimensions, the block-wise histogram size, and the overlap ratio between adjacent blocks. These parameters play a crucial role in extracting highly discriminative and representative features that effectively characterize the input FKP image, thereby directly influencing and potentially enhancing the system's overall recognition accuracy. In practical implementations, the selection of these parameters is typically guided by empirical evaluation, experimental validation, and performance analysis on relevant datasets.

3.4.3 xDNN: Explainable Deep Neural Networks

The xDNN [63] proposes a dynamic feedforward architecture that autonomously adapts and evolves its structure through continuous learning. Unlike traditional static neural networks, this approach continuously updates its layered configuration to better accommodate changing pattern recognition requirements. As illustrated in Figure 3.7, the model is composed of five dedicated processing layers that operate collaboratively to enable:

- Adaptive reconstruction of the feature space.
- On-demand generation of neural modules.
- Formation of transparent decision-making pathways.

These five functionally specialized layers collaborate in a coordinated and complementary way to successfully perform the overall pattern recognition task within the proposed framework. To provide a clearer and more detailed understanding of the system, we systematically present the computational role of each layer, describe its structural organization, and further detail its internal operations and interactions within the complete classification pipeline.

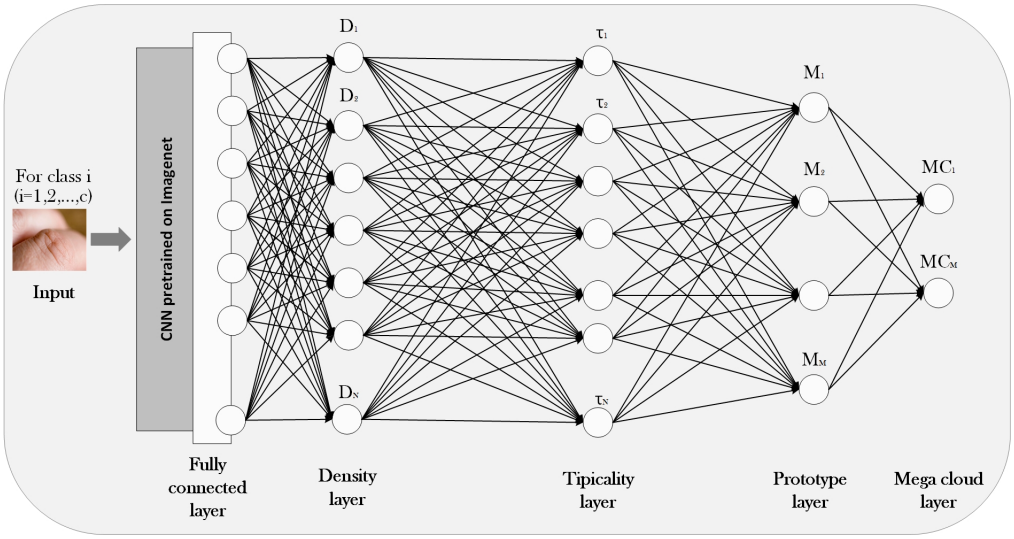


Figure 3.7: xDNN Classifier's architecture.

3.4.3.1 Feature descriptor layer: Extraction of High-Level Features

The initial stage of the xDNN pipeline consists of extracting features from the input data, where raw input images are transformed into compact and informative numerical representations suitable for learning. Although classical approaches such as PCA or handcrafted descriptors like Gabor filters have been extensively used in the literature, they are often limited in their ability to model complex nonlinear structures and high-level semantic variations present in the data. To address this limitation, we adopt VGG-VD-16, a pre-trained deep convolutional neural network (DCNN) that is widely recognized for its strong representational power and its ability to extract highly discriminative and robust visual features from images. As shown by Simonyan and Zisserman [64], this architecture is more effective than shallower networks in learning hierarchical representations. Following the approach of Ren et al., we retain only the first fully connected layer (FCL), which provides a good balance between feature dimensionality and discriminative power, resulting in a 1×4096 feature vector for each image. In addition, and in line with Ioffe and Szegedy [65], feature normalization is applied to improve training stability and convergence.

Feature standardization and normalization are fundamental preprocessing operations within this framework. They are formally expressed according to Equations 3.17 and 3.18, respectively:

$$\hat{x}_{i,j} = \frac{x_{i,j} - \mu(x_{i,j})}{\sigma(x_{i,j})} \tag{3.17}$$

$$\hat{x}_{i,j} = \frac{\hat{x}_{i,j} - \min\limits_{i}(\hat{x}_{i,j})}{\max\limits_{i}(\hat{x}_{i,j}) - \min\limits_{i}(\hat{x}_{i,j})} \tag{3.18}$$

Here, $\hat{x}$ denotes the standardized and normalized feature vector obtained from x, where x corresponds to the original feature values extracted from the fully connected layer (FCL) of the input image I. The term $\bar{x}$ refers to the normalized feature vector. The index i corresponds to the image ID within the dataset, while j denotes the j-th feature component of the vector x. In this context, I represents the current input image under processing, and N is the total number of images available in the dataset. This preprocessing step is essential as it ensures that all extracted features are brought to a comparable scale, allowing each feature to contribute fairly and effectively during the subsequent classification process.

The xDNN model dynamically initializes its meta-parameters, thereby eliminating the need for manual tuning or heuristic adjustment. Upon receiving the initial data sample, the system autonomously adapts and updates its internal parameters based on the observed data distribution, in accordance with Equation 3.19:

$$P \leftarrow 1; \quad \mu \leftarrow x_i; \quad C_1 \leftarrow x_1; \quad p_1 \leftarrow x_1; \quad Support_1 \leftarrow 1; \quad r_1 \leftarrow r^*; \quad \hat{I}_1 \leftarrow I_1 \tag{3.19}$$

Here, P signifies the prototype, C denotes the class label, and the support indicates the quantity of samples linked to the specified model. The parameter r specifies the radius of the effect region for class C_i, whereas μ represents the global mean of the class. I denotes the current input image, while $\hat{I}$ signifies the recognized prototype.

The 30° angular similarity threshold is motivated by neurophysiological evidence reported by Tanaka [66], which indicates that this range provides an optimal balance for discrimination

sensitivity in primate visual cortical neurons. In addition, the dynamic radius computation follows the data-driven approach proposed by Angelov et al. [67], enabling autonomous parameter estimation within the model.

Following the methodology outlined in [68], the critical threshold is computed as follows:

$$r^* = \sqrt{2 - 2\cos(30^\circ)}.$$

The metric r^* is a boundary derived analytically, not a parameter adjusted empirically. It is stipulated that the angle between two vectors is less than 30° and that both vectors possess the same orientation, indicated by d.

Building on this principle, two feature vectors are considered *similar* when the angle between them is less than 30°. The direction d is determined using Equation 3.20.

$$d(x_i, p_i) = \left\| \frac{x_i}{\|x_i\|} - \frac{p_i}{\|p_i\|} \right\| \tag{3.20}$$

3.4.3.2 Density layer

This layer is responsible for modeling the shared proximity between images in the feature space generated by the previous layer. When Euclidean distance is used, the data distribution follows a Cauchy-like behavior, as shown in [68, 63]. Unlike Gaussian kernels, the Cauchy formulation provides enhanced robustness to outliers and noisy observations, making it particularly well suited for handling real-world biometric data with inherent variability. In this context, the data density D is estimated using either Equation 3.21 or Equation 3.22, depending on the specific formulation adopted within the model.

$$D(x_i) = \frac{1}{1 + \frac{\|x_i - \mu_N\|^2}{\sigma_N^2}} \tag{3.21}$$

$$D(x_i) = \frac{1}{1 + \|x_i \mu_i\|^2 + \sum_i - \|\mu_i\|^2} \tag{3.22}$$

The scalar coefficient $\sum_i$ can be updated as in 3.23, where μ_i is involved.

$$\mu_i \leftarrow \frac{i-1}{i}\mu_{i-1} + \frac{1}{i}x_i \quad \sum_i = \frac{1}{1-i}\sum_{i-1} + \frac{1}{i}\|x_i\|^2 \quad \sum_1 = \|x_1\|^2 \tag{3.23}$$

Elevated density values signify enhanced cluster cohesion, demonstrating a robust reciprocal influence among images that exist in the data space owing to their common proximity.

3.4.3.3 Typicality layer

Typicality τ quantifies the degree to which a sample aligns with its class distribution, as determined by the associated probability density function, as articulated in Equation 3.24. It may be construed as a probabilistic confidence metric.

$$\tau(x_i) = \frac{\sum_i^c Support_i D(x_i)}{\sum_i^c Support_i \int_{-\infty}^{+\infty} D(x_i)dx} \tag{3.24}$$

Typicality values τ lie within the interval $(0,1)$, reaching their maximum near prototype vectors and decreasing toward outliers. A high τ indicates strong confidence in the classification, while its value always remains strictly below 1.

3.4.3.4 Prototypes or models layer

The xDNN framework generates inherently interpretable IF–THEN rules for each class by employing a transparent and data-driven rule-based learning strategy. This rule induction process is grounded in an interpretable fuzzy rule extraction approach, which enables a clear mapping between input patterns and class decisions while preserving model transparency. In addition, its dynamic and adaptive model expansion mechanism extends the classical notion of novelty detection, as originally introduced by Markou and Singh [69], allowing the system to continuously incorporate new patterns and evolve over time without retraining from scratch.

The xDNN architecture proposes a novel, flexible, and inherently interpretable paradigm within explainable artificial intelligence, centered around its Models Layer, which functions as a dynamic, adaptive, and self-organizing mechanism. This layer is designed to directly learn complex and highly non-linear data distributions from raw visual inputs in a fully data-driven manner, without requiring predefined statistical models, handcrafted features, or any prior domain-specific assumptions.

Its modular and extensible structure enables each classification model to operate independently in an autonomous way, which in turn supports the seamless incorporation of new classes while preserving previously acquired knowledge and preventing performance degradation. Such characteristics make the framework particularly suitable for scalable and real-world biometric scenarios, where systems are expected to evolve continuously, manage growing data diversity, and maintain both robustness and stability over time.

The Models Layer constitutes the fundamental interpretable element of the xDNN biometric system. It functions independently of preconceived notions on data distributions, instead acquiring knowledge directly from the visual patterns seen in the input photos. This architecture's primary advantage is its modular independence, facilitating the integration of new models without affecting existing ones, thus allowing for scalable growth while maintaining system stability.

At the learning stage, xDNN operates through a class-wise processing mechanism in order to construct distinct and well-structured sets of models, each one tailored to capture the essential data density characteristics extracted during the preceding learning phases. These models are then systematically used to derive clear, organized, and inherently interpretable decision rules, generally formulated in a logical IF–THEN structure, thereby improving both the transparency and explainability of the classification process:

$$\text{IF (input } \sim \text{ prototype } K_1\text{) THEN classify as Class C}$$

The symbol $\sim$ denotes both similarity and the corresponding degree of membership. In this context, multiple rules may be generated for a single model, however, those associated with the same class are subsequently combined in a unified manner using the logical disjunction *OR*, as follows:

$$\text{IF (input } \sim \text{ prototype } K_1\text{) OR} \ldots \text{OR (input } \sim \text{ prototype } K_m\text{) THEN classify as Class C}$$

Each individual model defines a Data Cloud [68], which represents a dynamic and adaptive

region that groups together feature vectors exhibiting high similarity in the feature space. In contrast to conventional approaches that depend on statistical averages or parametric assumptions, these clouds are directly anchored to real and representative data samples, ensuring a more faithful representation of the underlying distribution. The system also includes an adaptive assignment mechanism that iteratively compares each incoming input with the existing set of prototypes, using a minimum distance criterion:

$$j^* = \arg \min_{j=1,2..p} \left(\|x_i - p_j\|^2 \right) \tag{3.25}$$

The system incrementally establishes new recognition clusters whenever at least one of the following density-related conditions is fulfilled:

$$\begin{aligned} &\text{IF } (D(x) > \max_{j=1,2..P} D(p_j)) \text{ OR} (D(x) < \min_{j=1,2..P} D(p_j)) \text{ THEN (add a new data cloud} \\ &(P \leftarrow P+1)) \end{aligned} \tag{3.26}$$

Cluster Initialization:

$$P \leftarrow P+1; C_P \leftarrow x_i; p_P \leftarrow I_i; Support_P \leftarrow 1; r_P \leftarrow r_o; \hat{I_P} \leftarrow I_i; \tag{3.27}$$

For existing clusters, the system applies incremental update mechanisms:

$$\begin{aligned} &C_{j^*} \leftarrow C_{j^*}+1; p_{j^*} \leftarrow \frac{Support_{j^*}}{Support_{j^*}+1} p_{j^*} + \frac{Support_{j^*}}{Support_{j^*}+1} x_{j^*}; Support_{j^*} \leftarrow Support_{j^*}+1; \\ &r_{j^*}^2 \leftarrow \frac{r_{j^*}^2 + (1 - \|p_{j^*}\|^2)}{2} \end{aligned} \tag{3.28}$$

This layer's adaptive and self-organizing characteristics facilitate a smooth and organic system expansion while preserving high classification accuracy and stability. As a result, it is particularly well suited for dynamic and evolving biometric applications, where the system may need to incorporate new classes or update existing representations over time without retraining

from scratch or degrading performance.

3.4.3.5 Mega cloud MC layer

The cloud fusion technique enhances the hierarchical aggregation method by incorporating angular similarity restrictions. To improve interpretability, data clouds of the same kind are consolidated into larger entities known as Mega Clouds (MCs):

$$R_c : \text{IF } (\bar{v}_i \sim MC_1) \text{ OR } (\bar{v}_i \sim MC_2) \text{ OR } \ldots \text{ OR } (\bar{v}_i \sim MC_m) \text{ THEN Class}_c \tag{3.29}$$

This step is designed to reduce the complexity of the generated rules while simultaneously preserving a high level of transparency and interpretability for end-users. By simplifying the decision structure without compromising performance, the model remains both efficient and explainable. We conclude this section by providing a concise summary of the overall procedure in Algorithm 1.

Algorithm 1 xDNN Classifier's Algorithm [63]

Step A: Initialization

Read the first feature vector sample x_i representing the image I_i of class c

Initialize: $i \leftarrow 1; n \leftarrow 1; P_1 \leftarrow 1; p_1 \leftarrow x_i; \mu \leftarrow x_i; \text{Support} \leftarrow 1; r_1 \leftarrow r_o; \hat{I}_1 \leftarrow I_i;$

Step B: Execution

for each $i = 2, 3, \ldots$ **do**
    Read the new feature vector x_i
    Compute $\bar{x}$ and r^* according to Equation 3.25
    if Equation 3.28 holds **then**
        Generate rule according to Equation 3.29
    else
        Search for the closest data cloud according to Equation 3.27
        Update rule according to Equation ...
    end if
end for

3.5 Related work on deep learning–based biometric systems

This section provides an overview of DL–based approaches for biometric systems. These systems have also been developed using handcrafted feature extraction and conventional machine learning algorithms, including directional filter banks combined with dimensionality reduction techniques such as LDA and distance-based classifiers, as well as illumination normalization methods such as Retinex combined with HOG descriptors and similarity measures [70, 71]. Hybrid methods integrating handcrafted features (e.g., Gabor filters and BSIF) with rule-based classifiers have also demonstrated strong performance [22]. However, these approaches rely heavily on manually designed features and offer limited capability for automatic representation learning. In contrast, deep learning has emerged as a powerful paradigm, enabling highly accurate recognition across multiple biometric modalities.

This section reviews the state-of-the-art approaches highlighting both the architectures and training strategies that have been applied. We first discuss training strategies for deep networks in biometric applications. Next, we review specific DL architectures: CNNs for image-based biometrics, RNNs for sequential data, autoencoders for unsupervised feature extraction, and GANs for synthetic data generation and augmentation. We also examine lightweight and explainable models, which address efficiency and interpretability challenges of conventional deep networks.

3.5.1 Training deep networks for biometrics

Training deep networks for biometric recognition involves specific considerations that distinguish it from generic image classification. The objective of biometric recognition—distinguishing between many different identities while remaining invariant to within-identity variations requires specialized loss functions and training strategies [72].

The conventional approach treats biometric recognition as a classification problem, with each enrolled identity corresponding to a distinct class. A deep network is trained to predict identity labels from biometric samples, using a softmax output layer and cross-entropy loss. The features from the penultimate layer serve as the biometric template, capturing the representations learned through classification training. This approach works well when the training set includes many samples per identity and when the test identities are drawn from the same

distribution as training identities.

For applications where the system must generalize to identities not seen during training the typical scenario in deployed biometric systems classification-based training must be supplemented with strategies that promote template discrimination. Metric learning approaches, including contrastive loss, triplet loss, and their variants, directly optimize the relative distances between templates. Triplet loss, in particular, encourages that samples from the same identity (anchors and positives) be closer together than samples from different identities (anchors and negatives) by at least a specified margin. Training with triplets requires careful sampling to select informative triplets that drive learning, but can produce highly discriminative embeddings that generalize well to new identities [72]. Chattopadhyay et al. [23] demonstrated the effectiveness of dual triplet loss for signature verification, sampling negative samples from both within the same writer class and from other writers to achieve robust discriminative learning.

Margin-based loss functions including ArcFace, CosFace, and AdaFace have become the dominant approach for training face recognition models. These losses introduce angular margins between classes in the embedding space, encouraging compact intra-class distributions while maintaining separation between different classes. ArcFace, proposed by Deng et al. [73], has become the industry standard for training face recognition models due to its effectiveness and relative simplicity. For biometric applications, margin-based losses have demonstrated superior performance compared to traditional softmax or metric learning approaches, particularly when combined with large-scale training datasets.

Data augmentation plays a critical role in training robust biometric networks. By applying transformations that simulate real-world variations, rotation, scaling, cropping, illumination changes, and noise addition augmentation expands the effective training set and teaches the network to ignore nuisance factors. For biometric applications, augmentation must preserve identity-discriminative information while introducing realistic variations. Face recognition networks, for example, are typically trained with augmentations simulating pose variation, expression changes, and illumination conditions that the system will encounter in deployment [72]. Huang et al. [74] proposed patch-level data augmentation specifically designed for Vision Transformer-based face recognition, demonstrating that augmentations preserving facial structure while modifying dominant patches could effectively address the patch-level overfitting issue inherent in transformer architectures.

Few-shot learning techniques address the challenge of training with limited labeled data, a common scenario in biometric applications where collecting large-scale datasets may be impractical. Transfer learning, as discussed above, leverages pre-trained models to reduce the amount of task-specific data required. Meta-learning approaches, including prototypical networks and model-agnostic meta-learning (MAML), train models on multiple small-sample tasks so they can quickly adapt to new identities with minimal data [72]. Generative models, including GANs and autoencoders as we will see later, can synthesize additional training samples to augment small datasets, creating realistic variations in pose, illumination, or expression that improve model generalization.

3.5.2 CNN-Based biometric Systems

CNNs have emerged as the predominant DL architecture for image-based biometric recognition, including face, fingerprint, iris, and palmprint modalities. CNNs are specifically designed to process data with grid-like topology, exploiting three key ideas that align naturally with the properties of biometric images [30].

Local connectivity restricts each neuron to receive inputs from only a small spatial region of the previous layer, rather than from all neurons. This reflects the observation that local patterns, edges, textures, and minutiae are the building blocks of biometric images, and their statistics are relatively stationary across the image. By focusing on local receptive fields, CNNs dramatically reduce the number of parameters compared to fully connected networks and encode the prior knowledge that nearby pixels are more relevant to each other than distant pixels.

Weight sharing means that the same set of weights (a filter or kernel) is applied across all spatial locations. A filter that detects a particular pattern, say, a ridge ending in a fingerprint will be useful regardless of where that pattern appears in the image. Weight sharing further reduces parameters and ensures that the network learns translation-invariant features: a pattern can be recognized anywhere in the image using the same detector.

Pooling (subsampling) reduces the spatial dimensions of feature maps, creating a coarse-to-fine hierarchy of representations. Pooling aggregates information over local regions, providing tolerance to small translations and distortions while reducing computational load. Max pooling, which retains the maximum value within each pooling window, is particularly common and

has been shown to preserve the most salient features while discarding irrelevant variations. A typical CNN for biometric recognition comprises alternating convolutional and pooling layers, followed by one or more fully connected layers that produce the final representation or classification. Early layers respond to simple patterns such as oriented edges or color blobs; deeper layers combine these primitives into modality-specific features corresponding to discriminative characteristics minutiae configurations in fingerprints, textural patterns in iris, or facial components in face recognition. The application of CNNs to biometric recognition has followed two primary approaches. The first approach trains specialized architectures from scratch on biometric datasets, allowing the network to learn representations optimized specifically for the target modality and population. This approach requires substantial labeled data, which may not be available for all biometric applications. The second approach, transfer learning, adapts networks pre-trained on large-scale generic image datasets (such as ImageNet) to biometric tasks. By initializing with weights learned from generic image recognition and fine-tuning on biometric data, transfer learning achieves excellent performance with substantially less training data and computational resources [72]. Aung et al. [75] demonstrated the effectiveness of transfer learning with deep CNNs for multimodal biometric recognition in surveillance videos.

The evolution of CNN architectures has produced increasingly sophisticated networks that have been successfully adapted for biometric applications. AlexNet, which catalyzed the modern DL revolution through its 2012 ImageNet performance, established the template for modern convolutional networks with its combination of convolutional layers, ReLU activations, and dropout regularization. For biometric recognition, AlexNet provides a proven architecture capable of learning rich feature representations when sufficient training data are available or when fine-tuned from ImageNet pre-training.

The Visual Geometry Group (VGG) network demonstrated the value of increased depth, using very small convolutional filters (3×3) throughout a 16- or 19-layer architecture. The consistent use of small filters enables deeper networks while maintaining manageable parameter counts, and the architecture's simplicity has made it a popular choice for transfer learning. Modified VGG architectures have been applied to both unimodal and multimodal biometric recognition tasks, demonstrating the versatility of the design.

ResNet (Residual Network) introduced the transformative innovation of skip connections that enable effective training of substantially deeper architectures [30]. By allowing gradients

to flow directly through identity mappings, ResNet overcame the degradation problem that previously limited network depth, enabling architectures with dozens or hundreds of layers. For biometric applications, this depth enables learning of hierarchical representations spanning multiple levels of abstraction, from local texture details to global structural patterns. ResNet-based architectures have demonstrated state-of-the-art performance across diverse biometric modalities, including face, fingerprint, and palmprint recognition. Kadhim and Abdulameer [76] developed a multimodal biometric database and conducted case studies for face recognition using deep learning, demonstrating the effectiveness of ResNet architectures.

More recent architectures continue to push the boundaries of performance and efficiency. DenseNet connects each layer to every other layer in a feed-forward fashion, encouraging feature reuse and reducing parameter counts. EfficientNet systematically scales network dimensions to achieve optimal performance under computational constraints. These architectures offer particular value for biometric applications deployed on resource-constrained devices such as mobile phones or embedded systems.

The Vision Transformer (ViT) has emerged as an alternative to CNNs, applying transformer architectures originally developed for natural language processing to image recognition tasks. Unlike CNNs, ViT lacks built-in inductive biases such as locality and translation equivariance, making it highly dependent on large-scale training data to learn these properties. Huang et al. [74] proposed TransFace, a ViT-based face recognition framework that addresses the patch-level overfitting issue inherent in transformer architectures. Their approach, which combines patch-level data augmentation with entropy-based hard sample mining, demonstrated that ViT could achieve state-of-the-art face recognition performance when properly calibrated.

3.5.3 RNN-Based biometric systems

Diverse recurrent architectures have been employed in biometric systems for individual identification and identity verification. A range of biometric traits has been investigated, including eye motions [77], ECG [78], BCG [78], touchscreen passwords [79], visual speech [80], and gait [81], among others.

For example, V. Chandrabanshia and S. Domnic [80] proposed a resilient liveness detection framework utilizing Visual Speech Recognition (VSR) as a supplementary authentication layer.

The method examines lip motion dynamics from video sequences to authenticate user liveness with an encoder–decoder architecture designed for sequential modeling. The encoder integrates a 3D-CNN for spatial feature extraction with a hybrid recurrent architecture that mixes Bidirectional Gated Recurrent Units (BiGRU) and Bidirectional Long Short-Term Memory (BiLSTM) networks to capture spatio-temporal relationships in lip motions. To improve accuracy in prediction, the decoder utilizes a BiGRU–BiLSTM architecture supplemented with Multi-Head Attention (MHA), facilitating efficient emphasis on pertinent temporal variables during sequence-to-text decoding. This RNN-based architecture proficiently captures intricate sequential patterns in visual speech. Experimental findings indicate a substantial enhancement, attaining a word error rate of 0.79%, surpassing current VSR-based authentication techniques and bolstering resistance against facial spoofing assaults.

P. Hossain Progga et al. [81] concentrated on individual identification through gait analysis utilizing sequences of body landmarks. A recurrent Siamese architecture was utilized to represent temporal dynamics for biometric recognition. Sequential gait landmarks are obtained by MediaPipe, geometrically aligned using Procrustes analysis, and analyzed by a Siamese bidirectional GRU dual-stack network that learns distinctive temporal representations and a similarity score for identity matching. The proposed method demonstrates robust performance on various large-scale datasets, indicating elevated accuracy in recognition compared to various gait-based identification benchmarks.

H. Qin et al. [82] introduced ABLSTM-TSVT, an RNN-enhanced transformer for multi-view finger-vein recognition, specifically aimed at improving biometric identification and verification efficacy. The model enhances LSTM with attention for temporal feature extraction, incorporates a local temporal attention module across patches and views, and integrates a spatial attention module to create a temporal-spatial transformer that enhances identification and verification metrics on multi-view datasets. The suggested ABLSTM-TSVT surpasses state-of-the-art techniques on two multi-view finger-vein datasets, enhancing identification accuracy and diminishing verification errors in vein classifiers.

