PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

UNIVERSITY FERHAT ABBAS SETIF-1

FACULTY OF TECHNOLOGY

Domain: Science and Technology

THESIS

Presented to the Department of Electronics

For the award of the degree of

DOCTORATE

Field: Telecommunications **Option: Networks and Telecommunications**

By

Boudjellal Sif Eddine

THEME

Biometric Security System: Unimodal Identification Using Finger Veins.

Defended on 23/05/2024 before the jury composed of:

Ferhat Hamida Abdelhak	**Professor**	**Univ. F. Abbas Sétif 1**	**President**
Boukezzoula Naceur-Eddine	**Professor**	**Univ. F. Abbas Sétif 1**	**Thesis Supervisor**
Ziet Lahcene	**Professor**	**Univ. F. Abbas Sétif 1**	**Examiner**
Messaoudi Noureddine	**M.C.A**	**Univ. Boumerdes**	**Examiner**

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE FERHAT ABBAS SETIF-1

FACULTE DE TECHNOLOGIE

Domaine : Sciences et Technologie

THESE

Présentée au Département d'Electronique

Pour l'obtention du diplôme de

DOCTORAT

Filière : Télécommunications **Option : Réseaux et télécommunications**

Par

Boudjellal Sif Eddine

THEME

Système de sécurité biométrique :
Identification uni-modale par veines des doigts

Soutenue le 23/05/2024 devant le Jury:

Ferhat Hamida Abdelhak	Professeur	Univ. F. Abbas Sétif 1	Président
Boukezzoula Naceur-Eddine	Professeur	Univ. F. Abbas Sétif 1	Directeur de thèse
Ziet Lahcene	Professeur	Univ. F. Abbas Sétif 1	Examinateur
Messaoudi Noureddine	M.C.A	Univ. Boumerdes	Examinateur

Acknowledgment

First of all, I thank Allah the almighty, for allowing me to reach this modest scientific level and for giving me the courage and the patience to carry out the work done in this thesis.

I am indebted to my esteemed Supervisor, **Pr. Boukezzoula Naceur-Eddine**, for his unwavering guidance, supervision, motivation, and steadfast support. His exceptional mentorship has been instrumental in steering my efforts in the right direction.

I am also indebted to **Dr. Boudjelal Abdelwahhab** from Normandie University in Caen, France, whose invaluable guidance and constructive comments have been an honor and a pleasure to incorporate into my research.

The wisdom and insights shared by my thesis committee members, **Pr. Ferhat Hamida Abdelhak**, **Pr. Ziet Lahcene**, and **Dr. Messaoudi Nordine**, have significantly enriched the discourse of my thesis. Their engaging discussions, thoughtful suggestions, and constructive feedback have played a pivotal role in refining my work.

Special acknowledgment is extended to **Dr. Bousaadia Baadji** for his eminent support and unwavering encouragement throughout the entirety of my research project. I am deeply grateful for his impactful contributions.

I express genuine gratitude to my friends within the DZ-Writers community, as well as my friends and colleagues, for embracing me, demonstrating patience, offering steadfast support, and providing invaluable advice.

Finally, my deepest gratitude goes to my family, whose consistent support, encouragement, and assistance have been the bedrock of my academic journey.

Sif Eddine Boudjellal

Abstract

This thesis, "**Biometric Security System: Unimodal Identification Using Finger Veins**," explores the development and application of finger vein identification as a secure and efficient unimodal biometric recognition method. Leveraging advanced deep learning models, including InceptionResnet-V2 and a hybrid Convolutional Transformer-based approach (FVCT), the research establishes the potential for enhanced security and accuracy in biometric systems. Methodology involved the customization of deep learning architectures for finger vein identification, utilizing transfer learning and fusion of convolutional and transformer paradigms. Key findings demonstrate the superiority of these models, showcasing lower error rates and exceptional performance in comparison to existing state-of-the-art methods. Finger vein identification emerges as a reliable solution for diverse applications, from security to access control. Implications of these findings signify a path toward more secure and efficient biometric security systems. The fusion of deep learning paradigms and advancements in local feature extraction hold the promise of further innovation in the field. This research contributes to the ongoing development of robust and reliable personal identification solutions, ensuring enhanced security in critical domains.

Keywords: Biometrics, Finger Vein, Deep Learning, convolutional neural network (CNN), vision transformer.