X. Zhang et al. [78] examined the application of RNNs for modeling cardiac biometric signals, emphasizing architectures including basic RNNs, LSTM, and GRU to capture temporal dependencies in physiological waveforms. The study illustrates that RNN-based models can proficiently acquire discriminative representations for identity recognition and verification by

utilizing the sequential characteristics of ECG and BCG data. The results underscore the potential of integrating various cardiac modalities into a cohesive RNN architecture to improve biometric robustness and reliability.

R. Tolosana et al. [79] improved conventional password-based authentication by implementing a two-factor approach wherein users sketch password characters on touchscreens rather than typing them, therefore mitigating the intrinsic security vulnerabilities of traditional passwords. It introduces a unique Time-Aligned Recurrent Neural Network (TA-RNN) that combines Dynamic Time Warping (DTW) with recurrent neural networks to effectively address temporal fluctuations in stroke sequences and enhance resilience against misalignment in practical mobile environments. The methodology is assessed using extensive datasets, including MobileTouchDB [1], a publicly accessible collection gathered in unconstrained mobile environments, comprising over 64,000 character samples from 217 individuals spanning 94 smartphone models and various acquisition sessions, in addition to e-BioDigitDB[2]. The system is compared with classic DTW-based methods and typical RNN-based approaches, illustrating the benefits of integrating alignment with deep temporal modeling. Experimental findings indicate that the proposed TA-RNN markedly enhances authentication performance and resilience, surpassing current state-of-the-art techniques in demanding mobile contexts.

3.5.4 Autoencoder-based biometric systems

AEs serve multiple roles across biometric modalities. AEs are applied for denoising and enhancement, feature compression for matching, and feature extraction for classification, with architectures tuned to preserve fine spatial patterns or to condense global signatures. Architectures integrate domain cues (wavelet subbands for fingerprints, atrous/dilated convolutions for ridge detail, patch-wise training for vein patterns, and bottleneck vectors for gait maps) to retain discriminative biometric information.

Many studies on fingerprint biometrics have been proposed that leverage the capabilities of AEs. Y. Qi et al. [83] proposed a Dense Dilated Convolution ResNet Autoencoder (DDC-ResNet), where standard convolutions are replaced with atrous convolutions to preserve edge details and expand the receptive field. Evaluated across 54 feature extractor–classifier combi-

[1] https://github.com/BiDAlab/MobileTouchDB
[2] https://github.com/BiDAlab/eBioDigitDB

nations, the model achieved an average accuracy of 96.5% (97.52% for males, 95.48% for females). The study also showed that the right ring finger is the most informative for gender classification and identified key discriminative features, including loops and whorls, bifurcations, and line shapes. Y. Liang and W. Liang [84] addressed the limitations of biometric authentication in IoT devices, where fingerprint images are often degraded by noise and conventional deep learning denoising models are too computationally heavy for deployment. To solve this, the authors presented a lightweight Residual Wavelet-Conditioned Convolutional Autoencoder (Res-WCAE) specifically engineered for fingerprint denoising. The architecture employs two encoders (an image encoder and a wavelet encoder) to extract complementary information: spatial features from the raw picture and frequency-based details from wavelet decomposition. The representations are amalgamated and input into a singular decoder, which reconstructs a clean fingerprint image while maintaining intricate ridge structures. Residual connections between the encoder and decoder facilitate the preservation of nuanced features. Assessed on Additive White Gaussian Noise (AWGN) and synthetic noisy datasets, Res-WCAE surpasses current denoising techniques, especially in high noise environments, while maintaining a lightweight profile suitable for IoT implementation. P. Selvaraj et al. [85] tackled the issue of poor fingerprint image quality, a significant contributor to biometric matching failures and noted demographic performance disparities. These failures frequently result from noise, acquisition conditions, and physiological characteristics (e.g., age, gender, and finger type), rather than inherent variations in fingerprints. The authors propose an autoencoder to analyze and address this issue while maintaining privacy, which decomposes each fingerprint into two latent representations: a structure code that captures identity-related ridge patterns and a style code that represents noise, damage, and other non-identity variations. The model can improve degraded fingerprints by reconstructing images solely from the structural code, while the integration of both codes reproduces the original noisy image. This facilitates privacy-preserving data sharing, wherein only style codes are disclosed to emulate realistic degradations for benchmarking and bias research.

T. Arican et al. [86] examined the insufficient study about the interoperability of finger vein recognition across various devices. Although finger vein biometrics provide many benefits, including enhanced privacy, durability, and counterfeit resistance, identification efficacy frequently diminishes in cross-device contexts because to variations in image quality and lighting conditions. The authors present a cross-device finger vein dataset to examine this issue and as-

sess various methodologies, including a traditional technique, a convolutional neural network (CNN), and a novel patch-based convolutional autoencoder (CAE). The findings underscore the necessity for standardization in finger vein systems to provide dependable interoperability, akin to that of fingerprints and iris biometrics. The suggested CAE exhibits commendable performance in difficult cross-device scenarios, indicating its potential to enhance resilience in practical biometric applications.

C.-C. Wu et al. [87] introduced a convolutional neural network autoencoder (CNN-AE) to enhance gait recognition utilizing plantar pressure data, specifically for open-set recognition involving the identification of unknown individuals. The autoencoder compresses foot pressure maps into a 64-dimensional latent feature vector, thereafter utilized for identity comparison using feature distances. This representation more effectively differentiates authorized individuals from unauthorized ones compared to conventional CNN-only methods. The model is assessed under standard walking settings and while subjects bear additional weight (500 g), demonstrating robust performance although a significant decline under load variation. The CNN-AE enhances resilience in gait-based authentication and facilitates compact feature encoding for fast classification in open-set contexts.

3.5.5 GAN-Based biometric systems

GANs have also been applied to various biometric modalities, including fingerprints, irises, and faces, using architectures such as attention-enhanced DCGANs, StyleGAN variants, CycleGAN, and disentanglement-based GANs. In this section, we present several related studies illustrating such applications across different modalities.

Starting with fingerprint studies, S. K. Abbas et al. [88] introduced a GAN-based framework for producing high-quality synthetic fingerprint datasets, aiming to mitigate issues with privacy, cost, and restricted data availability. The methodology integrates StyleGAN2-ADA and StyleGAN3 to create authentic live fingerprints based on designated finger types, with CycleGAN to generate equivalent spoof fingerprints that mimic diverse attack materials. The produced datasets (DB2 and DB3) encompass several impressions for each finger, exhibiting significant realism and efficacy, attaining elevated matching accuracy while safeguarding privacy, with no indications of identity compromise. The methodology offers a scalable and privacy-preserving framework for the training and assessment of fingerprint recognition and

spoof detection systems. J. Ma and X. Chen [89] examined the security of DL-based fingerprint identification systems by assessing their susceptibility to synthetic fingerprint attacks. The authors proposed an attention-enhanced Deep Convolutional Generative Adversarial Network (AE-DCGAN) to produce highly realistic fingerprint images, subsequently employed to evaluate a Siamese network for matching purposes. The findings indicate that the proposed GAN enhances image quality relative to conventional DCGANs, however also exposes possible vulnerabilities in Siamese matching system, as synthetic fingerprints may result in misidentification. The findings underscores significant security vulnerabilities and the necessity for more resilient biometric systems.

Regarding iris-oriented proposals, A. Kordas et al. [90] explored the use of StyleGAN3 to produce high-quality synthetic iris images in both Cartesian and polar formats, directly aligned with iris recognition systems. A total of 1327 unique synthetic irises were generated and assessed using standard recognition algorithms, indicating that they possess adequate distinctive characteristics for identification while maintaining privacy, with no identity leakage detected. The findings underscore the significant potential of synthetic iris data to enhance datasets, address privacy limitations, and bolster the resilience and efficacy of iris recognition systems, particularly by facilitating the development of extensive and varied training data. S. Yadav and A. Ross [91] introduced iWarpGAN, a GAN-based system designed to generate synthetic iris images, addressing prevalent issues such as limited variety and resemblance to training data. The approach separates identity and style via two transformation routes, facilitating the creation of images with both inter-class (different identities) and intra-class (same identity, varying styles) variants. Experimental findings indicate that the produced images are both realistic and varied, and their incorporation in training enhances the efficacy of iris recognition systems. The methodology improves data diversity and scalability for biometric applications.

Ending with face modality, A. R. P N et al. [92] presented MLSD-GAN, a StyleGAN-based technique for producing high-quality face morphing attacks capable of deceiving facial recognition systems (FRS). The method employs disentangled latent representations and spherical interpolation to generate realistic and varied morphing faces. Experiments demonstrate that MLSD-GAN markedly enhances the susceptibility of contemporary FRS models, surpassing conventional morphing methods and presenting a substantial security threat, particularly in unsupervised enrollment contexts.

3.5.6 Lightweight and explainable architectures

While large-scale supervised networks dominate the literature, several alternative DL architectures offer distinctive advantages for biometric applications where labeled data are limited, computational resources are constrained, or interpretability is valued.

As previously discussed, **PCANet** represents a streamlined DL approach that cascades principal component analysis (PCA) filters through multiple stages, followed by binary hashing and histogram-based feature pooling. Unlike conventional CNNs that learn filters through backpropagation, PCANet computes filters as the principal components of training patches, producing an architecture that requires minimal parameter tuning and no iterative optimization. The unsupervised nature of PCANet offers particular advantages in biometric applications where labeled training data may be limited or where generalization across diverse acquisition conditions is required. By learning filters that capture the dominant variations within training data, PCANet produces representations that emphasize statistically significant patterns while suppressing noise and irrelevant variations. The architecture has been successfully applied to palmprint and other hand-based biometrics, demonstrating competitive performance with substantially simpler training procedures. Belhocine et al. [93] demonstrated that PCANet-based feature extraction, when combined with deep rule-based classification, achieves excellent identification rates for contactless palmprint recognition. In a related work on FKP recognition, PCANet-based feature learning combined with a linear multiclass SVM classifier and a score-level fusion strategy was shown to achieve high recognition accuracy and strong robustness using multimodal fusion of finger knuckle images. In a related work by R. Chlaoua et al. [94], a PCANet-based FKP recognition system combined with a linear multiclass SVM classifier and score-level fusion was proposed, achieving high recognition accuracy and demonstrating the effectiveness of multimodal fusion strategies.

ICANet extends the PCANet framework by replacing PCA with independent component analysis (ICA) for filter learning, pursuing representations that maximize statistical independence rather than variance. This shift in objective function reflects a different conception of discriminative features: where PCA identifies directions of maximum variance that may capture both signal and noise, ICA seeks components that are maximally independent, potentially isolating distinct generative factors underlying biometric images. ICANet may excel when the underlying data distribution exhibits non-Gaussian structure that PCA fails to capture optimally.

Transform-based networks including DCTNet and DSTNet employ fixed transforms - discrete cosine transform (DCT) and discrete sine transform (DST)- as filter banks. These approaches completely eliminate the data-dependent filter learning stage, instead relying on the proven image representation capabilities of frequency-domain transforms. Andén and Mallat (2014) developed the deep scattering spectrum, a transform-based approach that cascades wavelet transforms with nonlinear modulus operations to create representations invariant to translations and deformations. For biometric applications, transform-based networks offer the compelling advantage of immediate deployment without training data, making them particularly suitable for scenarios where labeled data are scarce or where computational resources for training are limited. While typically yielding slightly lower accuracy than data-dependent counterparts, transform-based networks provide predictable computational requirements and guaranteed performance independent of training data characteristics.

Sparse autoencoders offer an alternative approach to unsupervised feature learning that emphasizes reconstruction-based representation. A sparse autoencoder comprises an encoder network that maps input data to a latent representation, and a decoder network that reconstructs the original input from this representation. By constraining the latent representation to be sparse - containing relatively few active units - the network learns to capture the most salient features of the input distribution in a compact, interpretable form. For biometric applications, sparse autoencoders offer the advantage of learning representations that can be interpreted as indicating which basis features are present in a given biometric sample, providing a degree of explainability lacking in conventional CNN features. Bachay and Abdulameer [12] demonstrated that hybrid deep learning models combining autoencoders with CNNs achieved superior performance for palmprint authentication, showing the complementary strengths of reconstruction-based and discriminative learning. The learned sparse representations have been applied to various biometric modalities, often serving as initializations for supervised fine-tuning or as compact templates for efficient storage and matching.

3.6 Conclusion

This chapter has provided the necessary theoretical foundations for understanding the role of DL in modern biometric systems. It first introduced the fundamental principles of artificial neural networks, before presenting the main deep learning architectures commonly used in bio-

metric applications. The chapter also highlighted the advantages of DL as well as the challenges associated with conventional deep CNN models. To address these limitations, lightweight and interpretable approaches such as PCANet and xDNN were discussed as promising alternatives. Finally, a review of recent related works demonstrated the growing importance and diversity of DL techniques in the field of biometrics, thereby establishing the context for the methods and contributions presented in the following chapters.

Chapter 4

Explainable xDNN-PCANet with Illumination-Invariant Features for FKP Recognition

4.1 Introduction

Due to the swift progression of information technology, biometric systems have become a dependable substitute for conventional identification methods, such ID cards, passwords, or PIN codes. These methods enable the identification of people by utilizing their distinct physiological and behavioral traits [95, 96]. Consequently, numerous biometric modalities have been extensively examined, including fingerprint, iris, ear, palmprint, facial recognition, and Finger Knuckle Print (FKP) [97, 98, 99]. Among them, finger knuckle print (FKP), a hand-based biometric trait, has garnered considerable attention owing to its unique anatomical structures, intricate textural features, high stability, and robust discriminative ability. It may be acquired using inexpensive, compact imaging devices without the need for further gear, rendering it exceptionally practical for real-world applications [100, 101]. However, despite these advantages, existing FKP recognition systems still face two major challenges: their sensitivity to illumination variations, which can significantly degrade recognition accuracy, and the absence of interpretability in deep learning models, which often operate as "black-box" systems without providing clear explanations for their decisions. These limitations reduce the reliability

and trustworthiness of biometric systems, particularly in security-critical applications where transparent and robust decision-making is essential.

Recently, explainable deep neural networks have gained popularity to improve the transparency and interpretability of intricate learning models. In contrast to conventional methodologies, these approaches elucidate the decision-making process, which is especially critical in sensitive contexts like biometric identification. To improve the reliability and robustness of biometric decision-making, the proposed framework integrates an xDNN classifier for user identification. This approach relies on prototype-based reasoning together with feature-relevance visualization, which makes the decision process more transparent and easier to interpret. As a result, it not only supports accurate classification but also enhances the explainability and trustworthiness of the obtained categorization outcomes.

The proposed framework jointly leverages robust feature extraction and interpretable decision-making to achieve improved performance in FKP recognition. In the initial stage, the Self-Quotient Image (SQI) technique is utilized to decompose input FKP images into reflectance components that are invariant to illumination changes, thereby reducing the impact of lighting variations and ensuring more consistent and reliable input representations for further processing. Then, a two-stage PCANet is applied to extract deep and discriminative hierarchical representations from these enhanced components, benefiting from its computational efficiency and strong ability to capture both structural information and local texture patterns.

To strengthen the interpretability of the overall system, an xDNN classifier is incorporated within the pipeline, enabling prototype-based reasoning while also providing visual insights into feature relevance. This integration leads to a decision-making process that is both more transparent and inherently explainable. The complete framework is rigorously assessed on the PolyU FKP database using four distinct finger modalities (left index, left middle, right index, and right middle), each evaluated independently under a unimodal setting. Moreover, the main parameters of both SQI and PCANet are carefully optimized through extensive ablation experiments, resulting in identification accuracies ranging from 91% to 96%. Overall, the experimental findings demonstrate consistent improvements over existing state-of-the-art methods while preserving interpretability, thereby confirming the suitability of the proposed approach for security-critical and real-world biometric applications.

In this chapter, we present the xDNN framework, a learning approach designed to combine high predictive performance with transparent and interpretable decision-making. While conventional deep learning models often achieve strong accuracy, they usually operate as black-box systems, making it difficult to understand how predictions are produced an important limitation in security critical applications.The xDNN model addresses this issue by incorporating explainability inside its architecture, allowing users to track, examine, and validate the rationale behind each classification choice. This chapter therefore describes the main principles of the xDNN framework, including its adaptive feature representation and the generation of interpretable decision rules that link input data to the final output.

4.2 The proposed biometric system-based XDNN classifier architecture

The proposed methodology is built around an image-based explainable deep learning framework designed for FKP biometric recognition. The main goal of this framework is not only to achieve reliable recognition performance but also to ensure that the decision-making process remains transparent and interpretable. Unlike many conventional deep learning approaches that operate as black-box models, the adopted xDNN architecture provides a structure where each stage of the learning and classification process can be understood and interpreted by human users. This transparency is particularly important in biometric systems, where trust, traceability, and the ability to justify decisions are essential. As illustrated in Figure 4.1, the proposed biometric recognition system is organized into four principal stages, starting from image acquisition and ending with explainable classification. Each stage contributes to improving the quality of the biometric data and enhancing the reliability of the recognition process.

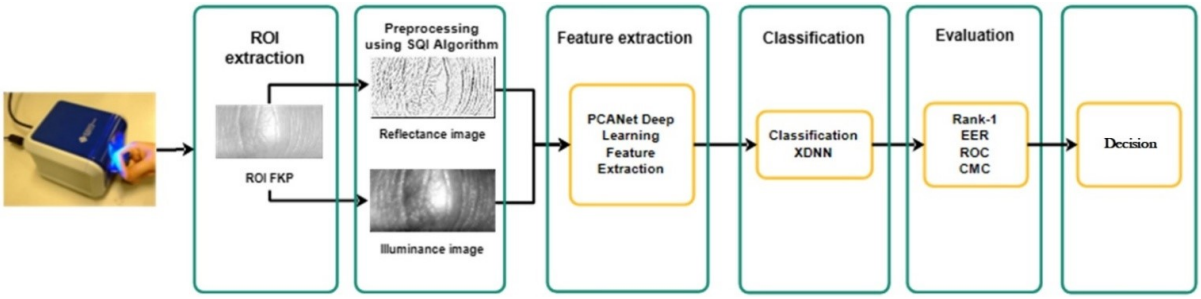


Figure 4.1: Architecture of the proposed biometric system.

- **Step 1: Image acquisition and normalization**

The first stage involves collecting the FKP images from the selected biometric databases. Since raw images may vary in terms of scale, orientation, or position, a normalization procedure is applied. This operation aligns the images and standardizes their size and orientation so that all samples share a consistent representation. Such normalization is essential because it reduces geometric variations between images and allows the subsequent processing stages to focus on the actual biometric patterns rather than irrelevant differences.

- **Step 2: Illumination-invariant preprocessing**

  After normalization, the images undergo a preprocessing stage aimed at reducing the influence of illumination conditions. Variations in lighting during image acquisition can significantly affect the appearance of FKP patterns and may degrade recognition accuracy. To address this issue, the SQI algorithm is applied. This method decomposes the original image into reflectance components, allowing the system to emphasize intrinsic texture information while suppressing lighting effects. As a result, the biometric features become more stable and robust when images are captured under different illumination conditions.

- **Step 3: Hierarchical feature extraction**

  Once the images are preprocessed, the system extracts meaningful features that describe the distinctive structures present in the FKP patterns. This task is performed using a PCANet-based deep learning framework. PCANet relies on principal component analysis to learn filters directly from the training data, enabling the model to capture discriminative patterns in an efficient manner. Through a hierarchical structure, PCANet progressively transforms the input images into feature representations that highlight relevant texture and structural information, which is crucial for accurate biometric identification

- **Step 4: Clarification of classification through xDNN**

  In the last stage, the retrieved features are supplied to the xDNN classifier, which executes the recognition task while preserving interpretability. The xDNN architecture consists of multiple conceptual levels that outline the data analysis process and how classification decisions are made:

  - **Feature layer**: This layer represents the high-level feature vectors generated during

feature extraction and provides input to the classification stage.

- **Density layer**: At this level, the distribution of the feature vectors is analyzed in order to identify representative data regions and prototypes that summarize the training samples.

- **Typicality layer**: This layer evaluates how similar a new input sample is to the learned prototypes, measuring the degree to which the sample can be considered typical of a particular class.

- **Mega-cloud layer**: Finally, multiple related prototypes are grouped into larger structures known as mega clouds, which represent broader class patterns and support the final classification decision.

By combining these layers, the xDNN classifier produces decisions that can be traced back to specific prototypes and data structures. This prototype-based reasoning enhances the transparency of the system and allows users to understand why a particular biometric sample has been assigned to a given identity.

4.2.1 ROI extraction

The extraction of ROI from FKP images is performed through a sequence of preprocessing and localization operations designed to isolate the most informative knuckle area. The original image is initially smoothed with a Gaussian filter to mitigate noise and minor intensity fluctuations that could impact edge recognition. The filtered image is then downsampled to a resolution of approximately 150 dpi to decrease computational complexity while preserving the main structural patterns of the knuckle.Subsequent to this preparatory phase, the horizontal reference axis (X-axis) is defined by identifying the bottom border of the finger by the Canny edge detection technique [102], which discerns prominent intensity gradients and delineates the finger shape. Based on this horizontal alignment, a sub-region of the image is selected and processed again with the same edge detection technique to determine the vertical orientation. A convex direction coding strategy is then applied to analyze the orientation of the detected contours and define the vertical reference axis (Y-axis). Once both axes are established, the system extracts a rectangular region centered on the knuckle area, which contains the most discriminative texture patterns used for recognition. This normalized ROI ensures consistent

alignment of FKP samples and improves the robustness of subsequent feature extraction and classification stages. The ROI is identified and defined as a rectangular region, as illustrated in Figure 4.2.

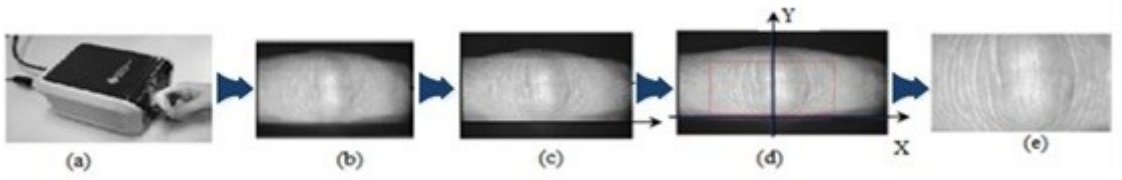


Figure 4.2: The process of extracting the FKP ROI.

4.2.2 Feature extraction

Feature extraction is an essential phase in pattern recognition systems, as the efficacy of the final classification stage significantly relies on the successful acquisition of pertinent information from the data. The choice of a suitable feature extraction approach is crucial, since it influences the system's capacity to depict the inherent patterns in the images. In addition, the discriminative power and stability of the extracted features are essential for accurately separating different classes and ensuring reliable recognition results. For this reason, the proposed approach employs a combination of the SQI technique and the PCANet framework to generate feature vectors from each FKP image. The SQI method enhances the robustness of the images by reducing illumination variations, while PCANet learns meaningful representations that capture the structural characteristics of the knuckle patterns, providing informative features for the subsequent classification stage [103][104, 105].

4.2.2.1 Self-Quotient Image Algorithm

The SQI technique, originally proposed by Wang Wei et al. [106], is designed to reduce the influence of illumination variations in images. The method aims to separate the intrinsic reflectance information of an image from the lighting conditions under which it was captured. By emphasizing reflectance while suppressing illumination effects, SQI improves the robustness of image-based recognition systems, particularly in biometric applications where lighting conditions may vary. A primary benefit of this approach is its ability to normalize lighting differences while preserving important structural details in the image. The SQI technique combines concepts from edge-preserving filtering with the illumination normalization principle

derived from the Retinex theory, allowing the algorithm to smooth illumination changes while maintaining significant texture patterns. The SQI computation generally involves two main stages: illumination estimation and illumination normalization. First, an approximation of the illumination component of the image is obtained by applying a smoothing operation that removes high-frequency texture details while keeping the general lighting distribution. After estimating the illumination component, the original image is normalized through division by its smoothed version. Denoting the original image by I and the resulting self-quotient image by Q, the relationship can be expressed as:

$$Q = \frac{I}{K} * I \tag{4.1}$$

where K represents a smoothing kernel used to estimate the illumination component, and the symbol $*$ denotes the convolution operation. The division is performed pixel by pixel, similar to the operation used in classical quotient image techniques. Unlike traditional quotient image approaches that necessitate several pictures of the same topic under varying lighting circumstances, the SQI method generates the quotient image directly from the input image itself, enhancing its practicality for real-world applications. Through this normalization process, the resulting image emphasizes reflectance features while significantly reducing the impact of illumination variations, thereby yielding more stable representations for subsequent stages such as feature extraction and classification. The PCANet framework, as illustrated in Figure 4.3, is briefly described as follows [61, 62]:

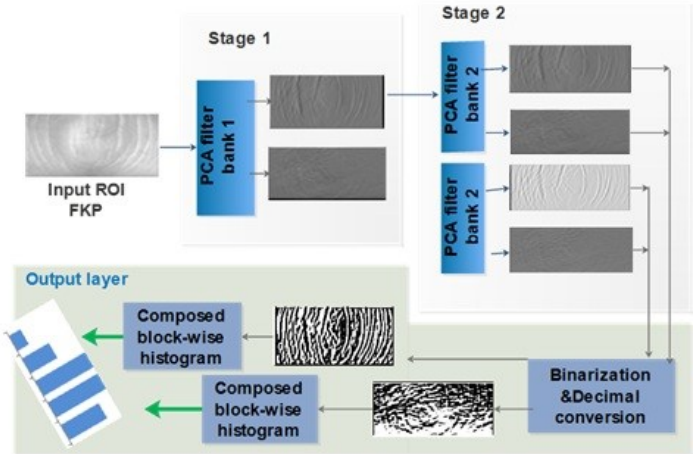


Figure 4.3: Feature extraction from FKP images using a two-stage PCANet scheme.

4.2.2.2 PCANet-based deep learning

Within the field of deep learning for image analysis, PCANet has emerged as a lightweight and efficient architecture for feature extraction. The method was introduced by Tsung-Han Chan and colleagues as a simple deep learning baseline that relies on classical statistical techniques instead of complex optimization procedures. Unlike traditional convolutional neural networks that often demand substantial labeled data and extensive parameter tuning, PCANet adopts a more straightforward learning mechanism based on PCA. This design makes the training process significantly easier while still enabling the extraction of discriminative image features. The PCANet framework is built around three main operations: cascaded PCA filtering, binary hashing, and histogram-based feature encoding. The first stage of PCANet involves the construction of PCA filter banks. In this step, small patches of size $k_1 \times k_2$ are extracted around each pixel of the input FKP image and rearranged into vectors. The mean value of each vector is computed and subtracted to normalize the data. Afterwards, PCA is applied to these vectors in order to identify the principal components representing the most prominent variations in the local image structure. The leading eigenvectors obtained from PCA are then reshaped into convolution kernels that serve as filters. Applying these filters to the input image via convolution produces a set of filtered images. If h_l represents a PCA-derived filter and I(x,y) the input image, the filtering operation can be written as:

$$T_l(x,y) = h_l(x,y) * I(x,y) \tag{4.2}$$

where $*$ denotes the discrete convolution operation. This process generates multiple feature maps corresponding to the principal components retained during PCA. In the second stage, the same procedure is repeated on the output images produced by the first stage, leading to a deeper hierarchical representation. New PCA filters are estimated from the intermediate feature maps, and convolution is again performed to generate additional feature responses:

$$I_{l,m}(x,y) = h_m(x,y) * I_l(x,y), \quad i \in [1 \dots l_{s2}] \tag{4.3}$$

Through this two-stage filtering mechanism, PCANet progressively captures more abstract structural characteristics of the FKP patterns. Following the convolution stages, the result-

ing feature maps are converted into a binary representation through a binary hashing process. A step function is applied to each pixel value in the feature maps so that positive responses are mapped to 1, while negative values are mapped to 0. The resulting binary maps are then combined by interpreting the set of binary bits around each pixel as a decimal number. This transformation compresses multiple feature maps into a single integer-valued image that summarizes the responses of all filters. The final step of PCANet consists of histogram-based feature encoding. Each integer-valued image is divided into several spatial blocks, which may either overlap or remain separate depending on the configuration. For every block, a histogram is calculated to describe the distribution of pixel values within that region. These histograms capture local structural information and are concatenated together to form the final feature vector representing the input image. Formally, if $B^h ist$ denotes the histogram computed for the $i^t h$ block, the feature representation of a filtered image can be expressed as:

$$I_{l,m}^{B}(i,j) = \begin{cases} 1, & \text{if } I_{l,m}(i,j) \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.4}$$

$$I_{l}^{D} = \sum_{m=1}^{L_{s2}} 2^{m-1} I_{l,m}^{B}(i,j) \tag{4.5}$$

where I_l^D denotes the hashed image, whose pixel values are integer values in the range $[0, 2^{LS2-1}]$.

The complete descriptor of the input FKP image is then obtained by concatenating the histograms produced for all output feature maps. The effectiveness of PCANet depends on several parameters that control the feature extraction process. These include the filter size, the number of filters per stage, the number of network stages, the size of histogram blocks, and the overlap ratio between neighboring blocks. In this work, these parameters were empirically selected in order to obtain discriminative representations of the FKP images while maintaining computational efficiency. Specifically, the network uses two stages, with two filters in each stage, a filter size of 7×7 , histogram blocks of size 21×21, and an overlap ratio of 75% between neighboring blocks. These settings allow the PCANet framework to effectively capture the distinctive texture patterns present in the finger-knuckle region, which are then used for the subsequent classification stage.

4.2.3 Validation of the xDNN Classifier

xDNN is a supervised learning methodology designed to ensure dependable pattern identification while maintaining the clarity of the decision-making process. In contrast to traditional deep learning models that frequently function as opaque "black-box" systems, the xDNN framework offers an interpretable architecture that allows for the tracing and comprehension of the rationale behind each classification choice. This study utilizes the xDNN classifier for FKP recognition. The model generates 165 class representations, each representing a permitted subject in the database. Each individual is characterized by six feature vectors derived from the PCANet-based feature extraction phase. These feature vectors represent different types of FKP images, including left index finger (LIF), left middle finger (LMF), right index finger (RIF), and right middle finger (RMF). The architecture of xDNN integrates hierarchical pattern analysis with transparent rule-based reasoning. Instead of relying solely on complex nonlinear transformations, the classifier organizes knowledge in the form of prototypes that represent typical samples within each class. During classification, the system evaluates similarities between the input feature vector and these stored prototypes. This prototype-based strategy allows the decision process to be interpreted and verified by human analysts. Moreover, the system performs similarity evaluation at both local (within-class) and global (across classes) levels before assigning the final label.

The validation stage of the xDNN classifier is organized into four successive layers:

- **Feature descriptor layer** At this stage, the feature representation of the input sample is generated using the same procedure applied during the training phase. The resulting feature vector constitutes the input to the classification module.

- **Prototype layer** In this layer, the similarity between the input feature vector x and each prototype p_i belonging to a given class is calculated. The similarity measure reflects how closely the new sample resembles the stored representative patterns. This similarity can be expressed as:

$$S(x, p_i) = \frac{1}{1 + \frac{(x - p_i)}{\sigma_i^2}} \tag{20}$$

  where σ denotes the variance associated with the prototype distribution. This measure

provides an indication of how typical the input sample is with respect to the learned prototypes.

- **Local decision layer (per class)**

  For each class, the system identifies the prototype that exhibits the highest similarity with the input sample. This process follows a winner-takes-all principle, where the maximum similarity value among the prototypes of the same class is selected:

$$\lambda_c = \max_{j=1,2,\ldots,P} \left(S_j\right), \quad \text{for } j = 1 \text{ to } P \tag{21}$$

  where P represents the number of prototypes associated with a particular class.

- **Global decision layer** After determining the best similarity score for each class, the classifier compares these scores across all classes. The final classification corresponds to the class with the highest similarity value:

$$\lambda_c^* = \max_{c=1,2,\ldots,C} (\lambda_c), \quad \text{for } c = 1,2,\ldots,C \tag{4.6}$$

  where C represents the total number of classes. The predicted label for the input image is then determined as:

$$\text{label} = \arg \max_{c=1,2,\ldots,C} (\lambda_c^*) \tag{4.7}$$

  Through this multi-layer validation process, the xDNN classifier ensures that the classification decision is supported by measurable similarities with representative prototypes. This architecture offers several important advantages, including the generation of interpretable rules, the evaluation of confidence at both local and global levels, and a mathematically transparent decision mechanism, which facilitates human understanding and verification of the recognition results.

4.3 Experimental Results and Discussion

4.3.1 Dataset

The performance of the proposed FKP-based recognition system is rigorously evaluated using the PolyU FKP Database, which was originally developed and subsequently made publicly accessible by The Hong Kong Polytechnic University for research purposes. This well-known and widely adopted benchmark dataset contains FKP images collected from a total of 165 distinct subjects, including 125 male and 40 female participants, and is extensively used in biometric research due to its richness, diversity, and overall data consistency. Among these individuals, 143 subjects fall within the age range of 20 to 30 years, while the remaining participants are distributed within the 30 to 50 age interval. For each subject, images are systematically captured from four different finger knuckle regions, namely LIF, LMF, RIF, and RMF, ensuring multi-modal coverage of biometric traits. Each finger modality includes 12 samples per individual, resulting in a total of 1980 images for each of the four FKP categories. For experimental evaluation, the dataset is carefully partitioned into two separate and strictly non-overlapping sessions, where six images per subject are used for training the model, while the remaining six images are reserved exclusively for testing and performance validation.

4.3.2 Configuration of SQI and PCANet Parameters for Decision Making in xDNN

The objective of this phase is to determine the optimal configuration of parameters for the SQI algorithm and the PCANet descriptor in the context of FKP-based individual authentication. Careful tuning of these parameters is essential, as they directly influence the performance of the SQI preprocessing stage as well as the effectiveness of PCANet feature extraction, thereby improving the overall classification accuracy of the xDNN model. This section analyzes the effect of several important parameters in the SQI and PCANet stages, including the standard deviation, block overlap ratio R_a, filter sizes (k_1, k_2), block size b, and number of filters N. The study is structured to evaluate how each factor influences the overall system performance. Experiments are conducted independently for each FKP modality (RMF, LMF, LIF, and RIF) to provide a detailed and fair assessment across all finger types.

4.3.3 Experimental results

4.3.3.1 Results for RMF modality

In this experiment, all remaining parameters are intentionally kept constant to ensure a fair, consistent, and controlled comparison, including the block overlap ratio $R_a = 75\%$, the filter size $(k_1, k_2) = (13 \times 13)$, the block size $b = [40, 40]$, and the number of filters $N = [5, 5]$. The system performance is evaluated using widely recognized biometric metrics, namely the Rank-1 identification rate, the equal error rate (EER), and the verification rate (VR) measured at false acceptance rate (FAR) levels of 1% and 0.1%.

Table 4.1: Performance of xDNN under different standard deviation (σ) settings in the RMF modality.

Standard Deviation	Identification RANK-1%	EER%	Verification VR@ 1% FAR%	Verification VR@ 0.1% FAR%
1	95.96%	3.34%	96.06%	91.92%
3	94.85%	3.53%	95.05%	92.12%
5	94.55%	3.23%	95.45%	92.02%
7	93.43%	3.64%	94.24%	89.80%
9	92.32%	4.46%	92.83%	87.78%

Table 4.1 indicates that the optimal performance is reached when $\sigma = 1$, achieving a Rank-1 accuracy of 95.96%, an EER of 3.34%, and verification rates of 96.06% and 91.92% at FAR levels of 1% and 0.1%, respectively. Although slightly better results are observed for some metrics with other values of σ, the improvements remain marginal. Therefore, $\sigma = 1$ is selected as it provides the most consistent performance. The CMC and ROC curves presented in Fig. 4.4 further confirm this choice.

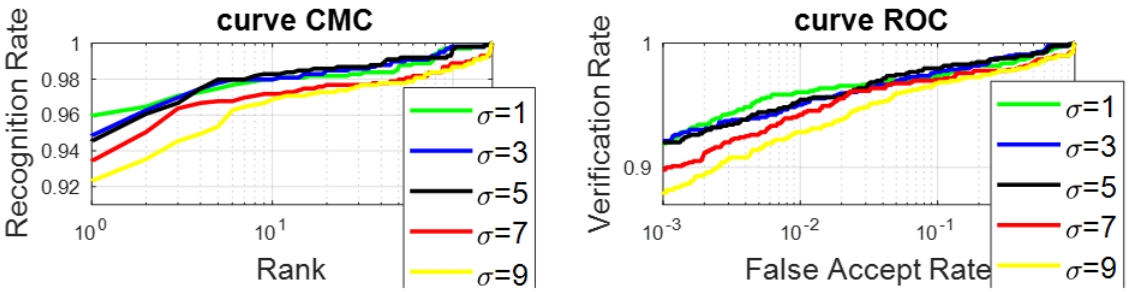


Figure 4.4: CMC and ROC curves for the RMF modality obtained using xDNN under different standard deviation values.

In the next step, the parameters of the PCANet descriptor are optimized. For this purpose, the standard deviation of the SQI algorithm is fixed at $\sigma = 1$, and the influence of the PCANet

parameters is investigated. The analysis focuses on four parameters: the overlap ratio R_a, the filter size (k_1, k_2), the block size b, and the number of filters N.

- The overlap ratio (R_a)

  With the standard deviation set to $\sigma = 1$, the block overlap ratio R_a is systematically varied between 0% (no overlap) and 75% in order to analyze its influence on performance. The corresponding results in terms of Rank-1 identification accuracy, EER, and verification rates at FAR levels of 1% and 0.1% are reported in Table 4.2.

Table 4.2: Results for the RMF modality obtained with different overlap ratios (R_a).

Overlap Ratio (R_a)%	Identification RANK-1%	EER%	Verification VR@ 1% FAR%	Verification VR@ 0.1% FAR%
0	91.01%	5.15%	91.72%	86.97%
25	94.65%	3.64%	94.85%	91.01%
50	93.23%	3.73%	94.55%	90.51%
75	94.34%	3.33%	95.35%	91.01%

As shown in Table 4.2, overlap ratios of $R_a = 25\%$ and $R_a = 75\%$ result in the best overall performance. Specifically, $R_a = 75\%$ achieves the lowest EER (3.33%) and the highest verification rate at 1% FAR (95.35%), whereas $R_a = 25\%$ attains the highest Rank-1 identification accuracy (94.65%). Both configurations also reach the best verification rate at 0.1% FAR (91.01%), demonstrating robust performance under strict operating conditions.

Figures 4.5 and 4.6 present the CMC and ROC curves of the FKP recognition system for different overlap ratios. The results indicate that the configuration with $R_a = 25\%$ provides the most favorable system performance.

- Filter size (k1, k2) selection

  For the RMF modality, the study focuses on determining the most appropriate filter size, while keeping the previously selected optimal parameters fixed, namely $\sigma = 1$ and $R_a = 25\%$. A set of filter sizes ranging from 3×3 to 15×15 is evaluated, and the corresponding results are reported in Table 4.3 in terms of Rank-1 identification accuracy, EER, and verification rates at FAR levels of 1% and 0.1%. The results indicate that the 13×13 filter achieves the best overall trade-off, with a Rank-1 accuracy of 95.45%, an EER of 3.43%,

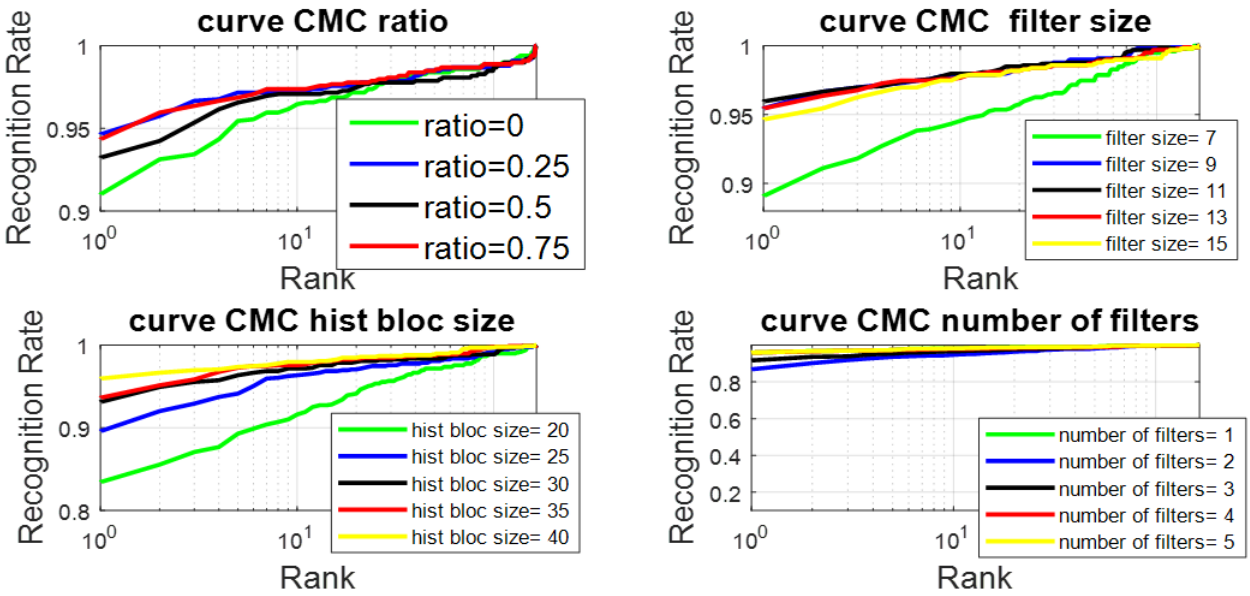


Figure 4.5: CMC curves for the RMF modality using xDNN and PCANet with varying parameters, including the overlap ratio, filter size, histogram block size, and number of filters.

and the highest verification rate at 0.1% FAR (95.86%). In comparison, although the 11×11 filter yields a slightly higher Rank-1 accuracy (95.96%), the 13×13 configuration provides superior verification performance, making it the most robust choice overall.

Table 4.3: Results for the RMF modality obtained using xDNN with different filter size configurations (k1, k2).

Filter Size	Identification RANK-1%	EER%	Verification VR@ 1% FAR%	Verification VR@ 0.1% FAR%
3×3	73.94%	13.14%	68.79%	52.93%
5×5	85.96%	8.20%	83.43%	75.15%
7×7	89.09%	6.78%	88.08%	83.54%
9×9	95.45%	3.41%	94.95%	91.52%
11×11	95.96%	3.44%	95.56%	93.13%
13×13	95.45%	3.43%	95.86%	95.86%
15×15	94.65%	3.54%	95.45%	91.01%

- The bloc Size (b) selection

  This stage is specifically devoted to identifying the optimal block size b for the RMF modality, while the remaining parameters are kept unchanged and fixed at $\sigma = 1$, $R_a = 75\%$, and filter size $(k_1, k_2) = (13 \times 13)$ to ensure a controlled evaluation setting. Block sizes in the range $[20, 20]$ to $[40, 40]$ are investigated. The performance is evaluated using widely adopted biometric metrics, namely Rank-1 identification accuracy, EER, and verification rates computed at FAR levels of 1% and 0.1%.

  As reported in Table 4.4, the $[40, 40]$ configuration yields the best overall results, achiev-

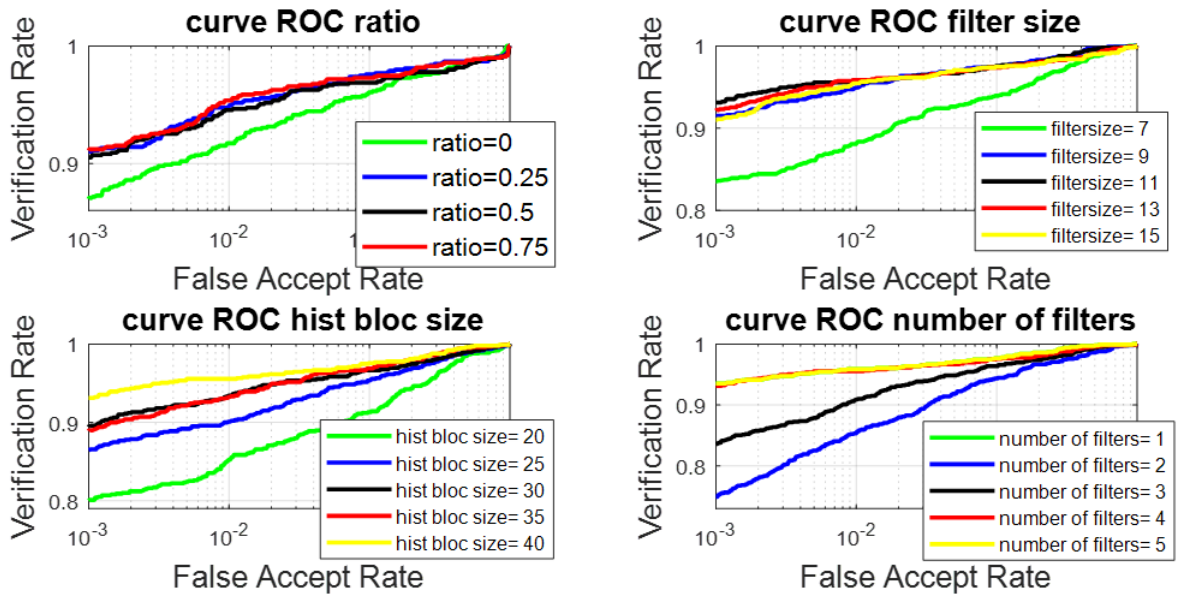


Figure 4.6: ROC curves for the RMF modality obtained using xDNN and PCANet under different parameter settings, including the overlap ratio, filter size, histogram block size, and number of filters.

ing a Rank-1 accuracy of 95.96%, an EER of 3.44%, and verification rates of 95.56% and 93.13% at 1% and 0.1% FAR, respectively.

Table 4.4: Results for the RMF modality obtained using xDNN with different block size values (b).

Block Size (h)	Identification RANK-1%	EER%	Verification VR@ 1% FAR%	Verification VR@ 0.1% FAR%
[20 20]	83.43%	8.88%	85.25%	80.00%
[25 25]	89.60%	5.55%	90.10%	86.46%
[30 30]	93.13%	4.45%	93.43%	89.49%
[35 35]	93.64%	3.94%	93.23%	88.89%
[40 40]	95.96%	3.44%	95.56%	93.13%

- Selection of the Number of Filters (N):

  The influence of the number of filters N is thoroughly examined by varying it from $[1,1]$ to $[5,5]$, while maintaining all other previously optimized parameters at fixed values in order to ensure a fair, stable, and consistent evaluation protocol, namely $\sigma = 1$, $R_a = 75\%$, filter size $(k_1,k_2) = (13 \times 13)$, and block size $b = [40,40]$. As illustrated in Table 4.5, the configuration $[5,5]$ consistently achieves superior performance across all evaluation criteria, yielding the lowest EER (3.33%), the highest verification rates (95.86% at 1% FAR and 93.54% at 0.1% FAR), along with a Rank-1 identification accuracy of 95.96%.

Table 4.5: Results for the RMF modality obtained using xDNN with different numbers of filters (N).

Number Filters (N)	Identification RANK-1%	EER%	Verification VR@ 1% FAR%	Verification VR@ 0.1% FAR%
[1 1]	61.41%	14.74%	61.01%	42.12%
[2 2]	86.87%	6.66%	85.35%	74.95%
[3 3]	91.72%	5.16%	90.91%	83.54%
[4 4]	95.96%	3.44%	95.56%	93.13%
[5 5]	95.96%	3.33%	95.86%	93.54%

4.3.3.2 Results for LMF modality

Table 4.6 summarizes the FKP recognition results for the LMF modality obtained by varying the standard deviation σ in the SQI algorithm from 1 to 9. During this evaluation, all other parameters are fixed, namely the block overlap ratio $R_a = 75\%$, filter size $(k_1, k_2) = (13 \times 13)$, block size $b = [40, 40]$, and number of filters $N = [5, 5]$.

Table 4.6: Results for the LMF modality obtained using xDNN with different standard deviation values (σ).

Standard Deviation	Identification RANK-1%	EER%	Verification VR@ 1% FAR%	Verification VR@ 0.1% FAR%
1	95.05%	3.63%	93.74%	90.20%
3	94.44%	3.71%	94.55%	89.49%
5	94.85%	3.23%	94.85%	90.40%
7	92.83%	5.26%	92.42%	88.59%
9	92.02%	5.55%	91.62%	88.48%

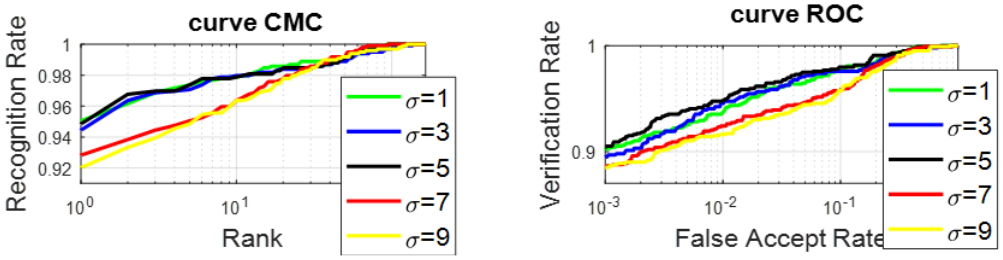


Figure 4.7: CMC curves (left) and ROC curves (right) for the LMF modality obtained using different standard deviation values.

Table 4.6 indicates that the optimal performance for the LMF modality is achieved at $\sigma = 5$, reaching a Rank-1 identification rate of 94.85%, the lowest EER of 3.23%, and verification rates of 94.85% at 1% FAR and 90.40% at 0.1% FAR. The corresponding CMC and ROC curves in Fig. 4.7 further validate that $\sigma = 5$ provides the best overall performance. Consequently, $\sigma = 5$ is retained for the LMF modality, and the subsequent analysis focuses on optimizing the

PCANet descriptor parameters, namely the overlap ratio R_a, filter size (k_1,k_2), block size b, and number of filters N.

- The overlap ratio (R_a)

  This section focuses on determining the optimal block overlap ratio R_a. Values from 0% (no overlap) up to 75% are evaluated, and system performance is measured using Rank-1 identification accuracy, EER, and verification rates at FARs of 1% and 0.1% (Table 4.7). The results show that $R_a = 75\%$ provides the best overall performance, achieving the lowest EER (4.57%) and the highest verification rate at 0.1% FAR (90.71%), while also maintaining strong identification accuracy (93.33%) and verification at 1% FAR (93.33%). This indicates that a 75% overlap ratio delivers the most balanced and effective performance across metrics.

Table 4.7: Results for the LMF modality obtained using xDNN with different overlap ratio values (R_a).

Block Overlap Ratio (*Ra*)%	Identification RANK-1%	EER%	Verification VR@ 1% FAR%	Verification VR@ 0.1% FAR%
0	91.62%	5.77%	91.41%	85.66%
25	93.43%	4.67%	93.64%	89.60%
50	92.53%	5.26%	92.63%	88.69%
75	93.33%	4.57%	93.33%	90.71%

Figures 4.8 and 4.9 show the CMC and ROC curves of the FKP recognition system for different block overlap ratios. The results indicate that $R_a = 75\%$ yields the best overall system performance.

- Filter size

  At this stage, the previously determined parameters are set, with $\sigma = 5$ and $R_a = 50\%$. The focus is now on identifying the most effective filter size, which is tested across a range from 3×3 to 15×15. The corresponding system performance for each filter size is summarized in Table 4.8.