Résumé

Cette thèse, intitulée "**Système de sécurité biométrique : Identification unimodale par les veines des doigts**", explore le développement et l'application de l'identification par les veines des doigts en tant que méthode de reconnaissance biométrique unimodale sécurisée et efficace. En exploitant des modèles avancés d'apprentissage en profondeur, notamment InceptionResnet-V2 et une approche hybride basée sur la convolution et le transformateur (FVCT), la recherche établit le potentiel d'une sécurité améliorée et d'une précision accrue dans les systèmes biométriques. La méthodologie a impliqué la personnalisation des architectures d'apprentissage en profondeur pour l'identification par les veines des doigts, en utilisant l'apprentissage par transfert et la fusion des paradigmes de convolution et de transformateur. Les principales conclusions démontrent la supériorité de ces modèles, avec des taux d'erreur plus faibles et des performances exceptionnelles par rapport aux méthodes de pointe existantes. L'identification par les veines des doigts émerge comme une solution fiable pour diverses applications, de la sécurité au contrôle d'accès. Les implications de ces conclusions signifient une voie vers des systèmes de sécurité biométrique plus sécurisés et efficaces. La fusion des paradigmes d'apprentissage en profondeur et les avancées dans l'extraction des caractéristiques locales laissent entrevoir la possibilité d'une plus grande innovation dans ce domaine. Cette recherche contribue au développement continu de solutions d'identification personnelle robustes et fiables, garantissant une sécurité renforcée dans des domaines critiques.

Mots Clé: biométrie, veine du doigt, apprentissage profond, réseau neuronal à convolution (CNN), transformateur de vision.

ملخص

تستكشف هذه الرسالة، بعنوان "**نظام الأمان البيومتري: التعرف الأحادي باستخدام الأوعية في الأصابع**"، تطوير وتطبيق تقنية التعرف باستخدام الأوعية في الأصابع كطريقة أمان فعّالة وآمنة. من خلال استغلال نماذج تعلم عميق متقدمة، بما في ذلك InceptionResnet-V2 ونموذج مبني على التكامل بين الشبكات العصبية المتسلسلة والمداخل (FVCT)، يوضح البحث إمكانية تعزيز الأمان والدقة في أنظمة التعرف البيومتري. المنهجية تتضمن تخصيص الهياكل العميقة للتعلم لتعرف الأوعية في الأصابع، مستفيدة من التعلم بالنقل ودمج بين الطرز العصبية المتسلسلة والمداخل. تشير النتائج الرئيسية إلى تفوق هذه النماذج، مع معدلات خطأ أقل وأداء استثنائي مقارنة بالأساليب الحالية المتقدمة. يظهر التعرف بواسطة الأوعية في الأصابع كخيار موثوق لتطبيقات متعددة، بدءًا من الأمان حتى مراقبة الوصول. تدل الآثار المترتبة على هذه النتائج على طريق نحو أنظمة أمان بيومترية أكثر أمانًا وفعالية. يمكن أن يكون دمج النماذج المبنية على التكامل بين الشبكات العصبية المتسلسلة والمداخل والتقدم في استخراج المعلومات المحلية مفيدًا لتحسين الأداء. تسهم هذه البحث في تطوير مستمر لحلول التعرف الشخصي الموثوقة والمتينة، مما يضمن تعزيز الأمان في المجالات الحيوية.

الكلمات المفتاحية: القياسات الحيوية، وريد الإصبع، التعلم العميق، الشبكة العصبية التلافيفية (CNN) ، محول الرؤية.