  Table 4.8 shows that a filter size of 15×15 provides the most balanced and robust performance for the system. It achieves the highest Rank-1 identification accuracy (95.86%) and a low EER (3.83%), with verification rates of 94.65% at 1% FAR and 90.71% at 0.1% FAR. While the 11×11 filter slightly improves the EER (3.54%) and the 13×13

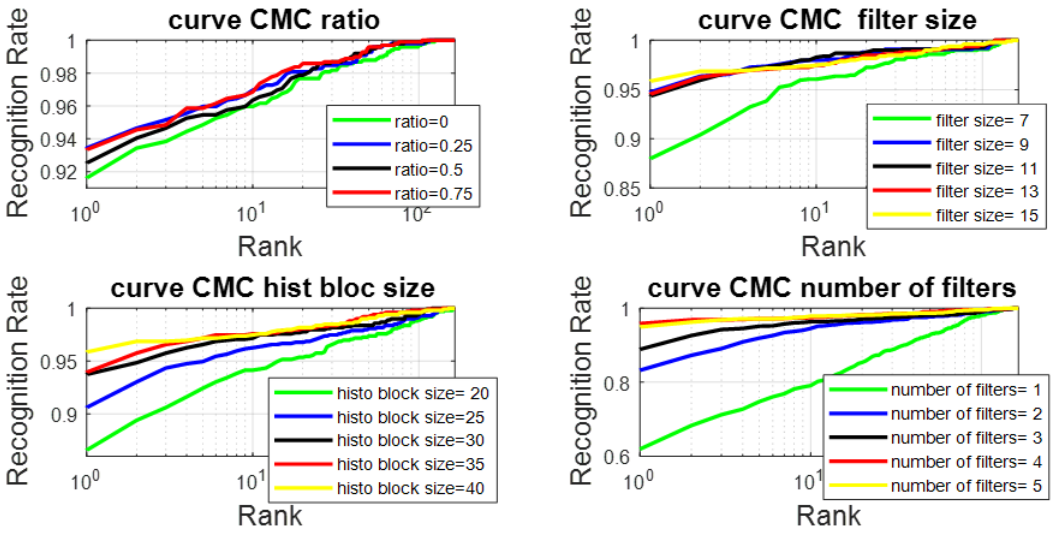


Figure 4.8: CMC curves for the LMF modality obtained using xDNN and PCANet under different parameter settings, including the overlap ratio, filter size, histogram block size, and number of filters.

Table 4.8: Results for the LMF modality obtained using xDNN with different filter size configurations (k1, k2).

Filter Size	Identification RANK-1%	EER%	Verification VR@ 1% FAR%	Verification VR@ 0.1% FAR%
3×3	72.02%	12.95%	66.06%	52.42%
5×5	83.64%	9.16%	81.31%	71.01%
7×7	87.98%	6.88%	88.08%	80.51%
9×9	94.75%	3.74%	94.14%	88.59%
11×11	94.34%	3.54%	94.75%	90.10%
13×13	94.55%	4.15%	93.74%	91.01%
15×15	95.86%	3.83%	94.65%	90.71%

filter gives a marginally higher verification rate at 0.1% FAR (91.01%), the 15×15 configuration delivers the most consistent performance across all metrics.

- The block size (b) selection

  In this stage, the previously optimized parameters $\sigma = 5$, $R_a = 50\%$, and filter size $(k_1, k_2) = (15 \times 15)$ are fixed, and the focus is on selecting the optimal block size for PCANet. Block sizes from $[20,20]$ to $[40,40]$ are evaluated, with results summarized in Table 4.9. The $[40,40]$ block size achieves the best overall performance, yielding a Rank-1 identification accuracy of 95.86%, an EER of 3.83%, and verification rates of 94.65% at 1% FAR and 90.71% at 0.1% FAR. This configuration provides a strong balance between spatial resolution and statistical stability, making it the most effective choice among the tested block sizes.

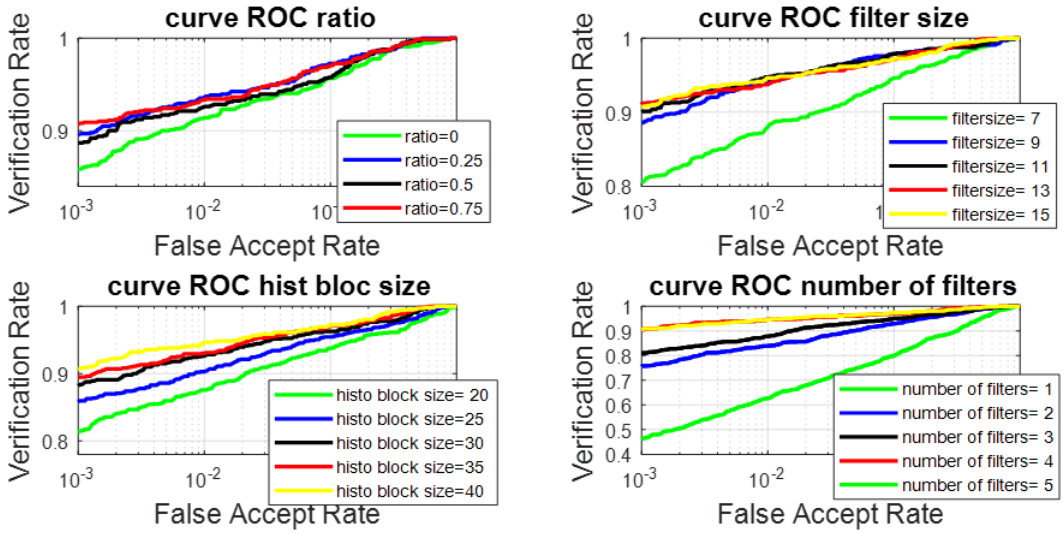


Figure 4.9: ROC curves for the LMF modality obtained using xDNN and PCANet under different parameter settings, including the overlap ratio, filter size, histogram block size, and number of filters.

Table 4.9: Results for the LMF modality obtained using xDNN with different block size values (b).

Block Size (b)	Identification RANK-1%	EER%	Verification VR @ 1% FAR%	Verification VR @ 0.1% FAR%
[20 20]	86.57%	6.96%	87.58%	81.41%
[25 25]	90.61%	5.56%	90.30%	85.76%
[30 30]	93.74%	4.75%	92.63%	88.28%
[35 35]	93.94%	4.23%	93.03%	89.39%
[40 40]	95.86%	3.83%	94.65%	90.71%

- The number of filters

  The impact of varying the number of filters N within the PCANet framework was investigated by considering configurations ranging from $[1,1]$ to $[5,5]$. During this analysis, all previously optimized parameters were kept unchanged to ensure a consistent and fair comparison, namely $\sigma = 5$, $R_a = 50\%$, filter size $(k_1,k_2) = (15 \times 15)$, and block size $b = [40,40]$. The corresponding performance results for each configuration are reported in Table 4.10.

  Among the evaluated settings, the $[4,4]$ configuration achieves the best overall performance, attaining a Rank-1 identification accuracy of 95.86%, an EER of 3.83%, and verification rates of 94.65% at 1% FAR and 90.71% at 0.1% FAR. Although the $[5,5]$ configuration yields slightly lower accuracy (94.95%) and a marginally higher EER (3.92%), its performance remains competitive and close to the optimal setting. Therefore, the $[4,4]$

configuration can be considered the most suitable choice, offering an effective trade-off between high recognition performance and computational efficiency.

Table 4.10: Results for the LMF modality obtained using xDNN with different numbers of filters (N).

Number Filters (N)	Identification RANK-1%	EER%	Verification VR @ 1% FAR%	Verification VR @ 0.1% FAR%
[1 1]	61.92%	15.57%	62.83%	46.36%
[2 2]	83.23%	7.98%	83.94%	75.56%
[3 3]	88.89%	6.15%	87.78%	80.91%
[4 4]	95.86%	3.83%	94.65%	90.71%
[5 5]	94.95%	3.92%	94.55%	90.71%

4.3.3.3 Results for RIF modality

For the RIF modality, the standard deviation σ of the SQI algorithm is optimized while the PCANet parameters remain fixed at $R_a = 75\%$, filter size $(k_1, k_2) = (13 \times 13)$, block size $b = [40, 40]$, and $N = [5, 5]$. Table 4.11 indicates that $\sigma = 3$ delivers the best overall performance, achieving a Rank-1 identification accuracy of 93.43%, an EER of 4.46%, and verification rates of 92.73% and 88.89% at FAR levels of 1% and 0.1%, respectively.

Table 4.11: Results for the RIF modality obtained using xDNN with different standard deviation values (σ).

Standard Deviation	Identification RANK-1%	EER%	Verification VR @ 1% FAR%	Verification VR @ 0.1% FAR%
1	92.93%	4.45%	93.43%	89.60%
3	93.43%	4.46%	92.73%	88.89%
5	92.22%	4.64%	93.03%	88.08%
7	91.31%	4.34%	92.53%	86.06%
9	0.9101	4.87%	91.52%	83.84%

Figure 4.10 shows the CMC and ROC curves of the FKP recognition system for different values of the standard deviation σ.

- The overlap ratio (R_a)

  The block overlap ratio R_a was varied from 0% to 75%, and performance was evaluated in terms of Rank-1 accuracy, EER, and verification rates at 1% and 0.1% FAR (Table 4.12). The best overall balance is achieved at $R_a = 25\%$, yielding 91.92% Rank-1 accuracy, an EER of 4.65%, and verification rates of 91.62% at 1% FAR and 85.45% at 0.1% FAR.

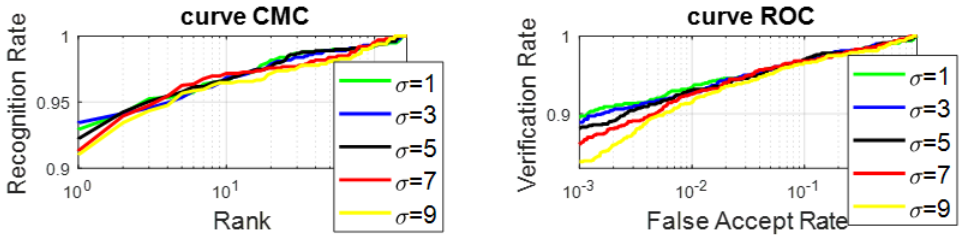


Figure 4.10: CMC curves (left) and ROC curves (right) for the RIF modality obtained using different standard deviation values.

Although $R_a = 75\%$ shows slightly higher verification at 1% FAR, 25% provides the most consistent performance. Figures 4.11 and 4.12 present the corresponding CMC and ROC curves.

Table 4.12: Results for the RIF modality obtained using xDNN with different block overlap ratio values (R_a).

Block Overlap Ratio (R_a)%	Identification RANK-1%	EER%	Verification VR @ 1% FAR%	Verification VR @ 0.1% FAR%
0	88.48%	5.65%	87.17%	79.09%
25	91.92%	4.65%	91.62%	85.45%
50	90.71%	4.43%	90.81%	83.94%
75	91.72%	4.65%	92.12%	86.57%

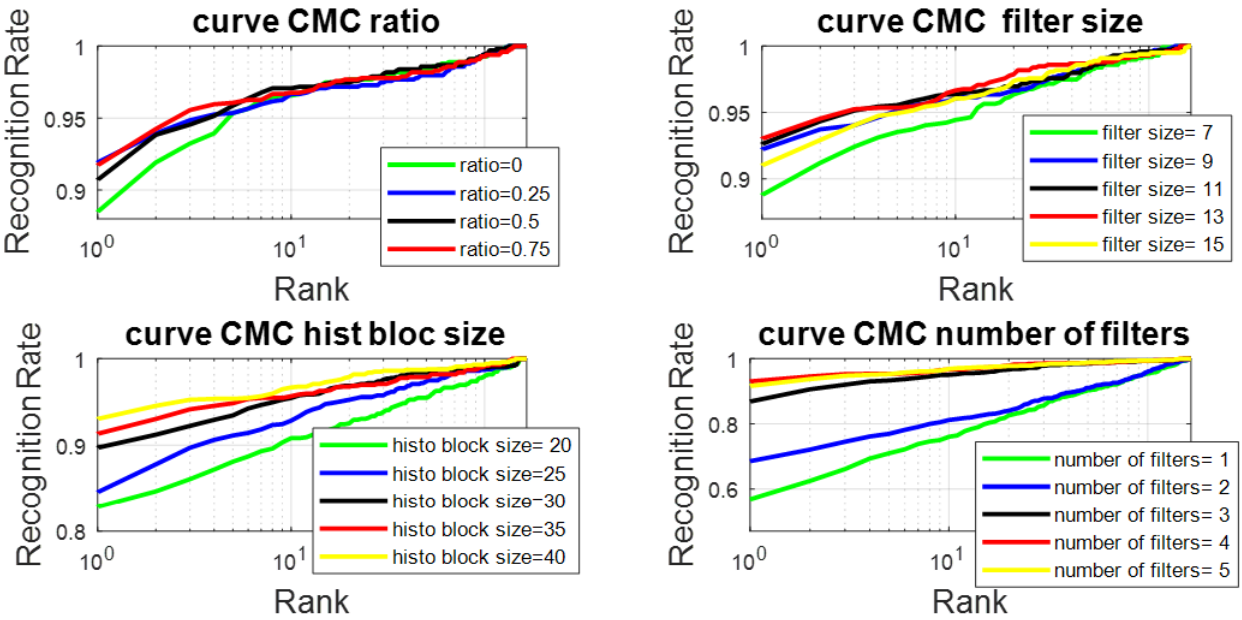


Figure 4.11: CMC curves for the RIF modality obtained using xDNN and PCANet under different parameter settings, including the overlap ratio, filter size, histogram block size, and number of filters.

- Filter size

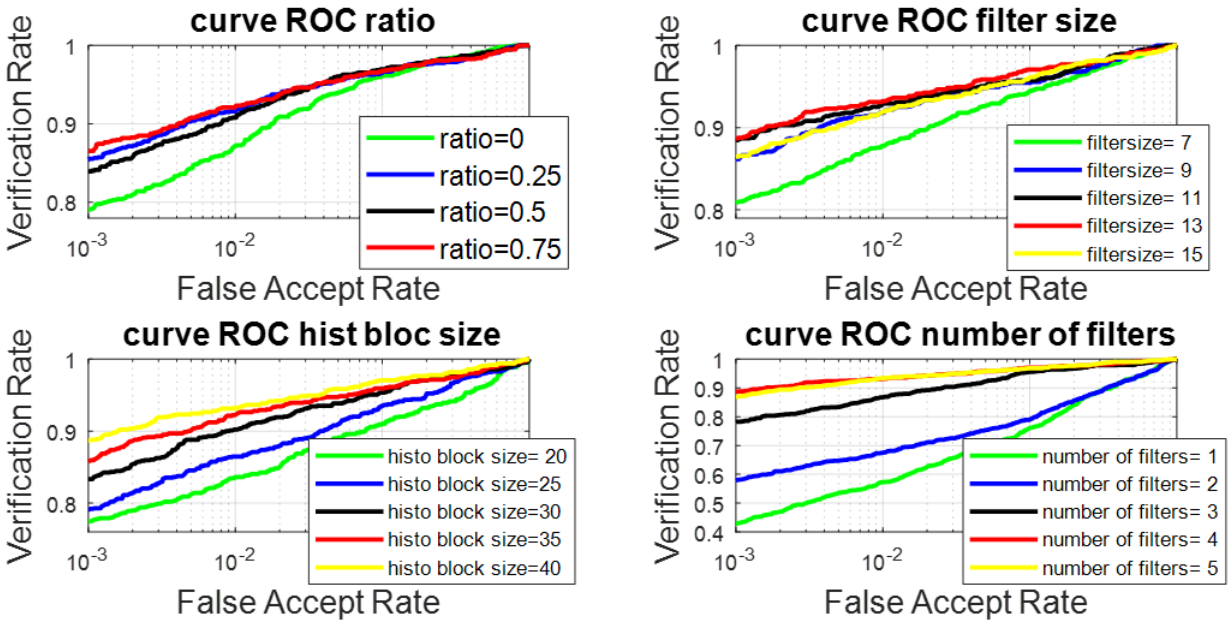


Figure 4.12: ROC curves for the RIF modality obtained using xDNN and PCANet under different parameter settings, including the overlap ratio, filter size, histogram block size, and number of filters.

For the RIF modality, with $\sigma = 1$ and $R_a = 75\%$ kept fixed, PCANet filter sizes ranging from 3×3 to 15×15 are investigated, as summarized in Table 4.13. The results show that the 13×13 filter yields the best overall performance, achieving a Rank-1 accuracy of 93.03%, an EER of 4.34%, and verification rates of 93.23% and 88.69% at 1% and 0.1% FAR, respectively, making it the most suitable configuration.

Table 4.13: Results for the RIF modality obtained using xDNN with different filter size configurations (k1, k2).

Filter Size	Identification RANK-1%	EER%	Verification VR @ 1% FAR%	Verification VR @ 0.1% FAR%
3×3	60.61%	15.59%	55.96%	42.02%
5×5	78.99%	10.75%	75.86%	63.74%
7×7	88.79%	6.71%	87.78%	80.71%
9×9	92.22%	5.15%	91.82%	86.06%
11×11	92.63%	4.95%	92.93%	88.48%
13×13	93.03%	4.34%	93.23%	88.69%
15×15	91.01%	5.15%	91.92%	86.46%

- The bloc size

For the RIF modality, with $\sigma = 1$, $R_a = 75\%$, and filter size $(k_1, k_2) = 13 \times 13$ fixed, different block sizes from $[20, 20]$ to $[40, 40]$ were evaluated (Table 4.14). The $[40, 40]$ block size achieves the best performance, with 93.03% Rank-1 accuracy, an EER of 4.34%, and verification rates of 93.23% at 1% FAR and 88.69% at 0.1% FAR, making it

the most effective choice.

Table 4.14: Results for the RIF modality obtained using xDNN with different block size values (b).

Block Size (h)	Identification RANK-1%	EER%	Verification VR @ 1% FAR%	Verification VR @ 0.1% FAR%
[20 20]	82.83%	9.21%	83.64%	77.47%
[25 25]	84.55%	7.49%	86.46%	79.09%
[30 30]	89.70%	5.79%	90.20%	83.33%
[35 35]	91.31%	5.15%	92.22%	85.76%
[40 40]	93.03%	4.34%	93.23%	88.69%

- The number of filters

  For the RIF modality, with $\sigma = 1$, $R_a = 75\%$, filter size $(k_1, k_2) = 13 \times 13$, and block size $b = [40, 40]$ fixed, the number of filters N was varied from $[1, 1]$ to $[5, 5]$ (Table 4.15). The $[4, 4]$ configuration provides the best balance, achieving 93.03% Rank-1 accuracy, an EER of 4.34%, and verification rates of 93.23% at 1% FAR and 88.69% at 0.1% FAR.

  Table 4.15: Performance metrics obtained for different numbers of filters (N).

Number of Filters (N)	Identification RANK-1%	EER%	Verification VR @ 1% FAR%	Verification VR @ 0.1% FAR%
[1 1]	56.87%	17.71%	57.27%	42.93%
[2 2]	68.59%	16.08%	67.47%	57.98%
[3 3]	86.87%	6.56%	86.77%	78.18%
[4 4]	93.03%	4.34%	93.23%	88.69%
[5 5]	91.62%	4.52%	93.43%	86.97%

4.3.3.4 Results for LIF modality

For the LIF modality, with $R_a = 75\%$, filter size $(k_1, k_2) = (13 \times 13)$, block size $b = [40, 40]$, and $N = [5, 5]$ kept fixed, the standard deviation σ is varied from 1 to 9, as reported in Table 4.16. The best performance is obtained at $\sigma = 5$, achieving a Rank-1 accuracy of 91.82%, an EER of 4.85%, and a verification rate of 91.01% at 1% FAR. The corresponding CMC and ROC curves in Fig. 4.13 further confirm that $\sigma = 5$ provides the optimal performance.

- The overlap ratio (R_a)

  With $\sigma = 5$ fixed, the block overlap ratio R_a was varied from 0% to 75% (Table 4.17). The best performance occurs at $R_a = 75\%$, achieving 90.00% Rank-1 accuracy and verification rates of 90.10% at 1% FAR and 84.65% at 0.1% FAR. Figures 4.14 and 4.15

Table 4.16: Performance metrics obtained for different values of the standard deviation (σ).

Standard Deviation	Identification RANK-1%	EER%	Verification VR @ 1% FAR%	Verification VR @ 0.1% FAR%
1	90.91%	5.26%	90.71%	85.76%
3	91.31%	5.17%	90.40%	85.35%
5	91.62%	4.85%	91.01%	85.66%
7	89.49%	5.76%	89.80%	84.34%
9	89.39%	6.53%	88.69%	82.63%

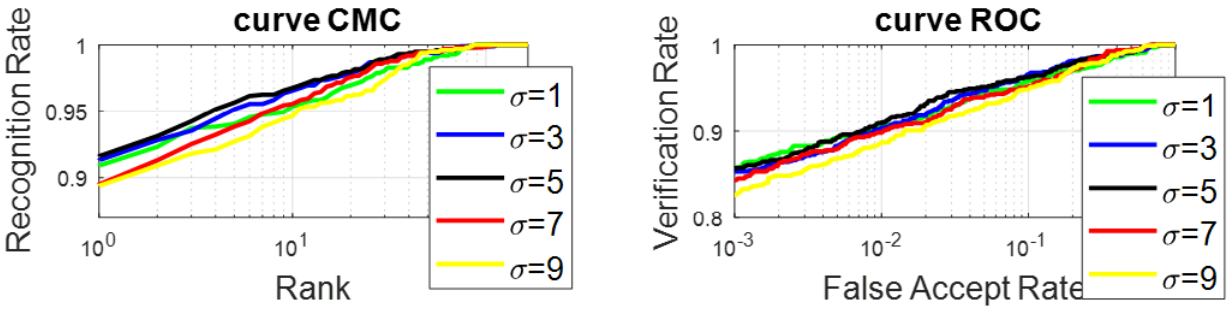


Figure 4.13: CMC curves (left) and ROC curves (right) for the LIF modality obtained using different standard deviation values.

show the corresponding CMC and ROC curves, confirming $R_a = 75\%$ as the optimal setting.

Table 4.17: Results for the LIF modality obtained using xDNN with different block overlap ratio values (R_a).

Block Overlap Ratio (R_a)%	Identification RANK-1%	EER%	Verification VR @ 1% FAR%	Verification VR @ 0.1% FAR%
0	89.19%	5.96%	88.89%	82.32%
25	89.09%	6.04%	88.99%	83.94%
50	89.49%	5.76%	89.80%	84.34%
75	90.00%	5.98%	90.10%	84.65%

- Filter size

  With $\sigma = 5$ and $R_a = 75\%$ fixed, different filter sizes from 3×3 to 15×15 were evaluated (Table 4.18). The 11×11 filter achieves the best performance, with 91.82% Rank-1 accuracy, an EER of 4.44%, and 86.36% verification at 0.1% FAR, making it the optimal choice for the PCANet descriptor.

- The block Size (b)

  With $\sigma = 5$, $R_a = 75\%$, and filter size $(k_1, k_2) = 11 \times 11$ fixed, block sizes from $[20, 20]$ to

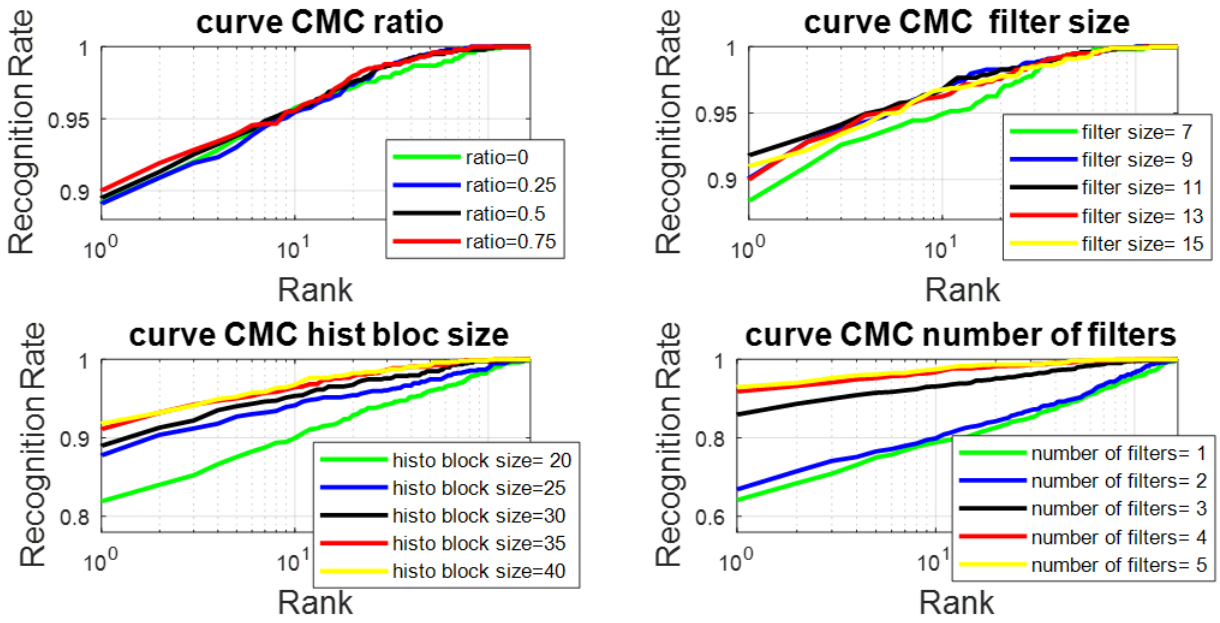


Figure 4.14: CMC curves for the LIF modality obtained using xDNN and PCANet under different parameter settings, including the overlap ratio, filter size, histogram block size, and number of filters.

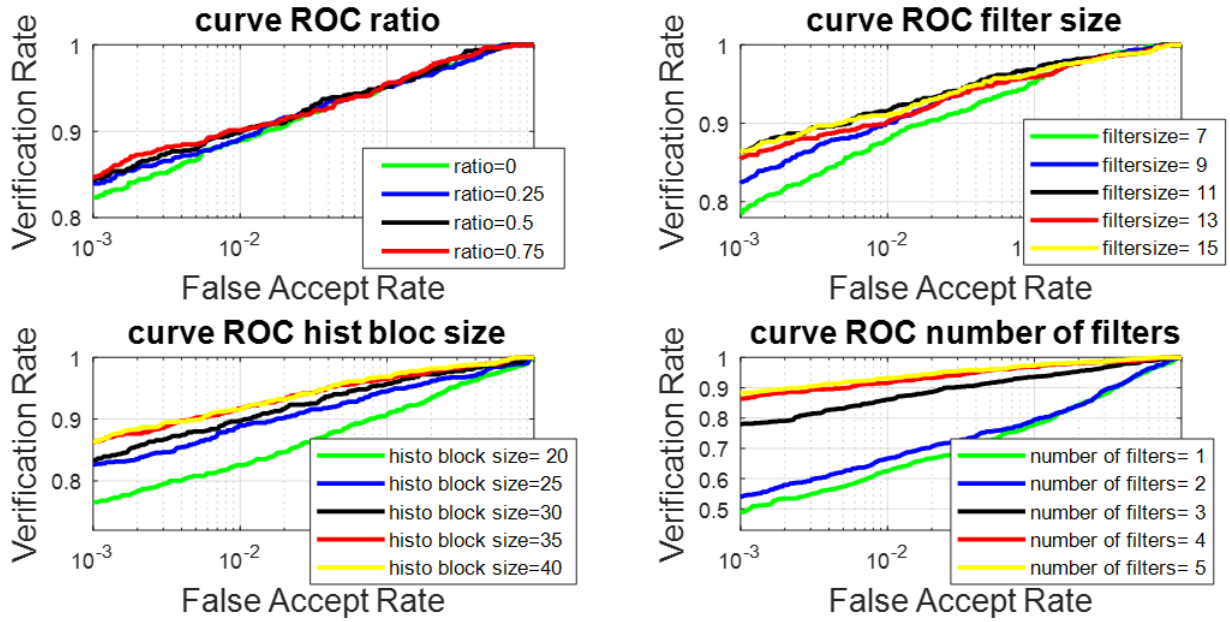


Figure 4.15: ROC curves for the LIF modality obtained using xDNN and PCANet under different parameter settings, including the overlap ratio, filter size, histogram block size, and number of filters.

$[40,40]$ were tested (Table 4.19). The $[40,40]$ block size provides the best performance, with 91.82% Rank-1 accuracy, an EER of 4.44%, and 86.36% verification at 0.1% FAR.

- The number of filters

Table 4.18: Results for the LIF modality obtained using xDNN with different filter size configurations (k1, k2).

Filter Size	Identification RANK-1%	EER%	Verification VR @ 1% FAR%	Verification VR @ 0.1% FAR%
3×3	91.72%	5.28%	91.82%	87.17%
5×5	76.46%	11.22%	74.65%	63.33%
7×7	88.38%	6.24%	87.98%	78.59%
9×9	90.10%	4.73%	90.10%	82.42%
11×11	91.82%	4.44%	91.62%	86.36%
13×13	90.00%	5.18%	90.20%	85.56%
15×15	91.01%	4.65%	91.11%	86.26%

Table 4.19: Results for the LIF modality obtained using xDNN with different numbers of filters (N).

Block Size (b)	Identification RANK-1%	EER%	Verification VR @ 1% FAR%	Verification VR @ 0.1% FAR%
[20 20]	81.92%	9.51%	82.53%	76.57%
[25 25]	87.78%	6.77%	88.99%	82.63%
[30 30]	88.99%	5.47%	89.70%	83.33%
[35 35]	91.11%	4.55%	91.72%	86.26%
[40 40]	91.82%	4.44%	91.62%	86.36%

For the LIF modality, the number of PCANet filters N is evaluated over the range $[1,1]$ to $[5,5]$, while keeping $\sigma = 5$, $R_a = 75\%$, filter size $(k_1,k_2) = (11 \times 11)$, and block size $b = [40,40]$ fixed. As shown in Table 4.20, performance improves as the number of filters increases, with $N = [5,5]$ achieving the best results in terms of Rank-1 accuracy (92.93%), EER (4.14%), and verification rates of 93.13% at 1% FAR and 88.18% at 0.1% FAR. Table 4.20 summarizes the optimal parameters of both the SQI algorithm and the PCANet descriptor derived from these experiments.

Table 4.20: Performance metrics obtained for different numbers of filters (N).

Number of Filters (N)	Identification RANK-1%	EER%	Verification VR @ 1% FAR%	Verification VR @ 0.1% FAR%
[1 1]	64.14%	17.27%	62.42%	48.59%
[2 2]	66.87%	16.97%	66.57%	54.04%
[3 3]	85.96%	7.17%	86.06%	77.98%
[4 4]	91.82%	4.44%	91.62%	86.36%
[5 5]	92.93%	4.14%	93.13%	88.18%

4.3.4 Comparative study

In this subsection, the proposed approach is evaluated against several well-established methods for FKP recognition in a unimodal setting. Each finger type (LIF, LMF, RIF, and

Table 4.21: Optimal parameter values for the SQI algorithm and the PCANet descriptor.

Parameters	RMF Modality	LMF Modality	RIF Modality	LIF Modality
Standard deviation σ	1	1	3	5
Block overlap ratio (R_a)	25	25	25	75
Filter Size	11×11	15×15	13×13	11×11
Block Size Number of Filters	[40 40] [5 5]	[40 40] [4 4]	[40 40] [4 4]	[40 40] [5 5]

RMF) is regarded as a distinct modality. Table 4.22 outlines the identification results, thereby underscoring the efficacy of the proposed framework in comparison to current methodologies.

Table 4.22: ROR (%) for different modalities

Modalities	ROR (%)
LMF	95.86
RMF	95.96
LIF	92.93
RIF	93.03

To thoroughly assess the robustness and discriminative power of the proposed illumination-invariant and explainable PCANet-based FKP recognition framework, a comprehensive comparative study is conducted against several recent state-of-the-art methods reported in [70, 71, 94, 22]. In order to guarantee a fair and meaningful comparison, all experiments are performed under the same evaluation protocol across the four finger modalities (LIF, LMF, RIF, and RMF), ensuring consistency in data usage and experimental conditions. The system performance is quantitatively evaluated using the Rank-1 recognition rate as the primary metric, reflecting identification accuracy, and the corresponding results are summarized and presented in Table 4.23.

Table 4.23: Performance comparison of the proposed system with state-of-the-art methods.

Methods	LIF ROR (%)	LMF ROR (%)	RIF ROR (%)	RMF ROR (%)
Ref [70]	93.80	94.70	92.20	94.80
Ref [71]	91.01	94.85	91.41	91.82
Ref [94]	97.30	95.75	96.83	95.15
Ref [22]	94.34	93.54	93.94	94.19
This work	92.93	95.86	93.03	95.96

The comparison results indicate that the proposed framework provides strong recognition

performance across all FKP modalities. The obtained Rank-1 recognition rates reach 92.93%, 95.86%, 93.03%, and 95.96% for LIF, LMF, RIF, and RMF, respectively. In particular, the system achieves the highest performance for the middle-finger modalities (LMF and RMF), outperforming the compared methods in these categories. For LIF and RIF, the obtained results remain competitive with previously reported approaches, showing only minor differences with the best existing methods. Beyond recognition accuracy, the proposed framework introduces additional advantages. The use of the SQI preprocessing improves robustness to illumination variations, while the xDNN enables transparent and interpretable classification decisions. These characteristics distinguish the proposed approach from conventional black-box recognition models. Consequently, the framework provides an effective balance between recognition accuracy, robustness to lighting variations, and interpretability, which makes it suitable for security-sensitive biometric applications.

4.4 Discussion

Figure 4.16 presents a clear example of the interpretable decision rules generated by the xDNN model during the classification process. The model expresses its decisions in the form of IF–THEN rules, where each condition is associated with a learned prototype (MegaCloud) extracted directly from the training data. These prototypes capture representative and discriminative patterns that establish a meaningful relationship between input features and the final class assignment. For instance, a rule of the form $IF\ (x \sim Prototype_1)\ OR\ (x \sim Prototype_2)\ OR\ (x \sim Prototype_3)\ THEN\ Class = "1"$ indicates that an input sample is assigned to a given class if it exhibits sufficient similarity to one or more representative prototypes.

In addition, these prototypes are also displayed as image patches within the rule structure, enabling intuitive visual interpretation and facilitating expert validation of the model's reasoning process. This dual representation enhances transparency by combining explicit decision logic with visual evidence derived from representative patterns, thereby significantly improving the interpretability, reliability, and trustworthiness of the model's predictions in practical applications.

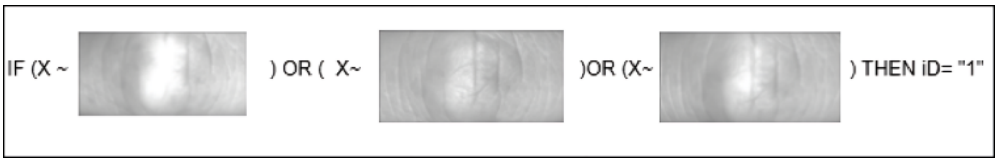


Figure 4.16: Explainability visualization

4.5 Conclusion

This chapter presented an explainable biometric recognition framework for FKP identification. The proposed approach combines illumination-invariant preprocessing based on the SQI technique with hierarchical feature extraction using PCANet and classification through the xDNN. This integration allows the system to achieve high recognition performance while maintaining full transparency in the decision-making process.

Extensive experiments conducted on the PolyU FKP database demonstrate the effectiveness of the proposed framework across the four finger modalities. The obtained results show strong identification performance, with recognition rates ranging from approximately 91% to 96%. These results confirm that the combination of SQI-based illumination normalization and PCANet feature extraction improves the robustness of the recognition process, while the xDNN classifier provides interpretable decisions through explicit IF–THEN rules and visual prototypes.

Beyond recognition accuracy, the main contribution of this work lies in introducing explainability into the FKP recognition pipeline. The generated decision rules and prototype-based visualizations enable users to understand how classification decisions are made, thereby increasing transparency and trust compared to conventional black-box models.

Future research will concentrate on expanding the suggested framework to multimodal biometric systems by incorporating FKP with additional modalities such as palmprint, fingerprint, and iris to enhance recognition reliability. Furthermore, hybrid optimization and feature selection methodologies will be examined to augment system performance and flexibility. User-centered evaluations will be performed to assess the influence of explainable mechanisms on human comprehension and confidence in biometric decision systems.

Chapter 5

Hierarchical Prototype Learning with Edge-Enhanced Features for FKP

5.1 Introduction

Biometric authentication has become an essential component of modern security systems as the demand for reliable digital identity verification continues to increase with the expansion of online services, electronic commerce, and cloud-based infrastructures. Conventional authentication mechanisms, such as passwords, PIN codes, or identity cards, present several limitations since they can be forgotten, stolen, or easily compromised. Biometric systems address these weaknesses by relying on inherent physiological or behavioral characteristics of individuals, providing a more reliable and convenient solution for identity verification.

Among the various biometric traits investigated in recent years, FKP has attracted growing attention due to its distinctive texture patterns and strong discriminative properties. The surface of the finger knuckle contains stable line and crease structures that remain relatively consistent over time, making them suitable for biometric recognition. In addition, FKP images can be acquired using simple and low-cost imaging devices, which makes this modality practical for real-world applications. Nevertheless, the performance of FKP recognition systems largely depends on the quality of the captured images and the effectiveness of the feature extraction process. Variations in illumination conditions, skin characteristics, and sensor noise may introduce distortions that affect the reliability of recognition. Therefore, preprocessing techniques

play a crucial role in enhancing image quality and highlighting meaningful structural patterns. In this context, edge detection methods are widely used to emphasize important boundaries and texture details within the image.

Recent advances in deep learning have significantly improved the performance of biometric recognition systems. CNNs are particularly effective at automatically learning hierarchical and discriminative representations from raw images, enabling robust feature extraction even under challenging conditions. By combining advanced preprocessing techniques with deep feature extraction and efficient classification strategies, more accurate and reliable FKP recognition systems can be developed. This chapter explores such an approach named **EDHP-FKP** to improve the effectiveness and robustness of FKP-based biometric identification.

5.2 The proposed EDHP-FKP approach

5.2.1 Image pre-processing

In biometric imaging, noise is a common issue that can significantly reduce image quality and negatively influence the accuracy and robustness of recognition systems. To mitigate this problem, the proposed framework employs three edge detection techniques: Canny, Sobel, and Laplacian of Gaussian (LoG). These operators were selected because they provide improved resistance to noise, effective edge localization at multiple scales, and stronger gradient representation. In contrast, classical operators such as Roberts and Prewitt are generally more sensitive to image noise and distortions, which can result in unstable or inaccurate edge extraction when applied to biometric images.

- **Canny Edge Detection**: Compared with simpler gradient-based operators such as the Roberts detector, which may struggle to identify accurate edges in the presence of noise or image artifacts, the Canny method improves detection reliability by incorporating a noise-reduction stage prior to gradient computation [107]. The first step consists of smoothing the biometric image using a Gaussian filter in order to suppress noise and small fluctuations. This operation can be expressed as:

$$g(x,y) = G_{\sigma}(x,y) * f(x,y) \tag{5.1}$$

where the Gaussian kernel is defined as

$$G_\sigma(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \tag{5.2}$$

After the smoothing stage, the gradient magnitude and orientation are calculated to determine the strength and direction of intensity changes within the image. These quantities are obtained using the following expressions:

$$M(x,y) = \sqrt{g_x^2(x,y) + g_y^2(x,y)} \tag{5.3}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{g_y(x,y)}{g_x(x,y)}\right) \tag{5.4}$$

To eliminate weak responses that are unlikely to correspond to true edges, a thresholding operation is applied to the gradient magnitude:

$$M_T(x,y) = \begin{cases} M(x,y), & \text{if } M(x,y) > T \\ 0, & \text{otherwise} \end{cases} \tag{5.5}$$

Let $M(x,y)$ represent the gradient magnitude at the pixel location (x,y) and $\theta(x,y)$ the associated gradient direction. Following thresholding, non-maximum suppression is performed on the resulting gradient map $M_T(x,y)$ to retain only the local maxima along the gradient direction, which correspond to potential edge points.

Subsequently, a double-thresholding procedure is applied using two thresholds τ_1 and τ_2 $(\tau_1 < \tau_2)$ to generate two binary edge maps denoted by T_1 and T_2. The higher threshold τ_2 is used to detect strong and reliable edges with minimal noise influence, whereas the lower threshold τ_1 allows the preservation of weaker edge responses that may still carry useful structural information.

In the final stage, edge tracking by hysteresis is carried out by connecting strong edges identified in T_2 with neighboring weak edges present in T_1. This step ensures the continu-

ity of edges and produces a coherent and well-defined edge map suitable for subsequent processing stages.

- **Sobel Edge Detection**:

  The Sobel operator is a gradient-based technique used to estimate the rate of intensity variation in an image. It determines edge information by approximating the first-order derivatives of the image intensity along two orthogonal directions. This operator employs two convolution kernels of size 3×3, designed to measure the horizontal (G_x) and vertical (G_y) intensity gradients. By convolving these masks with the image, the algorithm computes the directional gradient components at each pixel location, enabling the detection of regions where intensity changes rapidly, which typically correspond to edges.

  The vertical kernel G_y is derived from the horizontal kernel G_x through a 90° rotation, allowing the operator to capture edge structures oriented in both horizontal and vertical directions [108]. These convolution kernels are defined as follows:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \qquad G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{5.6}$$

  These kernels emphasize horizontal and vertical gradient variations, making the Sobel operator effective for highlighting edge structures while maintaining moderate robustness against noise.

- **Laplacian of Gaussian (LoG)**:

  The LoG operator, commonly referred to as the Marr–Hildreth or Mexican Hat operator, is a second-order derivative approach designed for edge detection. This technique integrates two operations within a single framework: Gaussian smoothing and Laplacian differentiation. Initially, the image is smoothed using a Gaussian filter to reduce high-frequency noise and minor fluctuations. Subsequently, the Laplacian operator is applied to the smoothed image to detect regions where the intensity changes sharply.

  Edges are identified by detecting zero-crossings in the resulting second-derivative re-

sponse, which correspond to locations where the image intensity undergoes significant transitions. By combining noise suppression with second-order differentiation, the LoG operator provides precise edge localization and can support sub-pixel edge estimation through interpolation techniques.

Mathematically, the Laplacian of Gaussian operator can be expressed as:

$$\mathrm{LoG}(x,y) = \frac{1}{\pi\sigma^4}\left(\frac{x^2+y^2}{\sigma^2} - 2\right)\exp!\left(-\frac{x^2+y^2}{\sigma^2}\right) \tag{5.7}$$

This formulation highlights the joint effect of Gaussian smoothing and Laplacian filtering, allowing the operator to enhance significant intensity variations while attenuating noise components in the image.

5.2.2 Feature Extraction

To obtain discriminative representations from the preprocessed images, several deep feature extraction models are utilized in the experimental framework.

- **VGG-16:** The VGG-16 network extracts hierarchical feature representations through a deep architecture composed of sequential convolutional layers with small receptive fields. This design enables the network to learn detailed spatial patterns and fine structural information, which is particularly beneficial for capturing subtle characteristics present in biometric images [109].

- **AlexNet:** AlexNet is employed to generate robust feature representations at intermediate and high abstraction levels. Owing to its comparatively shallow architecture, it provides an efficient balance between feature learning capability and computational complexity, enabling effective identification of discriminative visual patterns with reduced processing cost [110].

- **ResNet-50:** ResNet-50 introduces residual learning through skip connections that allow information to propagate across deep layers without degradation. This architecture enables the network to learn highly discriminative features while alleviating issues such as gradient vanishing during training, thereby improving the extraction of complex biometric patterns [111].

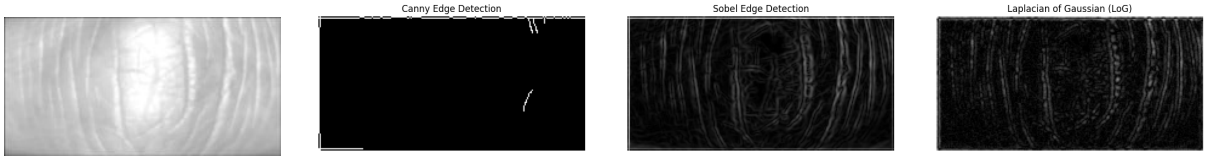
- **Gabor filter–based features:** In addition to deep representations, texture descriptors based on Gabor filters are used to capture localized spatial-frequency information. These filters are sensitive to specific orientations and frequencies, making them well suited for modeling fine texture details and directional structures commonly found in biometric patterns.

$$G(x,y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \phi\right) \tag{5.8}$$

where

$$x' = x\cos\theta + y\sin\theta, \quad y' = -x\sin\theta + y\cos\theta \tag{5.9}$$

Figure 5.1 presents an illustrative example of a finger knuckle print (FKP) image together with the corresponding edge detection outputs produced using the Canny, Sobel, and Laplacian of Gaussian (LoG) operators.



(a) (b) (c) (d)

Figure 5.1: Example of a finger knuckle print (FKP) image and the corresponding edge detection results obtained using different operators: (a) original FKP image, (b) Canny, (c) Sobel, and (d) Laplacian of Gaussian (LoG).

To ensure consistency among the extracted features, a min–max normalization procedure is applied to scale the feature values into a common interval. This normalization step reduces the influence of magnitude differences between features and facilitates subsequent processing stages. The normalization is performed as follows:

$$\tilde{X}_k = \frac{X_k - \min(X_k)}{\max(X_k) - \min(X_k)} \tag{5.10}$$

Following normalization, dimensionality reduction is performed using a combination of PCA and LDA. Initially, PCA is applied to eliminate redundant information and project the data into a lower-dimensional space while preserving the most significant variance. Subsequently,

LDA is utilized to further transform the features by maximizing the separability between different classes. This sequential application of PCA and LDA not only reduces computational complexity but also enhances the discriminative capability of the resulting feature representations, leading to improved recognition performance.

5.2.3 HP Classifier

The Hierarchical Prototype (HP) classifier is a prototype-driven classification framework organized in multiple layers. It performs classification by learning representative prototypes for each class and arranging them hierarchically from general representations to more detailed ones. During the training stage, the model constructs prototypes for every class in a layer-wise manner. At each level of the hierarchy, prototypes are progressively created and refined using the samples belonging to the corresponding class. Each prototype is obtained by averaging the feature vectors assigned to it, which can be expressed as:

$$P_l^c = \frac{1}{|S_l^c|} \sum_{x_i \in S_l^c} x_i, \quad x_i \in S_l^c \tag{5.11}$$

where S_l^c denotes the subset of samples from class c associated with layer l. This hierarchical representation allows the classifier to model both coarse-level structures and finer discriminative patterns present in the data.

During the inference stage, a test sample $x \in R$ is evaluated by comparing it with the learned prototypes across all layers. The comparison is performed using a similarity measure modulated by a learnable scaling parameter Y_l, defined as:

$$\text{sim}(x, P_l^c) = \exp(-\gamma_l |x - P_l^c|^2) \tag{5.12}$$

To determine the most probable class, the classifier aggregates the similarity scores computed at different hierarchical levels. Each layer contributes to the final decision according to a weight factor w_l, reflecting its relative importance. The confidence associated with a given class y is therefore computed as:

$$\text{Confidence}(y) = \sum_{l=1}^{L} w_l \cdot \text{sim}(x, P_l^y), \quad \sum_{l=1}^{L} w_l = 1 \tag{5.13}$$

The predicted class label corresponds to the class that achieves the highest aggregated similarity score:

$$y = \arg\max_{c} \sum_{l=1}^{L} w_l \cdot \text{sim}(x, P_l^c) \tag{5.14}$$

The HP classifier architecture includes three main components: the Hierarchical Prototype Layer (HPL), responsible for learning prototype representations at multiple levels of abstraction; the Adaptive Similarity Metric (ASM), which dynamically adjusts the similarity computation; and the Confidence Fusion Module (CFM), which combines the decisions obtained from different layers. Owing to this design, the classifier offers several advantages, including scalability, interpretability, efficient processing of large datasets, and the capability to support incremental learning.

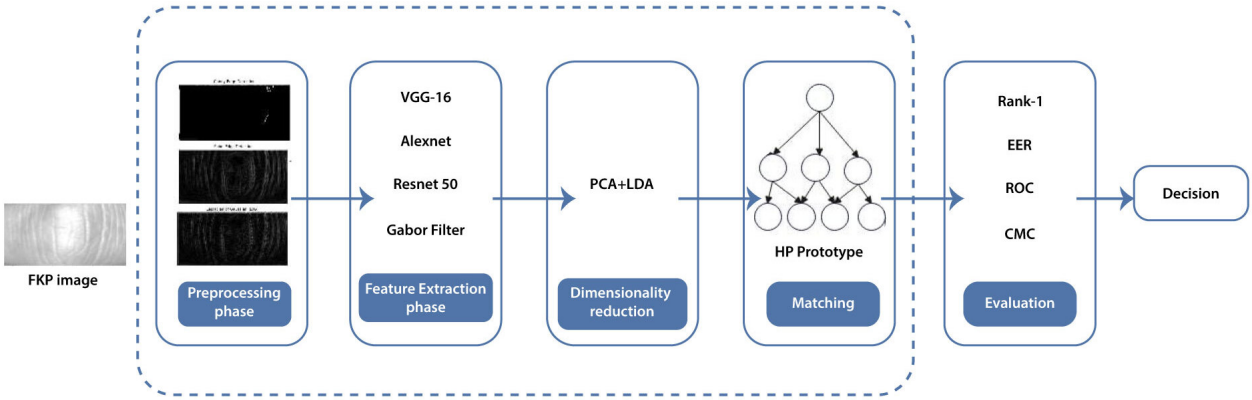


Figure 5.2: Proposed approach.

5.2.4 Nearest Neighbor (NN) classifier

Another classification strategy adopted in the matching stage is the NN classifier. This method assigns a query sample to the class of the most similar template available in the feature space. In the proposed framework, similarity between the query feature vector V_i and a stored template V_j is measured using the Mahalanobis distance, which takes into account the correlations between features through the covariance matrix C. The distance metric is defined

as follows [15]:

$$d_{\text{Mahalanobis}}(V_i, V_j) = \sqrt{(V_i - V_j)^\top C^{-1}(V_i - V_j)}, \tag{5.15}$$

where C^{-1} denotes the inverse of the covariance matrix estimated from the training data. This metric provides a more reliable similarity measure than simple Euclidean distance, as it incorporates the statistical relationships among the feature dimensions.

5.3 Experimental results

5.3.1 Database

The performance of the proposed EDHP-FKP framework is assessed using a publicly available FKP database developed by the Hong Kong Polytechnic University [112]. This dataset contains dorsal finger images collected from 503 subjects, focusing on the middle finger, and includes 2515 captured samples. A large proportion of the participants (approximately 88%) are younger than 30 years old. The images are provided in .JPG format.

Additionally, the database provides predefined regions of interest (ROIs) extracted from both the major and minor knuckle areas of each dorsal finger image. These ROIs correspond to different finger types, namely RMF, LMF, RIF, and other associated finger regions. For each finger category, five samples are available per subject, resulting in a total of 7545 FKP images used in the experimental evaluation.