Contents

List of Figures

List of Tables

Acronyms

ACC: Accuracy

ANFIS: Adaptive Neuro-Fuzzy Inference System

ATM: Automated Teller Machines

AUC: Area Under the Curve

CCD: Charged Coupled Device

CNN: Convolutional Neural Network

CSS: Curvature Scale Space

DET: Detection Error Tradeoff

DNA: Deoxyribo Nucleic Acid

EER: Equal Error Rate

FN: False Negative

FAR: False Acceptance Rate

FRR: False Rejection Rate

FP: False Positive

FV: Finger Vein

FVT: Finger Vein Transformer

FVCT: Finger Vein Convolution-Transformer

GAR: Genuine Acceptance Rate

ID: Identification Card

kNN: K-Nearest Neighbor

LBP: Local Binary Pattern

LDA: Linear Discriminant Analysis

LSTM: Long Short-Term Memory

NIR: Near Infra-Red

NLP: Natural Language Processing

PCA: Principal Component Analysis

PIN: Personal Identification Number

PR: Precision-Recall

ResNet: Residual Neural Network

ROC: Receiver Operating Characteristic

R&D: Research and Development

ROI: Region Of Interest

SE: Squeeze-Excitation

SOTA: State-of-the-Art

SVM: Support Vector Machines

SVR: Support Vector Regression

TN: True Negative

TP: True Positive

ViT: Vision Transformer

CHAPTER 1

Introduction

In an era characterized by unprecedented digital connectivity, the protection of sensitive information and personal identities has never been more crucial. With the rapid expansion of online transactions, the proliferation of digital records, and the persistent threats of identity theft and cybercrime, the imperative for robust and secure authentication and identification methods is undeniable. Traditional means of identity verification, including passwords, personal identification numbers (PINs), and physical identification cards, have long been fraught with vulnerabilities that malicious actors can exploit. In response to these weaknesses, the field of biometric recognition has emerged as a pioneering solution, offering an unparalleled level of security through the analysis of an individual's unique physiological or behavioral traits[1, 2].

1.1 The Overarching Topic and Aims of the Thesis

The central focus of this Ph.D. thesis is the development and exploration of a biometric security system centered on the unimodal identification of individuals using the intricate patterns of finger veins. This innovative approach, termed "Finger Vein Identification," leverages the distinct patterns of blood vessels beneath the skin's surface in the human finger, making it a compelling and secure biometric modality. Our overarching aim is to advance the understanding and application of finger vein identification as a secure and efficient means of person recognition, addressing critical challenges and pushing the boundaries of its utility.

To ensure clarity and precision, let us define key terms and the scope of this thesis:

- Biometric Security System: A security system that relies on an individual's unique physiological or behavioral traits for the purpose of authentication and identification. In this thesis, we focus on the unimodal identification using finger vein patterns.

- Unimodal Identification: The process of identifying individuals using a single biometric modality, specifically the patterns of finger veins.

- Finger Veins: The intricate patterns of blood vessels beneath the skin's surface in the human finger, which are used as a biometric feature for identification.

The use of biometric recognition methods has proliferated in various authentication scenarios, including mobile devices, online payments, criminal investigations, and secure financial services. While fingerprint, iris, and facial recognition have been at the forefront of this technological transformation, finger vein identification has gained increasing prominence [3]. The following critical evaluation of the current state of the literature elucidates the unique attributes of finger vein identification and underscores the need for its further exploration.

Biometric recognition technology has evolved into an indispensable component of modern security and authentication systems. It offers enhanced efficiency compared to traditional identification methods, owing to its convenience and steadfast security. Unlike traditional methods that often rely on easily compromised means such as passwords or physical tokens, biometric systems authenticate individuals based on their distinctive physiological or behavioral traits. These traits can include, but are not limited to, fingerprints, iris patterns, facial features, voice characteristics, and finger vein patterns. The focus of this thesis is unimodal identification using finger vein patterns, a biometric modality that holds particular promise.

The technology of finger vein identification operates by analyzing the unique patterns of blood vessels located beneath the surface of an individual's finger, using near-infrared light for imaging. First pioneered by Hitachi's *R&D* department [4], finger vein identification has found diverse applications in healthcare, finance, automobile security, and confidential systems such as automated teller machines (ATMs). It is distinguished by several inherent advantages [5]:

- Enhanced Stability: The dispersed nature of finger veins under the skin's surface results in less variation depending on factors like an individual's age or weight. Additionally, the finger vein is shielded directly by the skin, preventing contamination by external factors and reducing susceptibility to damage.
- Inherent Difficulty of Usurpation: Due to the specific distribution and imaging circumstances of finger veins, obtaining finger vein images without the owner's consent is significantly challenging, making it resistant to spoofing.
- User-Friendly Operation: Finger vein authentication entails a straightforward process where users need only place one of their fingers on a finger vein device to perform fast and effortless identification.
- Liveliness Detection Capability: Finger vein imaging exhibits a distinct distribution of gray levels due to the veins' ability to absorb near-infrared light at a different rate than other finger tissues. This unique feature enables the detection of liveliness during the authentication process.
- Portability: Finger vein identification devices are designed to be compact, slightly larger than the size of a finger, making them easily portable and convenient.

While traditional biometric identification methods like fingerprint and iris recognition have seen substantial development, finger vein identification offers additional security features and advantages. These advantages include an enhanced level of stability, as finger veins are not prone to significant changes over time and resist spoofing attempts. Furthermore, finger vein identification does not require direct contact with the sensor, making it more hygienic in applications such as healthcare.