5.3.2 Experimental Evaluation

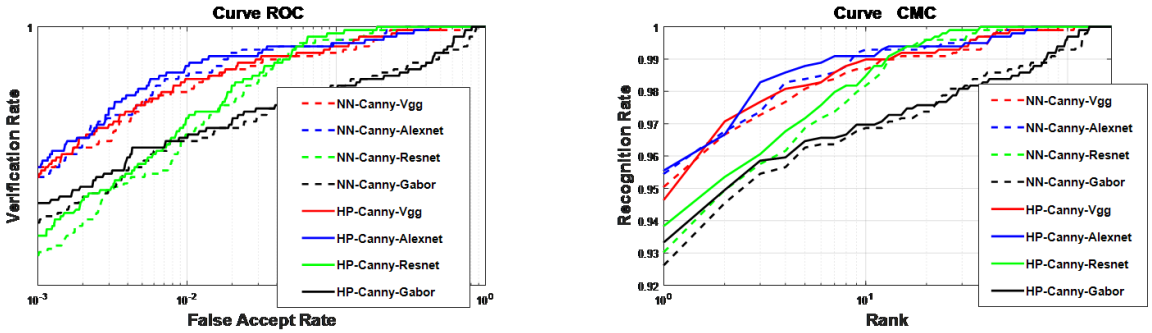
To assess the effectiveness of the EDHP-FKP framework, a series of experiments were conducted using the PolyU FKP database. The evaluation focused on four different finger modalities: RMF, LMF, RIF, and LIF, with classification performed using the HP classifier. In all experiments, the images were preprocessed with Canny, Sobel, and Laplacian of Gaussian (LoG) edge detection operators. Deep features were then extracted using VGG-16, AlexNet, ResNet-50, and Gabor filters, followed by classification with the HP classifier.

5.3.2.1 Experiment 1: RMF Modality with HP Classifier

- **Canny Edge Detection:** The performance metrics for this configuration are summarized in Table 5.1. For better visualization, the ROC and CMC curves are shown in Fig. 2. Among the feature extractors, AlexNet achieved the highest Rank-1 recognition of 95.56% and an EER of 1.11% using the HP classifier. In comparison, the KNN classifier reached a Rank-1 accuracy of 95.45% with an EER of 1.41%.

Table 5.1: Performance of HP and KNN classifiers for RMF modality using Canny edge detection

Deep learning extractor	HP Classifier		KNN Classifier	
	Rank-1 (%)	EER (%)	Rank-1 (%)	EER (%)
VGG-16	94.65	1.52	95.05	1.52
AlexNet	**95.56**	**1.11**	**95.45**	**1.41**
ResNet-50	93.84	2.00	93.03	2.22
Gabor	93.33	2.62	92.63	2.91



(a) ROC curve

(b) CMC curve

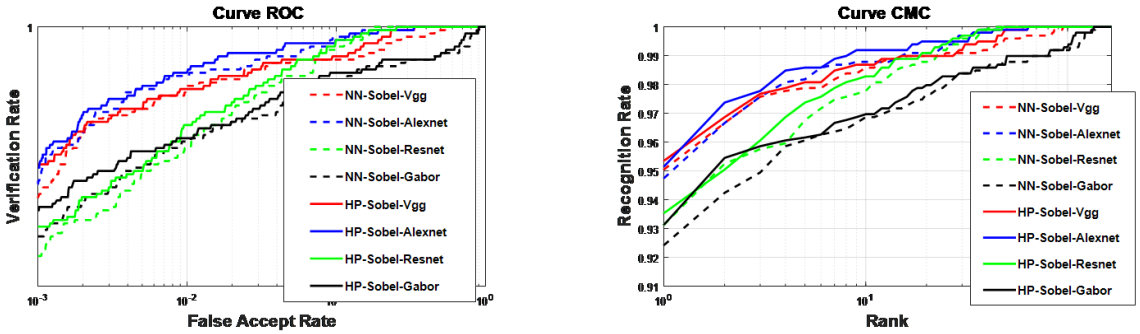
Figure 5.3: ROC and CMC curves of the RMF modality using HP classifier and Canny edge detection.

- **Sobel Edge Detection:** Table 5.13 lists the results when Sobel preprocessing was applied. Fig. 3 presents the ROC and CMC curves for this setup. The AlexNet extractor again outperformed the others, achieving a Rank-1 recognition rate of 95.15% and an EER of 1.22%, while the KNN classifier yielded 94.75% Rank-1 and 1.39% EER.

- **LoG Edge Detection:** The results obtained with LoG preprocessing are displayed in Table 5.3, and the corresponding ROC and CMC curves are shown in Fig. 4. Using AlexNet, the HP classifier achieved a Rank-1 recognition rate of 95.76% and an EER of 1.11%, whereas the KNN classifier reached the same Rank-1 rate of 95.76% with a slightly higher EER of 1.30%.

Table 5.2: Performance of HP and KNN classifiers for RMF modality using Sobel edge detection

Deep learning extractor	HP Classifier		KNN Classifier	
	Rank-1 (%)	EER (%)	Rank-1 (%)	EER (%)
VGG-16	95.35	1.61	95.05	1.62
AlexNet	**95.15**	**1.22**	**94.75**	**1.39**
ResNet-50	93.54	2.22	93.13	2.33
Gabor	93.13	2.63	92.42	2.83

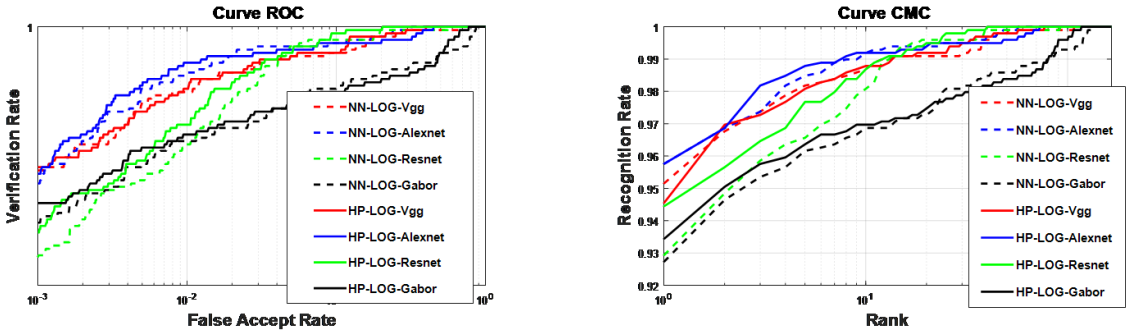


(a) ROC curve (b) CMC curve

Figure 5.4: ROC and CMC curves of the RMF modality using HP Classifier and Sobel Edge Detection.

Table 5.3: Performance of HP and KNN classifiers for RMF modality using LoG edge detection

Deep learning extractor	HP Classifier		KNN Classifier	
	Rank-1 (%)	EER (%)	Rank-1 (%)	EER (%)
VGG-16	94.55	1.62	95.15	1.51
AlexNet	**95.76**	**1.11**	**95.76**	**1.30**
ResNet-50	94.44	1.92	92.93	2.12
Gabor	93.43	2.72	92.73	2.93



(a) ROC curve (b) CMC curve

Figure 5.5: ROC and CMC curves of the RMF modality using HP Classifier and LoG Edge Detection.

Across all three edge detection methods, AlexNet consistently provided the most discriminative features, confirming its suitability for RMF FKP recognition within the proposed EDHP-FKP framework. The results also demonstrate that the HP classifier slightly outperforms the KNN classifier in terms of both recognition accuracy and equal error rate.

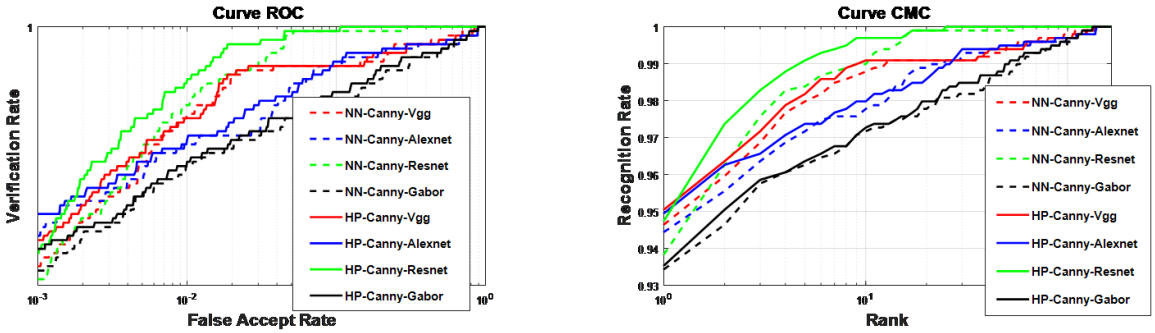
5.3.2.2 Experiment 2: LMF Modality Evaluation Using HP Classifier

This experiment investigates the performance of the EDHP-FKP system on the LMF modality. The images were preprocessed with three different edge detection methods Canny, Sobel, and LoG and features were extracted using VGG-16, AlexNet, ResNet-50, and Gabor filters. Classification was performed using both the HP and KNN classifiers.

- **Canny Edge Detection:** The performance results with Canny preprocessing are summarized in Table 5.4. The ROC and CMC curves for this configuration are depicted in Fig. 5. Among the tested feature extractors, ResNet-50 combined with the HP classifier achieved the highest Rank-1 recognition of 94.75% and an EER of 1.11%. For the KNN classifier, VGG-16 provided the best outcome, with a Rank-1 accuracy of 94.65% and an EER of 1.61%.

Table 5.4: Performance of HP and KNN classifiers for LMF modality using Canny edge detection

Deep learning extractor	**HP Classifier**		**KNN Classifier**	
	Rank-1 (%)	**EER (%)**	**Rank-1 (%)**	**EER (%)**
VGG-16	95.05	1.59	**94.65**	**1.61**
AlexNet	94.95	2.11	94.44	2.52
ResNet-50	**94.75**	**1.11**	93.84	1.34
Gabor	93.54	2.52	93.43	2.63



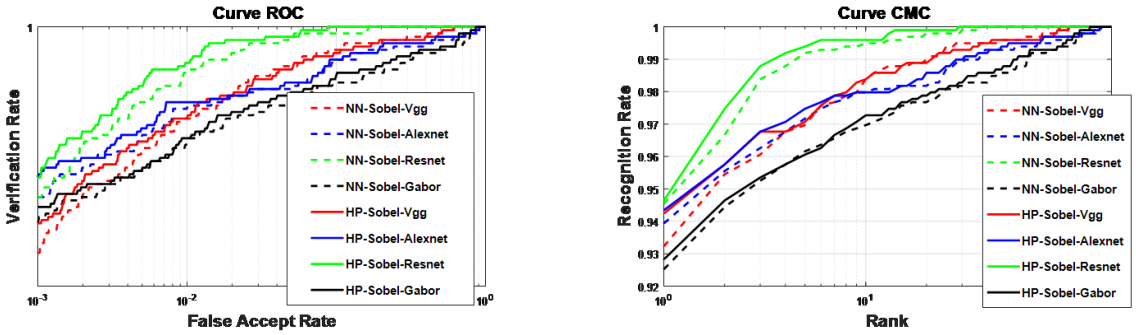
(a) ROC curve (b) CMC curve

Figure 5.6: ROC and CMC curves of the LMF modality using HP Classifier and Canny edge detection.

- **Sobel Edge Detection:** Table 5.5 presents the results for Sobel-based preprocessing, while Fig. 6 shows the corresponding ROC and CMC curves. Once again, ResNet-50 delivered the highest performance using the HP classifier, achieving 94.65% Rank-1 and 1.03% EER, whereas the KNN classifier reached a Rank-1 rate of 94.55% with an EER of 1.21%.

Table 5.5: Performance of HP and KNN classifiers for LMF modality using Sobel edge detection

Deep learning extractor	HP Classifier		KNN Classifier	
	Rank-1 (%)	EER (%)	Rank-1 (%)	EER (%)
VGG-16	94.24	2.04	93.23	2.04
AlexNet	94.34	2.12	93.94	2.12
ResNet-50	**94.65**	**1.03**	**94.55**	**1.21**
Gabor	92.83	2.53	92.53	2.73



(a) ROC curve (b) CMC curve

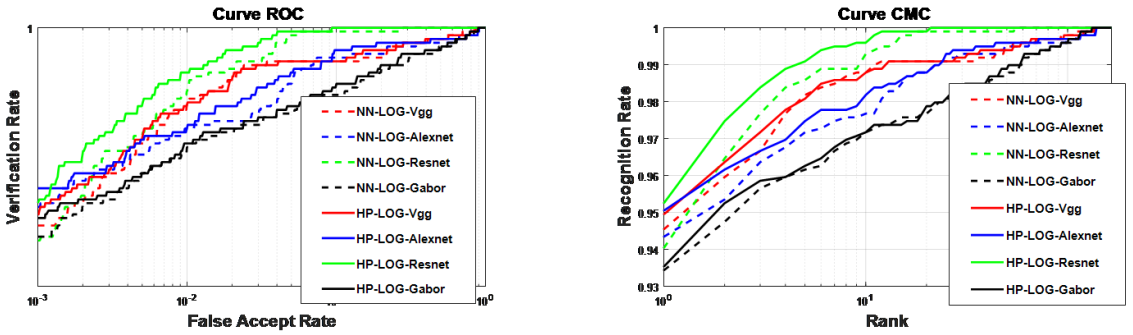
Figure 5.7: ROC and CMC curves of the LMF modality using HP Classifier and Sobel Edge Detection.

- **LoG Edge Detection:** The results obtained using LoG preprocessing are reported in

Table 5.6, with additional visual analysis in Fig. 7. The combination of ResNet-50 features with the HP classifier yielded the highest recognition performance, reaching 95.25% Rank-1 and an EER of 1.11%. Using the KNN classifier, the best Rank-1 rate was 94.04%, with an EER of 1.31%.

Table 5.6: Performance of HP and KNN classifiers for LMF modality using LoG edge detection

Deep learning extractor	HP Classifier		KNN Classifier	
	Rank-1 (%)	EER (%)	Rank-1 (%)	EER (%)
VGG-16	94.95	1.71	94.55	1.61
AlexNet	95.05	2.12	94.34	2.52
ResNet-50	**95.25**	**1.11**	**94.04**	**1.31**
Gabor	93.54	2.52	93.43	2.63



(a) ROC curve (b) CMC curve

Figure 5.8: ROC and CMC curves of the LMF modality using HP Classifier and LoG Edge Detection.

Overall, the experiments indicate that ResNet-50 provides the most discriminative features for the LMF modality, and the HP classifier consistently outperforms KNN, demonstrating the advantage of hierarchical prototype-based classification in handling LMF FKP recognition tasks.

5.3.2.3 Experiment 3: RIF Modality Evaluation Using HP Classifier

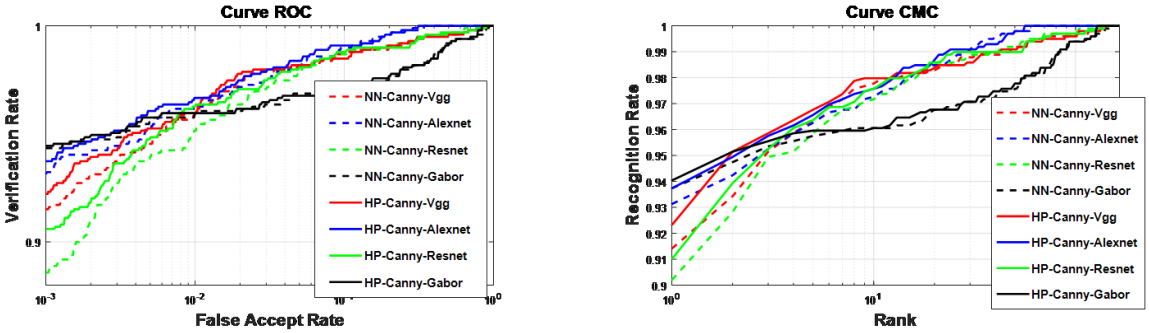
This experiment assesses the performance of the proposed EDHP-FKP system on the RIF modality. The input images were processed using three edge detection methods: Canny, Sobel, and LoG, and features were extracted using VGG-16, AlexNet, ResNet-50, and Gabor filters. Classification was performed with both the HP and KNN classifiers to compare recognition effectiveness.

- **Canny Edge Detection:** The results obtained with Canny preprocessing are summarized in Table 5.7, while Fig. 8 illustrates the corresponding ROC and CMC curves. Among the

tested feature extractors, VGG-16 combined with the HP classifier achieved the highest performance, reaching 92.32% Rank-1 and an EER of 2.12%. For the KNN classifier, the best performance was 91.41% Rank-1 with an EER of 2.32%.

Table 5.7: Performance of HP and KNN classifiers for RIF modality using Canny edge detection

Deep learning extractor	HP Classifier		KNN Classifier	
	Rank-1 (%)	EER (%)	Rank-1 (%)	EER (%)
VGG-16	**92.32**	**2.12**	**91.41**	**2.32**
AlexNet	93.74	2.33	93.13	2.63
ResNet-50	91.01	2.76	90.20	2.93
Gabor	94.04	3.43	93.74	3.63



(a) ROC curve (b) CMC curve

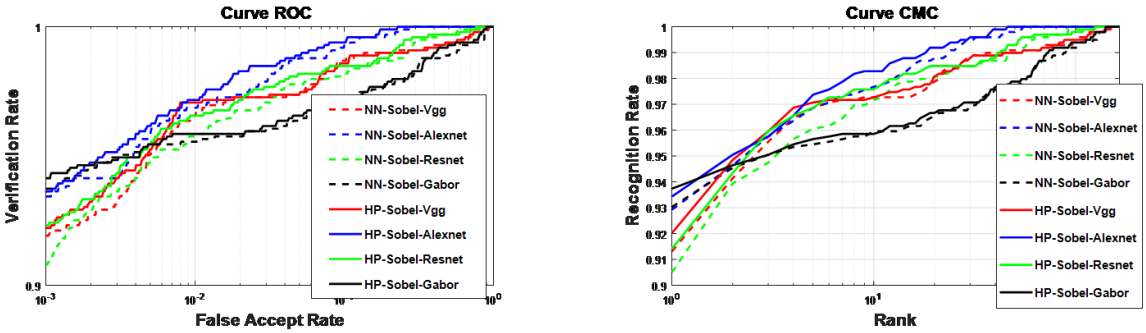
Figure 5.9: ROC and CMC curves of the RIF modality using HP Classifier and Canny Edge Detection.

- **Sobel Edge Detection:** Table 5.8 presents the performance metrics for Sobel preprocessing, with Fig. 9 displaying the ROC and CMC curves. AlexNet with the HP classifier achieved the top performance in this configuration, yielding 93.43% Rank-1 and an EER of 1.81%. Using KNN, the Rank-1 rate was 92.93%, with an EER of 2.42%.

Table 5.8: Performance of HP and KNN classifiers for RIF modality using Sobel edge detection

Deep learning extractor	HP Classifier		KNN Classifier	
	Rank-1 (%)	EER (%)	Rank-1 (%)	EER (%)
VGG-16	92.02	2.72	91.31	2.72
AlexNet	**93.43**	**1.81**	**92.93**	**2.42**
ResNet-50	91.41	2.43	90.51	2.94
Gabor	93.74	3.63	93.03	3.85

- **LoG Edge Detection:** The outcomes for LoG-based preprocessing are shown in Table 5.9, while Fig. 10 presents the ROC and CMC curves. ResNet-50 in combination with



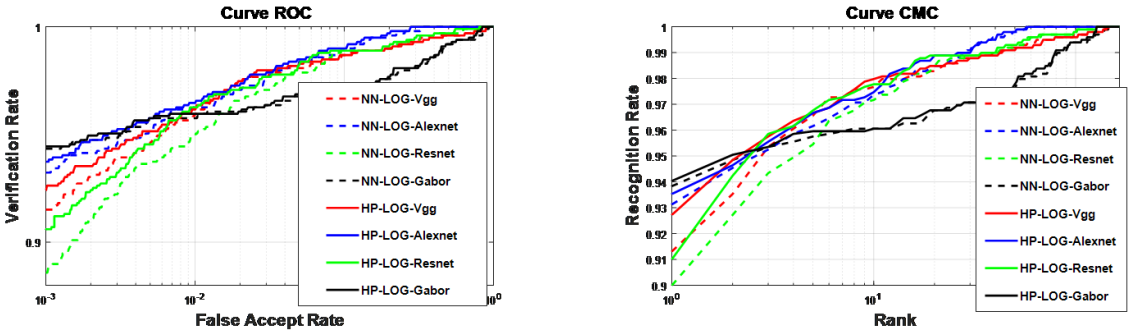
(a) ROC curve (b) CMC curve

Figure 5.10: ROC and CMC curves of the RIF modality using HP Classifier and Sobel Edge Detection.

the HP classifier demonstrated the best overall performance, achieving 93.54% Rank-1 and an EER of 2.31%. For the KNN classifier, VGG-16 reached 91.31% Rank-1 with an EER of 2.40%.

Table 5.9: Performance of HP and KNN classifiers for RIF modality using LoG edge detection

Deep learning extractor	HP Classifier		KNN Classifier	
	Rank-1 (%)	EER (%)	Rank-1 (%)	EER (%)
VGG-16	92.73	2.33	**91.31**	**2.40**
AlexNet	**93.54**	**2.31**	93.13	2.72
ResNet-50	91.01	2.63	90.00	2.93
Gabor	94.04	3.43	93.84	3.63



(a) ROC curve (b) CMC curve

Figure 5.11: ROC and CMC curves of the RIF modality using HP Classifier and LoG Edge Detection.

These experiments indicate that different feature extractors provide varying advantages depending on the edge detection technique, and that the HP classifier generally surpasses KNN in terms of Rank-1 accuracy and EER, confirming the robustness of hierarchical prototype-based classification for the RIF finger knuckle prints.

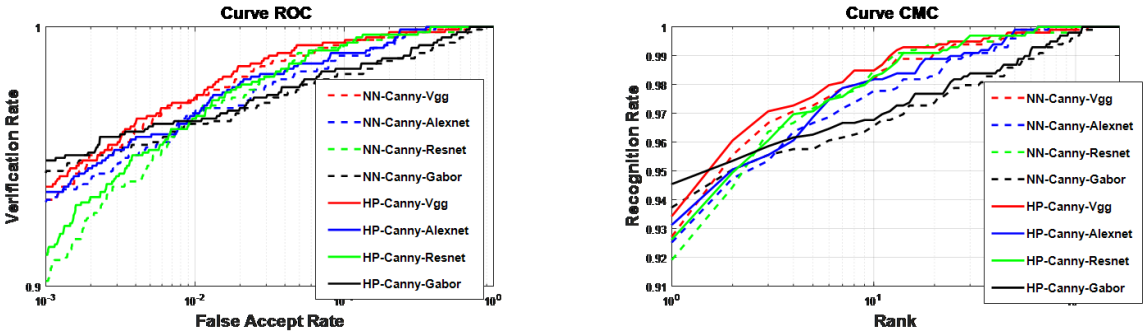
5.3.2.4 Experiment 4: LIF Modality Evaluation Using HP Classifier

The final experiment evaluates the proposed EDHP-FKP approach on the LIF finger knuckle prints. Preprocessing was performed using three edge detection methods Canny, Sobel, and LoG followed by feature extraction through VGG-16, AlexNet, ResNet-50, and Gabor filters. Both HP and KNN classifiers were employed to assess recognition performance.

- **Canny Edge Detection:** Table 5.10 summarizes the results for Canny-based preprocessing, while Fig. 11 presents the ROC and CMC curves. The VGG-16 extractor achieved the best performance with the HP classifier, attaining a Rank-1 accuracy of 93.43% and an EER of 1.80%. When using KNN, the highest performance was 92.73% Rank-1 with an EER of 2.00%.

Table 5.10: Performance of HP and KNN classifiers for LIF modality using Canny edge detection

Deep learning extractor	HP Classifier		KNN Classifier	
	Rank-1 (%)	EER (%)	Rank-1 (%)	EER (%)
VGG-16	**93.43**	**1.80**	**92.73**	**2.00**
AlexNet	93.13	2.11	92.53	2.45
ResNet-50	92.63	2.22	91.92	2.42
Gabor	94.55	2.72	93.74	2.93



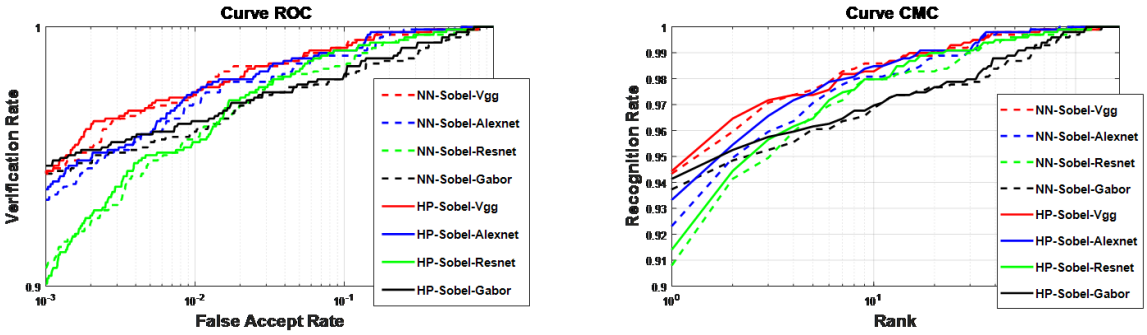
(a) ROC curve (b) CMC curve

Figure 5.12: ROC and CMC curves of the LIF modality using HP Classifier and Canny Edge Detection.

- **Sobel Edge Detection:** The results for Sobel preprocessing are shown in Table 5.11, with ROC and CMC curves provided in Fig. 12. Again, VGG-16 achieved the highest recognition with the HP classifier (94.44% Rank-1, 2% EER), whereas KNN achieved 94.34% Rank-1 and an EER of 1.71%.

Table 5.11: Performance of HP and KNN classifiers for LIF modality using Sobel edge detection

Deep learning extractor	HP Classifier		KNN Classifier	
	Rank-1 (%)	EER (%)	Rank-1 (%)	EER (%)
VGG-16	**94.44**	**2.00**	**94.34**	**1.71**
AlexNet	93.33	2.02	92.32	2.12
ResNet-50	91.41	2.53	90.81	2.52
Gabor	94.14	2.72	93.74	2.64



(a) ROC curve (b) CMC curve

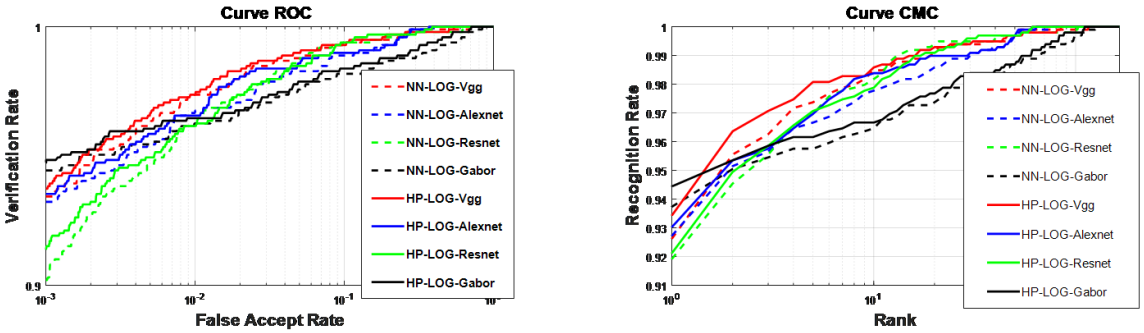
Figure 5.13: ROC and CMC curves of the LIF modality using HP Classifier and Sobel Edge Detection.

- **LoG Edge Detection:** Performance metrics for LoG preprocessing are reported in Table 5.12, with ROC and CMC curves shown in Fig. 13. VGG-16 continued to outperform the other extractors using the HP classifier, reaching 94.43% Rank-1 with an EER of 1.72%. For the KNN classifier, the best recognition was 92.63% Rank-1 with an EER of 1.94%.

Table 5.12: Performance of HP and KNN classifiers for LIF modality using LoG edge detection

Deep learning extractor	HP Classifier		KNN Classifier	
	Rank-1 (%)	EER (%)	Rank-1 (%)	EER (%)
VGG-16	**93.43**	**1.72**	**92.63**	**1.94**
AlexNet	93.03	2.00	92.73	2.53
ResNet-50	92.12	2.32	91.92	2.42
Gabor	94.44	2.80	93.74	2.93

These findings confirm that VGG-16 consistently provides superior feature representation across different edge detection techniques for the LIF modality, and that the HP classifier generally surpasses KNN in recognition accuracy and error rate, highlighting the effectiveness of hierarchical prototype-based classification.



(a) ROC curve

(b) CMC curve

Figure 5.14: ROC and CMC curves of the LIF modality using HP Classifier and LoG Edge Detection.

5.3.3 Comparative Analysis

The comparative evaluation highlights the significant impact of both the choice of feature extractor and the classification method on recognition performance. Across all tested modalities (RMF, LMF, RIF, and LIF) and preprocessing techniques, the HP classifier consistently demonstrates superior results compared to the KNN classifier, achieving higher Rank-1 recognition rates and lower EERs.

Specifically, deep feature extractors paired with the HP classifier show stronger discriminative power than when used with KNN. For instance, in the RMF modality, combining AlexNet with HP classification and LoG edge detection delivers the highest performance, attaining a Rank-1 accuracy of 95.76% with an EER of 1.11%, whereas KNN under the same conditions exhibits slightly higher EER values. A similar pattern is observed for the LMF modality, where ResNet-50 together with the HP classifier achieves a Rank-1 accuracy of 95.25% and an EER of 1.11%, outperforming the corresponding KNN-based approach.

While KNN can provide satisfactory recognition in certain settings, it is generally more sensitive to intra-class variations, resulting in increased EER values. The HP classifier, in contrast, leverages its hierarchical prototype learning mechanism to model class distributions more effectively, enhancing robustness and generalization. These results underline that the HP classifier represents a more reliable and effective classification strategy, particularly when combined with deep feature representations and edge-based preprocessing such as LoG.

As summarized in Table 5.13, the HP classifier consistently achieves higher Rank-1 recognition rates and lower EERs than KNN across all modalities, confirming the effectiveness of

hierarchical prototype-based learning in capturing discriminative features from FKP images.

Table 5.13: Comparison between HP and KNN classifiers using the best Performances per modality

Modality	Edge Detector	Feature Extractor	Classifier	Rank-1 (%)	EER (%)
RMF	LoG	AlexNet	HP	**95.76**	**1.11**
RMF	LoG	AlexNet	KNN	95.76	1.30
LMF	LoG	ResNet-50	HP	**95.25**	**1.11**
LMF	LoG	ResNet-50	KNN	94.04	1.31
RIF	Sobel	AlexNet	HP	**93.43**	**1.81**
RIF	Sobel	AlexNet	KNN	92.93	2.42
LIF	Sobel	VGG-16	HP	**94.44**	**2.00**
LIF	Sobel	VGG-16	KNN	94.34	1.71

5.4 Conclusion

In this chapter, we presented EDHP-FKP, a robust finger knuckle print recognition framework that combines edge-enhanced preprocessing, deep feature extraction, and hierarchical classification. The system leverages multiple edge detectors along with diverse deep feature extractors and the Hierarchical Prototype classifier to capture both global structures and fine-grained texture patterns. Experimental results and comparative analysis indicate that EDHP-FKP consistently outperforms existing methods, demonstrating its effectiveness and reliability for accurate FKP recognition in practical biometric scenarios.

Chapter 6

General Conclusion

This thesis set out to address two fundamental challenges that have limited the practical deployment of finger knuckle print recognition systems: sensitivity to illumination variations and the lack of interpretability in deep learning-based approaches. The work presented in the preceding chapters has demonstrated that these challenges can be overcome through a carefully designed framework that combines robust feature extraction with transparent decision-making. As we reach the conclusion of this investigation, it is worthwhile to reflect on what has been accomplished, consider the broader implications for the field of biometric recognition, and identify promising directions for future research.

The journey began with a recognition that existing FKP systems, despite their promising accuracy under controlled conditions, struggled when deployed in real-world environments where lighting cannot be standardized. It also acknowledged a growing concern within the security community about the opacity of deep learning models—systems that could identify individuals with remarkable accuracy but could not explain how they reached their conclusions. These twin problems formed the motivation for a unified approach that would not compromise either robustness or interpretability.

The framework developed in this thesis demonstrates that such unification is possible. By integrating illumination normalization through the Self-Quotient Image algorithm, hierarchical feature extraction through PCANet, and prototype-based classification through xDNN, we have created a system that maintains high recognition accuracy across varying lighting conditions while providing complete transparency in its decision-making process. The experimental

validation on the PolyU FKP database, encompassing four finger modalities and thousands of images, confirms that this approach achieves recognition rates between 91% and 96%—performance that is competitive with or superior to existing methods.

Beyond the technical achievements, this work represents a step toward a different philosophy of biometric system design—one that values not only what a system can do but also how it does it. In security-critical applications, accuracy alone is insufficient. Users must trust the system; operators must understand its behavior; regulators must be able to audit its decisions. The prototype-based explanations generated by xDNN address these needs by making the reasoning behind each classification visible and verifiable.

6.1 Summary of Contributions

The work presented in this thesis has yielded several contributions to the field of biometric recognition, particularly in the area of finger knuckle print identification with explainable deep learning. These contributions can be summarized as follows.

A unified framework for illumination-invariant and explainable FKP recognition. The primary contribution of this work is the development of a complete recognition pipeline that simultaneously addresses the two major limitations of existing systems. The integration of SQI preprocessing, two-stage PCANet feature extraction, and xDNN classification represents a novel approach that has not been previously explored in the FKP recognition literature. This framework demonstrates that robustness to real-world imaging conditions and transparency in decision-making can be achieved within a single, coherent architecture.

Systematic optimization of key parameters across multiple modalities. Through comprehensive ablation studies conducted on four finger types left index, left middle, right index, and right middle this work has identified optimal configurations for both the illumination normalization algorithm and the feature extraction framework. These investigations revealed that different modalities benefit from different parameter settings, with optimal standard deviation values ranging from 1 to 5, overlap ratios varying between 25% and 75%, and filter sizes differing across finger types. The systematic documentation of these findings provides valuable guidance for future research and practical implementations.

Demonstration of prototype-based explainability for biometric decisions. This thesis provides the first detailed demonstration of how xDNN's prototype-based reasoning can be applied to FKP recognition. The generation of interpretable IF-THEN rules that link classification decisions to specific learned patterns represents a significant advance over conventional black-box models. The visualization of these prototypes as image patches enables human experts to understand and verify the reasoning behind the system's decisions, addressing the critical need for transparency in security-sensitive applications. This contribution extends beyond FKP recognition to the broader challenge of explainable artificial intelligence in biometric systems.

Experimental validation against state-of-the-art methods. The proposed framework has been rigorously evaluated against existing FKP recognition approaches, demonstrating competitive or superior performance across all four finger modalities. For the right middle finger modality, the system achieved a recognition rate of 95.96%, surpassing all compared methods. For left middle finger recognition, the framework attained 95.86%, again outperforming existing approaches. These results validate the effectiveness of the proposed approach while establishing new benchmarks for explainable FKP recognition. The publication of these findings in a peer-reviewed journal confirms their significance within the research community.

Extension through edge-enhanced deep feature extraction. The investigation of edge detection techniques as complementary preprocessing methods has extended the capabilities of the basic framework. The comparison between Canny, Sobel, and Laplacian of Gaussian operators demonstrated how different approaches to enhancing structural patterns can affect recognition performance. The Hierarchical Prototype classifier, evaluated alongside conventional KNN classification, provided additional insights into the advantages of prototype-based approaches for biometric applications. This extension opens new directions for integrating classical image processing techniques with modern deep learning architectures.

6.2 Implications for Biometric Systems

The contributions of this thesis have implications that extend beyond the specific domain of finger knuckle print recognition. The broader significance of this work touches upon fundamental questions about how biometric systems should be designed, evaluated, and deployed in real-world applications.

Reconciling accuracy and interpretability. Perhaps the most important implication of this work is the demonstration that high recognition accuracy and full interpretability are not mutually exclusive goals. The prevailing assumption in the field has been that the most accurate systems would necessarily be the most opaque that the complexity required for superior performance precluded transparency. This thesis challenges that assumption by showing that a carefully designed architecture can achieve state-of-the-art accuracy while providing decisions that can be traced, understood, and verified by human users. For security-critical applications where both performance and accountability are essential, this finding has significant practical implications.

The value of prototype-based reasoning for biometric applications. The xDNN classifier's approach of representing each class through a collection of prototypes offers advantages that extend beyond interpretability. Unlike conventional classifiers that store statistical summaries or learned parameters, prototype-based systems retain actual representative samples that capture the natural variation within each class. This representation is particularly well-suited to biometric applications, where intra-class variation differences between multiple samples from the same individual is a fundamental challenge. By maintaining multiple prototypes per class, the system can accommodate the range of appearances that a single individual may present across different capture sessions, lighting conditions, and physiological states.

Robustness through illumination normalization. The integration of SQI preprocessing has implications for how biometric systems should handle the variability inherent in real-world deployment. Rather than attempting to learn invariance to illumination through massive training datasets an approach that requires extensive and carefully curated data the proposed framework explicitly normalizes lighting effects at the preprocessing stage. This strategy reduces the burden on subsequent learning components and improves generalization across diverse capture conditions. For practical deployments where controlled capture environments cannot be guaranteed, this approach offers a more reliable path to robust performance.

Transparency as a design principle. This work advocates for treating transparency as a fundamental design principle rather than an optional add-on. The xDNN architecture embeds explainability directly into its structure, ensuring that every classification decision can be understood in terms of the prototypes that influenced it. This stands in contrast to approaches that attempt to explain black-box models after the fact through approximation or visualiza-

tion techniques. For applications where decisions have real consequences border control, law enforcement, financial services the ability to audit and verify system behavior is not merely desirable but essential. By demonstrating that transparency can be achieved without sacrificing accuracy, this thesis provides a model for how future biometric systems might be designed.

Implications for user trust and system adoption. The explainability provided by the proposed framework has implications for how users perceive and interact with biometric systems. Research in human-computer interaction has consistently shown that users are more likely to trust and accept automated systems when they understand how decisions are made. The prototype-based explanations generated by xDNN presented as intuitive IF-THEN rules accompanied by visual examples make the system's reasoning accessible to non-experts. This transparency could accelerate the adoption of biometric technologies in applications where user acceptance has historically been a barrier.

6.3 Future Research Directions

While this thesis has addressed significant challenges in FKP recognition, it has also opened new questions and identified opportunities for further investigation. The following directions represent promising avenues for future research.

Extension to multimodal biometric fusion. The framework developed in this thesis has been evaluated on unimodal FKP recognition, treating each finger type as an independent modality. A natural extension would be to integrate multiple biometric traits combining FKP with palmprint, fingerprint, iris, or face recognition within the same explainable architecture. Multimodal fusion offers the potential for enhanced accuracy and robustness, as different modalities provide complementary information and are susceptible to different types of interference. The challenge lies in designing fusion strategies that maintain the interpretability of the overall system while effectively combining information from multiple sources. Feature-level fusion, score-level fusion, and decision-level fusion each offer different trade-offs between information preservation and architectural complexity. Investigating how xDNN's prototype-based reasoning can be extended to multimodal inputs represents a rich area for future work.

Integration of hybrid optimization and feature selection strategies. The parameter optimization conducted in this thesis, while comprehensive, relied on systematic ablation stud-

ies across a range of values. Future work could explore more sophisticated optimization approaches, including metaheuristic algorithms such as genetic algorithms, particle swarm optimization, or differential evolution. These techniques could automatically search the parameter space more efficiently, potentially identifying configurations that outperform manual exploration. Additionally, feature selection methods could be integrated to identify the most discriminative components of the PCANet feature vectors, reducing dimensionality and potentially improving both accuracy and computational efficiency. The combination of deep feature extraction with intelligent feature selection represents a promising direction for enhancing system performance.

Investigation of alternative deep feature extraction architectures. While PCANet has proven effective for FKP recognition, offering advantages in terms of simplicity and training efficiency, other lightweight architectures warrant investigation. ICANet, which uses independent component analysis rather than principal component analysis, might capture different aspects of the data distribution. Transform-based networks such as DCTNet or the deep scattering spectrum could provide representations with different invariance properties. Sparse autoencoders offer another approach to unsupervised feature learning that emphasizes reconstruction-based representation. Comparing these alternatives within the same explainable classification framework could reveal which feature extraction strategies are best suited to different biometric modalities and application requirements.

User-centered evaluation of explainability mechanisms. The explainability visualizations presented in this thesis demonstrate that the system can generate interpretable rules and prototype-based explanations. However, the ultimate value of these explanations depends on how effectively they communicate the system's reasoning to human users. Future work should include user-centered evaluations that assess whether the generated explanations actually improve understanding, trust, and decision-making in practical contexts. Such studies could compare different explanation formats rule-based, prototype-based, feature attribution based to determine which are most effective for different user populations and task requirements. Understanding how humans interact with and interpret explanations from AI systems is essential for designing transparent systems that truly serve their intended users.

Application to other biometric modalities and domains. While this thesis has focused on FKP recognition, the proposed framework is potentially applicable to other biometric modal-

ities. Face recognition, which similarly struggles with illumination variations and lacks interpretability in deep learning approaches, could benefit from the same combination of illumination normalization and prototype-based classification. Palmprint recognition, fingerprint identification, and even behavioral biometrics such as gait or keystroke dynamics might be addressed within the same architectural framework. Exploring the generalization of this approach across diverse biometric traits would provide valuable insights into its strengths and limitations.

Addressing emerging challenges in biometric security. As biometric systems become more widely deployed, new security challenges continue to emerge. Presentation attacks—attempts to fool the system using artificial replicas or recorded samples—represent a growing concern. Future work could investigate how the prototype-based representations learned by xDNN might be leveraged for liveness detection, distinguishing genuine biometric samples from spoofed presentations. Similarly, template security and cancelable biometrics—techniques for protecting stored templates and enabling revocation if compromised—could be integrated with the explainable framework. The transparency of xDNN might actually enhance template protection by making it easier to verify that transformation functions preserve discriminative information while preventing reverse engineering.

Real-world deployment and longitudinal evaluation. The experiments conducted in this thesis, while rigorous, have been performed on established benchmark databases under controlled conditions. Future work should include real-world deployment and longitudinal evaluation to assess how the system performs under the full range of conditions encountered in practical applications. Such studies would reveal whether the illumination normalization techniques generalize across diverse environments, how the system handles the natural aging of knuckle patterns over time, and whether the explainability mechanisms remain effective as the system scales to larger populations. Real-world deployment would also provide opportunities to gather user feedback and refine the system based on practical experience.

In conclusion, this thesis has demonstrated that finger knuckle print recognition can achieve both high accuracy and full interpretability through a carefully designed framework that integrates illumination-invariant preprocessing, hierarchical feature extraction, and prototype-based classification. The experimental results confirm that the proposed approach performs competitively with state-of-the-art methods while providing the transparency that conventional deep learning models lack. As biometric systems continue to play an increasingly important

role in security and identity management, the ability to understand and verify their decisions will become ever more critical. This work represents a step toward biometric systems that are not only powerful but also trustworthy—systems that can explain themselves to the people who rely on them.

References

[1] A. K. Jain, A. A. Ross, K. Nandakumar, and T. Swearingen, *Introduction to Biometrics*, 2nd ed. Springer, 2025.

[2] M. Fairhurst, *Biometrics: A Very Short Introduction*. Oxford University Press, 2018.

[3] Rutgers University, "Authentication: Biometric authentication," CS 419/Lecture Notes, 2025.

[4] ScienceDirect Topics, Engineering, "Biometric recognition," https://www.sciencedirect.com/topics/engineering/biometric-recognition.

[5] ScienceDirect Topics, Computer Science, "Biometric characteristic," https://www.sciencedirect.com/topics/computer-science/biometric-characteristic.

[6] M. E. Schuckers, *Computational Methods in Biometric Authentication: Statistical Methods for Performance Evaluation*, ser. Information Science and Statistics. Springer London, 2010. [Online]. Available: https://doi.org/10.1007/978-1-84996-202-5

[7] D. Zhang, Z. Guo, and Y. Gong, *Multispectral Biometrics: Systems and Applications*. Springer International Publishing, 2016. [Online]. Available: https://doi.org/10.1007/978-3-319-22485-5

[8] W. Yang, S. Wang, J. Hu, Z. Guanglou, J. Chaudhry, E. Adi, and C. Valli, "Securing mobile healthcare data: A smart card based cancelable finger-vein bio-cryptosystem," *IEEE Access*, vol. PP, pp. 1–1, 06 2018. [Online]. Available: https://doi.org/10.1109/ACCESS.2018.2844182

[9] T. Hafs, L. Bennacer, M. Boughazi, and A. Nait-Ali, "Empirical mode decomposition

for online handwritten signature verification," *IET Biometrics*, vol. 5, no. 3, pp. 190–199, 2016. [Online]. Available: https://doi.org/10.1049/iet-bmt.2014.0041

[10] F. u. A. A. Minhas, "Fingerprint based person identification and verification," Ph.D. dissertation, 08 2005. [Online]. Available: https://doi.org/10.13140/RG.2.2.30439.96165

[11] IDIAP Research Institute, "Structure of a biometric recognition system," https://www.idiap.ch/software/bob/docs/bob/bob.bio.base/v4.1.1/struct_bio_rec_sys.html, 2020.

[12] F. M. Bachay and M. H. Abdulameer, "Hybrid deep learning model based on autoencoder and cnn for palmprint authentication," *International Journal of Intelligent Engineering Systems*, vol. 15, no. 3, pp. 488–499, 2022. [Online]. Available: https://doi.org/10.22266/ijies2022.0630.41

[13] R. Dwivedi and S. Dey, "A novel hybrid score level and decision level fusion scheme for cancelable multi-biometric verification," *Applied Intelligence*, vol. 49, pp. 1016 – 1035, 2018. [Online]. Available: https://doi.org/10.1007/s10489-018-1311-2

[14] OpenBR Documentation, "Performance metrics," https://deepwiki.org/biometrics/openbr/Performance_metrics.

[15] N. E. Chalabi, A. Attia, A. Bouziane, and Z. Akhtar, "Face based person recognition mechanism using monogenic binarized statistical image features," *Multimedia Tools and Applications*, vol. 81, no. 18, pp. 25 657–25 674, 2022. [Online]. Available: https://doi.org/10.1007/s11042-022-12890-4

[16] T. Dunstone and N. Yager, Eds., *Individual Evaluation: The Biometric Menagerie*. Springer US, 2009, pp. 153–180. [Online]. Available: https://doi.org/10.1007/978-0-387-77627-9_8

[17] S. Minaee and Y. Wang, "Palmprint recognition using deep scattering network," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2017, pp. 1–4. [Online]. Available: https://doi.org/10.1109/ISCAS.2017.8050421

[18] M. Chaa, Z. Akhtar, and A. Attia, "3d palmprint recognition using unsupervised convolutional deep learning network and svm classifier," *IET Image Processing*, vol. 13,

no. 5, pp. 736–745, 2019. [Online]. Available: https://doi.org/10.1049/iet-ipr.2018.5642

[19] 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, 2020. [Online]. Available: https://doi.org/10.3390/s20195523

[20] S. Daas, A. Yahi, T. Bakir, M. Sedhane, M. Boughazi, and E.-B. Bourennane, "Multimodal biometric recognition systems using deep learning based on the finger vein and finger knuckle print fusion," *IET Image Processing*, vol. 14, no. 15, pp. 3859–3868, 2020. [Online]. Available: https://doi.org/10.1049/iet-ipr.2020.0491

[21] A. Attia, C. Mourad, Z. Akhtar, and Y. Chahir, "Finger kunckcle patterns based person recognition via bank of multi-scale binarized statistical texture features," *Evolving Systems*, vol. 11, 12 2020. [Online]. Available: https://doi.org/10.1007/s12530-018-9260-x

[22] A. Attia, Z. Akhtar, N. E. Chalabi, S. Maza, and Y. Chahir, "Deep rule-based classifier for finger knuckle pattern recognition system," *Evolving Systems*, vol. 12, no. 4, pp. 1015–1029, 2021. [Online]. Available: https://doi.org/10.1007/s12530-020-09359-w

[23] S. Chattopadhyay, S. Manna, S. Bhattacharya, and U. Pal, "Surds: Self-supervised attention-guided reconstruction and dual triplet loss for writer independent offline signature verification," 08 2022, pp. 1600–1606. [Online]. Available: https://doi.org/10.1109/ICPR56361.2022.9956442

[24] A. Rahman and M. Banavar, "A hybrid deep learning model for robust biometric authentication from low-frame-rate ppg signals," 2025. [Online]. Available: https://arxiv.org/abs/2511.04037

[25] Z. Qin, P. Zhao, T. Zhuang, F. Deng, Y. Ding, and D. Chen, "A survey of identity recognition via data fusion and feature learning," *Information Fusion*, vol. 91, pp. 694–712, 2023. [Online]. Available: https://doi.org/10.1016/j.inffus.2022.10.032

[26] R. Rasool, "Feature-level vs. score-level fusion in the human identification system," *Applied Computational Intelligence and Soft Computing*, vol. 2021, pp. 1–10, 06 2021. [Online]. Available: https://doi.org/10.1155/2021/6621772

[27] B. Ammour, L. Boubchir, T. Bouden, and M. Ramdani, "Face–iris multimodal biometric identification system," *Electronics*, vol. 9, no. 1, 2020. [Online]. Available: https://doi.org/10.3390/electronics9010085

[28] Y. Wang, D. Shi, and W. Zhou, "Convolutional neural network approach based on multimodal biometric system with fusion of face and finger vein features," *Sensors*, vol. 22, no. 16, 2022. [Online]. Available: https://doi.org/10.3390/s22166039

[29] S. Soleymani, A. Dabouei, H. Kazemi, J. Dawson, and N. Nasrabadi, "Multi-level feature abstraction from convolutional neural networks for multimodal biometric identification," 08 2018, pp. 3469–3476. [Online]. Available: https://doi.org/10.1109/ICPR.2018.8545061

[30] M. Martínez-Ramón, M. Ajith, and A. R. Kurup, *Deep Learning: A Practical Introduction*. John Wiley & Sons, 2024.

[31] S. Chakraverty, D. M. Sahoo, and N. R. Mahato, *Concepts of Soft Computing: Fuzzy and ANN with Programming*. Springer, 2019. [Online]. Available: https://doi.org/10.1007/978-981-13-7430-2

[32] A. Vannieuwenhuyze, *Intelligence artificielle vulgarisée : Le Machine Learning et le Deep Learning par la pratique*. ENI Editions, 2019.

[33] K.-L. Du, C.-S. Leung, W. H. Mow, and M. N. S. Swamy, "Perceptron: Learning, generalization, model selection, fault tolerance, and role in the deep learning era," *Mathematics*, vol. 10, no. 24, 2022. [Online]. Available: https://doi.org/10.3390/math10244730

[34] A. R. S. Parmezan, V. M. Souza, and G. E. Batista, "Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model," *Information Sciences*, vol. 484, pp. 302–337, 2019. [Online]. Available: https://doi.org/10.1016/j.ins.2019.01.076

[35] C. Ünsalan, B. Höke, and E. Atmaca, *Embedded Machine Learning with Microcontrollers: Applications on Arduino Boards*. Springer International Publishing, 2024. [Online].

Available: https://doi.org/10.1007/978-3-031-69421-9

[36] E. Charniak and A. Bohy, *Introduction au Deep Learning*. Dunod, 2021.

[37] D. Thakur, J. K. Saini, and S. Srinivasan, "Deepthink iot: The strength of deep learning in internet of things," *Artif. Intell. Rev.*, vol. 56, no. 12, 2023. [Online]. Available: https://doi.org/10.1007/s10462-023-10513-4

[38] F. He and D. Tao, *Foundations of Deep Learning*, ser. Machine Learning: Foundations, Methodologies, and Applications. Springer Singapore, 2024. [Online]. Available: https://doi.org/10.1007/978-981-16-8233-9

[39] M. Chen, Y. Bai, J. D. Lee, T. Zhao, H. Wang, C. Xiong, and R. Socher, "Towards understanding hierarchical learning: benefits of neural representations," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, ser. NIPS '20. Curran Associates Inc., 2020. [Online]. Available: https://arxiv.org/abs/2006.13436

[40] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [Online]. Available: https://doi.org/10.1038/nature14539

[41] O. S. Ekundayo and A. E. Ezugwu, "Deep learning: Historical overview from inception to actualization, models, applications and future trends," *Applied Soft Computing*, vol. 181, p. 113378, 2025. [Online]. Available: https://doi.org/10.1016/j.asoc.2025.113378

[42] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, and D. Zhang, "Biometrics recognition using deep learning: A survey," *Artificial Intelligence Review*, vol. 56, no. 8, pp. 8647–8695, 2023. [Online]. Available: https://doi.org/10.1007/s10462-022-10237-x

[43] J. Jin, "Convolutional neural networks for biometrics applications," in *SHS Web of Conferences*, vol. 144. EDP Sciences, 2022, p. 03013. [Online]. Available: https://doi.org/10.1051/shsconf/202214403013

[44] J. M. Ackerson, R. Dave, and N. Seliya, "Applications of recurrent neural network for biometric authentication & anomaly detection," *Information*, vol. 12, no. 7, p. 272, 2021. [Online]. Available: https://doi.org/10.3390/info12070272

[45] M. E. Oztemel and Ö. M. Soysal, "Eeg-based personal identification by special design

domain-adaptive autoencoder," *Sensors*, vol. 25, no. 20, p. 6457, 2025. [Online]. Available: https://doi.org/10.3390/s25206457

[46] S. B. V. Shweta Sinha, "A systematic review on generative adversarial networks (gans) for biometrics," *Library Progress International*, vol. 44, no. 3, 2024. [Online]. Available: https://doi.org/10.48165/bapas.2024.44.2.1

[47] I. D. Mienye and T. G. Swart, "A comprehensive review of deep learning: Architectures, recent advances, and applications," *Information*, vol. 15, no. 12, p. 755, 2024. [Online]. Available: https://doi.org/10.3390/info15120755

[48] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent neural networks: A comprehensive review of architectures, variants, and applications," *Information*, vol. 15, no. 9, p. 517, 2024. [Online]. Available: https://doi.org/10.3390/info15090517

[49] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735

[50] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder–decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2014, pp. 1724–1734. [Online]. Available: https://doi.org/10.3115/v1/D14-1179

[51] K. Berahmand, F. Daneshfar, E. S. Salehi, Y. Li, and Y. Xu, "Autoencoders and their applications in machine learning: a survey," *Artificial Intelligence Review*, vol. 57, no. 2, p. 28, 2024. [Online]. Available: https://doi.org/10.1007/s10462-023-10662-6

[52] P. Li, Y. Pei, and J. Li, "A comprehensive survey on design and application of autoencoder in deep learning," *Applied Soft Computing*, vol. 138, p. 110176, 2023. [Online]. Available: https://doi.org/10.1016/j.asoc.2023.110176

[53] A. Al-Yaari and Y. Deng, "Generative adversarial networks: A comprehensive survey," *Systems and Soft Computing*, vol. 8, p. 200460, 2026. [Online]. Available: https://doi.org/10.1016/j.sasc.2026.200460

[54] Z. Wang, Q. She, and T. E. Ward, "Generative adversarial networks in computer vision: A survey and taxonomy," *ACM Computing Surveys*, vol. 54, no. 2, pp. 37:1–37:38, 2021. [Online]. Available: https://doi.org/10.1145/3439723

[55] B. Ghosh, I. K. Dutta, M. W. Totaro, and M. A. Bayoumi, "A survey on the progression and performance of generative adversarial networks," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020, pp. 1–8. [Online]. Available: https://doi.org/10.1109/ICCCNT49239.2020.9225510

[56] O. Harris, M. Andrews, P. Allen, and A. Tripathi, "Effects and results of dropout layer in reducing overfitting with convolutional neural networks (cnn)," *World Journal of Advanced Engineering Technology and Sciences*, vol. 13, pp. 836–853, 12 2024. [Online]. Available: https://doi.org/10.30574/wjaets.2024.13.2.0584

[57] D. M. Montserrat, Q. Lin, J. Allebach, and E. J. Delp, "Training object detection and recognition cnn models using data augmentation," *Electronic Imaging*, vol. 29, no. 10, pp. 27–27, 2017. [Online]. Available: https://doi.org/10.2352/ISSN.2470-1173.2017.10.IMAWM-163

[58] L. Windrim, A. Melkumyan, R. J. Murphy, A. Chlingaryan, and R. Ramakrishnan, "Pretraining for hyperspectral convolutional neural network classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 5, pp. 2798–2810, 2018. [Online]. Available: https://doi.org/10.1109/TGRS.2017.2783886

[59] M. Krichen, "Convolutional neural networks: A survey," *Computers*, vol. 12, no. 8, p. 151, 2023. [Online]. Available: https://doi.org/10.3390/computers12080151

[60] R. Ibrahim and M. O. Shafiq, "Explainable convolutional neural networks: A taxonomy, review, and future directions," *ACM Computing Surveys*, vol. 55, no. 10, pp. 206:1–206:37, 2023. [Online]. Available: https://doi.org/10.1145/3563691

[61] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "Pcanet: A simple deep learning baseline for image classification?" *Trans. Img. Proc.*, vol. 24, no. 12, 2015. [Online]. Available: https://doi.org/10.1109/TIP.2015.2475625

[62] W. Xinshao and C. Cheng, "Weed seeds classification based on pcanet deep learning baseline," in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2015, pp. 408–415. [Online]. Available: https://doi.org/10.1109/APSPA.2015.7415304

[63] P. Angelov and E. Soares, "Towards explainable deep neural networks (xdnn)," *Neural Networks*, vol. 130, pp. 185–194, 2020. [Online]. Available: https://doi.org/10.1016/j.neunet.2020.07.010

[64] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1409.1556

[65] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, F. Bach and D. Blei, Eds., vol. 37. PMLR, 07–09 Jul 2015, pp. 448–456. [Online]. Available: https://proceedings.mlr.press/v37/ioffe15.html

[66] K. Tanaka, "Inferotemporal cortex and object vision." *Annual review of neuroscience*, vol. 19, pp. 109–39, 1996. [Online]. Available: https://doi.org/10.1146/ANNUREV.NE.19.030196.000545

[67] P. P. Angelov, X. Gu, and J. C. Príncipe, "Autonomous learning multimodel systems from data streams," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 4, pp. 2213–2224, 2018.

[68] X. G. Plamen P. Angelov, *Empirical Approach to Machine Learning*. Springer Cham, 2018. [Online]. Available: https://doi.org/10.1007/978-3-030-02384-3

[69] M. Markou and S. Singh, "Novelty detection: A review," *Signal Processing*, vol. 83, no. 12, pp. 2481–2497, 2003.

[70] B. Zeinali, A. Ayatollahi, and M. Kakooei, "A novel method of applying directional filter bank (dfb) for finger-knuckle-print (fkp) recognition," in *2014 22nd Iranian conference on electrical engineering (ICEE)*. IEEE, 2014, pp. 500–504. [Online]. Available: https://doi.org/10.1109/IranianCEE.2014.6999594

[71] M. Chaa, N.-E. Boukezzoula, and A. Meraoumia, "Features-level fusion of reflectance and illumination images in finger-knuckle-print identification system," *International Journal on Artificial Intelligence Tools*, vol. 27, no. 03, p. 1850007, 2018. [Online]. Available: https://doi.org/10.1142/S0218213018500070

[72] Tencent Cloud, "How do face recognition algorithms solve the problem of small sample training?" Technology Encyclopedia, 2025, https://www.tencentcloud.com/techpedia/120226.

[73] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4690–4699. [Online]. Available: https://doi.org/10.1109/CVPR.2019.00482

[74] J. Dan, Y. Liu, H. Xie, J. Deng, H. Xie, X. Xie, and B. Sun, "Transface: Calibrating transformer training for face recognition from a data-centric perspective," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 20 585–20 596. [Online]. Available: https://doi.org/10.1109/ICCV51070.2023.01887

[75] H. M. L. Aung, C. Pluempitiwiriyawej, K. Hamamoto, and S. Wangsiripitak, "Multimodal biometrics recognition using a deep convolutional neural network with transfer learning in surveillance videos," *Computation*, vol. 10, no. 7, 2022. [Online]. Available: https://doi.org/10.3390/computation10070127

[76] O. N. Kadhim and M. H. Abdulameer, "A multimodal biometric database and case study for face recognition based deep learning," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 1, pp. 677–685, 2024. [Online]. Available: https://doi.org/10.11591/eei.v13i1.6605

[77] S. Jia, D. H. Koh, A. Seccia, P. Antonenko, R. Lamb, A. Keil, M. Schneps, and M. Pomplun, "Biometric recognition through eye movements using a recurrent neural network," in *2018 IEEE International Conference on Big Knowledge (ICBK)*, 2018, pp. 57–64.

[78] X. Zhang, Y. Zhang, L. Zhang, H. Wang, and J. Tang, "Ballistocardiogram based person identification and authentication using recurrent neural networks," in *2018 11th International Congress on Image and Signal Processing, BioMedical*

Engineering and Informatics (CISP-BMEI), 2018, pp. 1–5. [Online]. Available: https://doi.org/10.1109/CISP-BMEI.2018.8633102

[79] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Biotouchpass2: Touchscreen password biometrics using time-aligned recurrent neural networks," *Trans. Info. For. Sec.*, vol. 15, p. 2616–2628, 2020.

[80] V. Chandrabanshi and S. Domnic, "A deep learning approach for strengthening person identification in face-based authentication systems using visual speech recognition," *The European Physical Journal Special Topics*, vol. 234, no. 3, pp. 1–15, 2025. [Online]. Available: https://doi.org/10.1140/epjs/s11734-025-01586-z

[81] P. H. Progga, M. J. Rahman, S. Biswas, M. S. Ahmed, A. R. Anwary, and S. Shatabda, "A bidirectional siamese recurrent neural network for accurate gait recognition using body landmarks," *Neurocomputing*, vol. 605, p. 128313, 2024.

[82] H. Qin, Z. Xiong, Y. Li, M. A. El-Yacoubi, and J. Wang, "Attention blstm-based temporal-spatial vein transformer for multi-view finger-vein recognition," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 9330–9343, 2024.

[83] Y. Qi, M. Qiu, H. Jiang, and F. Wang, "Extracting fingerprint features using autoencoder networks for gender classification," *Applied Sciences*, vol. 12, no. 19, 2022. [Online]. Available: https://doi.org/10.3390/app121910152

[84] Y. Liang and W. Liang, "Reswcae: Biometric pattern image denoising using residual wavelet-conditioned autoencoder," in *Neural Information Processing: 31st International Conference, ICONIP 2024, Auckland, New Zealand, December 2–6, 2024, Proceedings, Part X*. Berlin, Heidelberg: Springer-Verlag, 2025, p. 238–252. [Online]. Available: https://doi.org/10.1007/978-981-96-6603-4_17

[85] P. Selvaraj, A. Elliott, and C. Maple, "A dual-use autoencoder for fingerprint enhancement and feature extraction," *IET Conference Proceedings*, vol. 2025, pp. 200–203, 2025.

[86] T. Arıcan, R. Veldhuis, L. Spreeuwers, L. Bergeron, C. Busch, E. Jalilian, C. Kauba, S. Kirchgasser, S. Marcel, B. Prommegger, K. Raja, R. Ramachandra, and A. Uhl,

"A comparative study of cross-device finger vein recognition using classical and deep learning approaches," *IET Biometrics*, vol. 2024, no. 1, p. 3236602, 2024. [Online]. Available: https://doi.org/10.1049/2024/3236602

[87] C.-C. Wu, C.-W. Tsai, F.-E. Wu, C.-H. Chiang, and J.-C. Chiou, "Plantar pressure-based gait recognition with and without carried object by convolutional neural network-autoencoder architecture," *Biomimetics*, vol. 10, no. 2, 2025. [Online]. Available: https://doi.org/10.3390/biomimetics10020079

[88] S. K. Abbas, S. Purnapatra, M. G. Sarwar Murshed, C. Miller-Lynch, L. Igene, S. Dey, S. Schuckers, and F. Hussain, "Conditional synthetic live and spoof fingerprint generation," *IET Biometrics*, vol. 2026, no. 1, p. 7736489, 2026. [Online]. Available: https://doi.org/10.1049/bme2/7736489

[89] J. Ma and X. Chen, "Fingerprint image generation based on attention-based deep generative adversarial networks and its application in deep siamese matching model security validation," *Journal of Computational Methods in Engineering Applications*, vol. 4, p. 1–13, 2024. [Online]. Available: https://doi.org/10.62836/jcmea.v4i1.040107

[90] A. Kordas, E. Bartuzi-Trokielewicz, M. Ołowski, and M. Trokielewicz, "Synthetic iris images: A comparative analysis between cartesian and polar representation," *Sensors*, vol. 24, no. 7, 2024. [Online]. Available: https://doi.org/10.3390/s24072269

[91] S. Yadav and A. Ross, "iwarpgan: Disentangling identity and style to generate synthetic iris images," ser. 2023 IEEE International Joint Conference on Biometrics (IJCB). IEEE, 09 2023. [Online]. Available: https://doi.org/10.1109/ijcb57857.2023.10449250

[92] A. R. P N, R. Ramachandra, K. S. Rao, and P. Mitra, "Mlsd-gan - generating strong high quality face morphing attacks using latent semantic disentanglement," in *2023 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI)*, 2023, pp. 1–6. [Online]. Available: https://doi.org/10.1109/CVMI59935.2023.10464945

[93] Y. Belhocine, A. Meraoumia, K. Abderrazak, and M. Saigaa, "Efficient contactless palmprint recognition system based on deep rule-based classification," *Acta Informatica Pragensia*, vol. 13, no. 2, pp. 193–212, 2024. [Online]. Available: https://doi.org/10.18267/j.aip.236

[94] R. Chlaoua, A. Meraoumia, K. E. Aiadi, and M. Korichi, "Deep learning for finger-knuckle-print identification system based on pcanet and svm classifier," *Evolving Systems*, vol. 10, no. 2, pp. 261–272, 2019. [Online]. Available: https://doi.org/10.1007/s12530-018-9227-y

[95] G. Jaswal, A. Nigam, and R. Nath, "Deepknuckle: revealing the human identity," *Multimedia Tools and Applications*, vol. 76, no. 18, pp. 18 955–18 984, 2017. [Online]. Available: https://doi.org/10.1007/s11042-017-4475-6

[96] R. Hammouche, A. Attia, and S. Akhrouf, "Score level fusion of major and minor finger knuckle patterns based symmetric sum-based rules for person authentication," *Evolving Systems*, vol. 13, no. 3, pp. 469–483, 2022. [Online]. Available: https://doi.org/10.1007/s12530-022-09430-8

[97] M. Chaa, N.-E. Boukezzoula, and A. Attia, "Score-level fusion of two-dimensional and three-dimensional palmprint for personal recognition systems," *Journal of Electronic Imaging*, vol. 26, no. 1, pp. 013 018–013 018, 2017. [Online]. Available: https://doi.org/10.1117/1.JEI.26.1.013018

[98] J. Qian, J. Yang, Y. Tai, and H. Zheng, "Exploring deep gradient information for biometric image feature representation," *Neurocomputing*, vol. 213, pp. 162–171, 2016. [Online]. Available: https://doi.org/10.1016/j.neucom.2015.11.135

[99] R. Hammouche, A. Attia, S. Akhrouf, and Z. Akhtar, "Gabor filter bank with deep autoencoder based face recognition system," *Expert Systems with Applications*, vol. 197, p. 116743, 2022. [Online]. Available: https://doi.org/10.1016/j.eswa.2022.116743

[100] R. Cappelli, M. Ferrara, and D. Maltoni, "Minutia cylinder-code: A new representation and matching technique for fingerprint recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, pp. 2128–2141, 2010. [Online]. Available: https://doi.org/10.1109/TPAMI.2010.52

[101] A. Bera, D. Bhattacharjee, and M. Nasipuri, "Hand biometrics in digital forensics," in *Computational Intelligence in Digital Forensics: Forensic Investigation and Applications*. Springer, 2014, pp. 145–163. [Online]. Available: https://doi.org/10.1007/978-3-319-05885-6_8

[102] L. Zhang, L. Zhang, and D. Zhang, "Finger-knuckle-print: a new biometric identifier," in *2009 16th IEEE international conference on image processing (ICIP)*. IEEE, 2009, pp. 1981–1984. [Online]. Available: https://doi.org/10.5555/1818719.1819272

[103] K. Usha and M. Ezhilarasan, "Fusion of geometric and texture features for finger knuckle surface recognition," *Alexandria Engineering Journal*, vol. 55, no. 1, pp. 683–697, 2016. [Online]. Available: https://doi.org/10.1016/j.aej.2015.10.003

[104] A. Attia, A. Moussaoui, M. Chaa, and Y. Chahir, "Finger-knuckle-print recognition system based on features-level fusion of real and imaginary images," *Journal on Image and Video Processing*, 2018. [Online]. Available: https://doi.org/10.21917/ijivp.2018.0252

[105] Y. Belayadi, A. Attia, N. E. Chalabi, Z. Akhtar, and A. W. Mohamed, "Gabor filter bank features selection for face recognition based on particle swarm optimization," *Ingenierie des Systemes d'Information*, vol. 30, no. 3, p. 713, 2025. [Online]. Available: https://doi.org/10.18280/isi.300315

[106] H. Wang, S. Z. Li, Y. Wang, and J. Zhang, "Self quotient image for face recognition," in *2004 International Conference on Image Processing, 2004. ICIP'04.*, vol. 2. IEEE, 2004, pp. 1397–1400. [Online]. Available: https://doi.org/10.1109/ICIP.2004.1419763

[107] S. Kumar, M. Singh, and D. Shaw, "Comparative analysis of various edge detection techniques in biometric application," *International Journal of Engineering and Technology (IJET)*, vol. 8, no. 6, pp. 2452–2459, 2016. [Online]. Available: https://doi.org/10.21817/ijet/2016/v8i6/160806409

[108] A. Singh, M. Singh, and B. Singh, "Face detection and eyes extraction using sobel edge detection and morphological operations," in *2016 Conference on Advances in Signal Processing (CASP)*. IEEE, 2016, pp. 295–300. [Online]. Available: https://doi.org/10.1109/CASP.2016.7746183

[109] D. A. Muhtasim, M. I. Pavel, and S. Y. Tan, "A patch-based cnn built on the vgg-16 architecture for real-time facial liveness detection," *Sustainability*, vol. 14, no. 16, p. 10024, 2022. [Online]. Available: https://doi.org/10.3390/su141610024

[110] M. G. Alaslani, "Convolutional neural network based feature extraction for iris recognition," *International Journal of Computer Science & Information Technology (IJCSIT) Vol*, vol. 10, 2018. [Online]. Available: https://doi.org/10.5121/ijcsit.2018.10206

[111] B. Mandal, A. Okeukwu, and Y. Theis, "Masked face recognition using resnet-50," *arXiv preprint arXiv:2104.08997*, 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2104.08997

[112] A. Kumar, "Importance of being unique from finger dorsal patterns: Exploring minor finger knuckle patterns in verifying human identities," *IEEE transactions on information forensics and security*, vol. 9, no. 8, pp. 1288–1298, 2014. [Online]. Available: https://doi.org/10.1109/TIFS.2014.2328869

ملخص

أصبح التحقق البيومتري مكونا أساسيا في أنظمة الأمن الحديثة، حيث برز التعرف على بصمة مفصل الإصبع (FKP) كأحد الأساليب الواعدة نظرا لغنى وثبات الأنماط النسيجية التي يتميز بها، إضافة إلى إمكانية التقاطها باستخدام أجهزة منخفضة التكلفة. ومع ذلك، تعاني أنظمة FKP الحالية من حساسية عالية لتغيرات الإضاءة، إضافة إلى ضعف قابلية تفسير القرارات الناتجة عن نماذج التعلم العميق. تقدم هذه الأطروحة إطارا موحدا وقويا وقابلا للتفسير لمعالجة هذه التحديات من خلال دمج المعالجة المسبقة غير المتأثرة بالإضاءة، واستخراج الميزات بكفاءة، والتصنيف الشفاف. حيث يتم أولا استخدام خوارزمية صورة النسبة الذاتية (SQI) لتطبيع تأثيرات الإضاءة، ثم تطبيق نموذج (PCANet) ثنائي المراحل لاستخراج ميزات هرمية وتمييزية تجمع بين الخصائص المحلية والعالمية. بعد ذلك، يتم اعتماد شبكة عصبية عميقة قابلة للتفسير (xDNN) لإجراء التصنيف القائم على النماذج الأولية، مما يتيح إنتاج قواعد من نوع IF-THEN ونماذج مرئية تعزز فهم القرارات. ولتحسين الأداء بشكل إضافي، تم توسيع الإطار من خلال منهجية EDHP-FKP التي تدمج تقنيات كشف الحواف (Canny وSobel و LoG)، وتقيّم عدة مستخرجات ميزات عميقة (VGG-16 وAlexNet وResNet-50) إلى جانب واصفات Gabor، كما تعتمد مصنفا هرميا (HP) قائما على النماذج الأولية مع مقارنة أدائه بمصنف أقرب جار (NN). أظهرت النتائج التجريبية على قاعدة بيانات PolyU FKP أن الإطار المقترح يحقق دقة عالية (تصل إلى 95.96%) مع معدل خطأ منخفض، مع ضمان متانة في مواجهة تغيرات الإضاءة وتحسين ملحوظ في قابلية التفسير مقارنة بالنماذج التقليدية، مما يساهم في تطوير أنظمة تحقق بيومتري موثوقة وقابلة للتطبيق.

الكلمات المفتاحية: التعرف البيومتري، بصمة مفصل الإصبع (FKP)، صورة النسبة الذاتية (SQI)، PCANet، الشبكات العصبية العميقة القابلة للتفسير (xDNN)، كشف الحواف، المصنف القائم على النماذج الأولية الهرمية، استخراج الميزات العميقة، قابلية التفسير.

Résumé

L'authentification biométrique est devenue un élément fondamental des systèmes de sécurité modernes, et la reconnaissance des empreintes des articulations des doigts (FKP) apparaît comme une modalité prometteuse grâce à la richesse et à la stabilité de ses motifs texturaux, malgré une sensibilité aux variations d'éclairage et un manque d'interprétabilité dans les décisions des modèles d'apprentissage profond. Cette thèse propose un cadre unifié, robuste et explicable visant à surmonter ces limitations en intégrant un prétraitement invariant à l'éclairage, une extraction efficace des caractéristiques et une classification transparente. Plus précisément, l'algorithme Self-Quotient Image (SQI) est utilisé pour normaliser les effets d'illumination, suivi d'un modèle PCANet en deux étapes pour extraire des caractéristiques hiérarchiques et discriminantes capturant à la fois les structures locales et globales. Ensuite, un réseau neuronal profond explicable (xDNN) est employé pour une classification basée sur des prototypes, générant des règles interprétables de type IF–THEN et des prototypes visuels. Afin d'améliorer les performances, le cadre est étendu via l'approche EDHP-FKP, qui intègre un prétraitement basé sur la détection de contours (Canny, Sobel, LoG), évalue plusieurs extracteurs de caractéristiques profonds (VGG-16, AlexNet, ResNet-50) ainsi que des descripteurs de Gabor, et adopte un classifieur à prototypes hiérarchiques (HP) comparé au classifieur des plus proches voisins (NN). Les résultats expérimentaux obtenus sur la base de données PolyU FKP montrent que le cadre proposé atteint une précision élevée (jusqu'à 95,96%) avec un faible taux d'erreur, tout en garantissant une robustesse face aux variations d'éclairage et une meilleure interprétabilité par rapport aux modèles traditionnels, contribuant ainsi au développement de systèmes biométriques fiables et dignes de confiance.

Mots-clés: Reconnaissance biométrique, empreinte des articulations des doigts (FKP), image auto-quotient (SQI), PCANet, réseau neuronal profond explicable (xDNN), détection de contours, classifieur à prototypes hiérarchiques, extraction de caractéristiques, interprétabilité.

Abstract

Biometric authentication has become a fundamental component of modern security systems, with Finger Knuckle Print (FKP) recognition emerging as a promising modality due to its rich and stable texture patterns and low-cost acquisition. However, existing FKP systems suffer from sensitivity to illumination variations and a lack of interpretability in deep learning-based decisions. This thesis proposes a unified, robust, and explainable framework that addresses these challenges by integrating illumination-invariant preprocessing, efficient feature extraction, and transparent classification. Specifically, the Self-Quotient Image (SQI) algorithm is employed to normalize illumination effects, followed by a two-stage PCANet model to extract hierarchical and discriminative features capturing both local and global patterns. An Explainable Deep Neural Network (xDNN) is then used for prototype-based classification, providing interpretable IF–THEN rules and visual prototypes. To further enhance performance, the framework is extended through the EDHP-FKP approach, which incorporates edge-enhanced preprocessing using Canny, Sobel, and Laplacian of Gaussian operators, evaluates multiple deep feature extractors (VGG-16, AlexNet, ResNet-50) alongside Gabor descriptors, and integrates a hierarchical prototype (HP) classifier with comparative analysis against the Nearest Neighbor (NN) method. Extensive experiments on the PolyU FKP database demonstrate that the proposed framework achieves high recognition accuracy (up to 95.96%) with low Equal Error Rates, while ensuring robustness to illumination variations and significantly improving interpretability compared to conventional black-box models, thereby contributing to the development of reliable and trustworthy biometric authentication systems.

Keywords: Biometric recognition, Finger Knuckle Print (FKP), Self-Quotient Image (SQI), PCANet, Explainable Deep Neural Network (xDNN), Edge detection, Hierarchical Prototype classifier, Deep feature extraction, Interpretability.