This chapter critically evaluates the existing body of research related to biometric recognition, emphasizing the unique attributes and advantages of finger vein identification. While traditional biometric modalities have seen widespread adoption and extensive research, finger vein identification holds particular promise for unimodal identification systems. By critically assessing the current state of the literature, we identify a gap in the research landscape related to unimodal identification using finger vein patterns. The importance of this research lies in its contribution to the advancement of biometric security systems, with a specific focus on unimodal identification using finger veins. As biometric recognition systems continue to play pivotal roles in security contexts, financial services, and various authentication scenarios, our research addresses the pressing need for robust, convenient, and secure identification methods. The key contributions and significance of this thesis are as follows:

- Advancing the Understanding of Finger Vein Identification: This research aims to enhance the understanding and application of finger vein identification as a secure and efficient biometric recognition modality. By delving into the historical development, techniques for feature extraction, and matching, we aim to bolster the security and accuracy of unimodal identification using finger veins.

- Evaluation of Deep Learning Models: This thesis involves the investigation of deep learning models, including the InceptionResnet-V2 model and hybrid Convolutional Transformer-based networks. Through comprehensive experiments, we aim to assess the performance of these models in finger vein identification, paving the way for improved unimodal identification.

- Identification of Research Gap: By critically evaluating the current state of the literature, we identify a research gap related to unimodal identification using finger vein patterns. This thesis seeks to fill this gap by conducting in-depth research and experimentation.

The contribution of this research extends to the enhancement of security, efficiency, and convenience in biometric recognition systems. By focusing on finger vein identification, we aim to improve the understanding and application of this technology, further enhancing its security and accuracy. Our work addresses the need for unimodal identification methods that are robust and secure, making it a valuable addition to the field of biometric security systems.

1.2 Describe The Methodology and Main Findings

In this thesis, our methodology involves a multifaceted exploration of finger vein identification, encompassing historical development, feature extraction techniques, and the application of deep learning models. We critically evaluate the current state of the literature, identify a research gap, and embark on a comprehensive journey to fill this gap.

Specifically, we investigate the performance of deep learning models in finger vein identification. Chapter 3 focuses on the InceptionResnet-V2 model, providing an in-depth exploration of its architecture and performance. Chapter 4 delves into the development and evaluation of a hybrid Convolutional Transformer-based model tailored for finger vein identification. Our experiments and findings aim to assess the capabilities of these models and their potential to enhance the accuracy of finger vein identification.

1.3 Thesis organisation

This thesis is structured as follows:

- Chapter 1 -Introduction: Provides an overview of the field of biometric recognition, the significance of finger vein identification, and the aims of the thesis.

- Chapter 2 -Biometric Systems: provides a comprehensive overview of biometric systems and their performance evaluation.

- Chapter 3 - Finger Vein Identification: Explores the intricacies of finger vein identification, including its historical development, techniques for feature extraction and matching, and the role of databases.
- Chapter 4 - InceptionResnet-V2 Model: Discusses a deep learning model for finger vein recognition based on InceptionResnet-V2 and its performance.
- Chapter 5 - Hybrid Convolutional Transformer Model: Explores a hybrid Convolutional Transformer-based model for finger vein identification and its results.
- Chapter 6 - Conclusion and Future Directions: Summarizes the key findings and contributions of the thesis and outlines potential avenues for future research in the field of finger vein identification.

In this thesis, we aim to advance the understanding and application of finger vein identification as a secure and efficient unimodal biometric recognition modality. Through comprehensive exploration and experimentation, we seek to enhance the security, accuracy, and convenience of unimodal identification using finger veins. Our research fills a critical gap in the current state of the literature and contributes to the broader field of biometric security systems.

CHAPTER 2

Overview of Biometric Systems

2.1 Introduction

Biometrics, derived from the Greek words "Bio" (meaning life) and "Metric" (to measure), represents a pioneering field offering a compelling solution for person recognition. Biometric systems stand as robust, highly secure, and inherently natural alternatives for verifying one's identity. The central objective of these systems revolves around the automation of human identification processes. Unlike traditional methods reliant on easily manipulated or compromised means such as badges, personal identification numbers (PINs), passwords (which can be words or phrases), and ID cards, biometric systems rely on an individual's distinctive physiological traits (e.g., fingerprint, iris, vein patterns, hand geometry, and ear shape) or behavioral characteristics (e.g., gait, signature, and keystroke dynamics)[1, 2, 6, 7, 8, 9, 10].

Identity verification systems have become indispensable in various domains, encompassing account logins, online payments, and automated teller machines (ATMs). These technologies are designed to safeguard user privacy and information security. The classical password, though widely used, suffers from drawbacks such as protracted

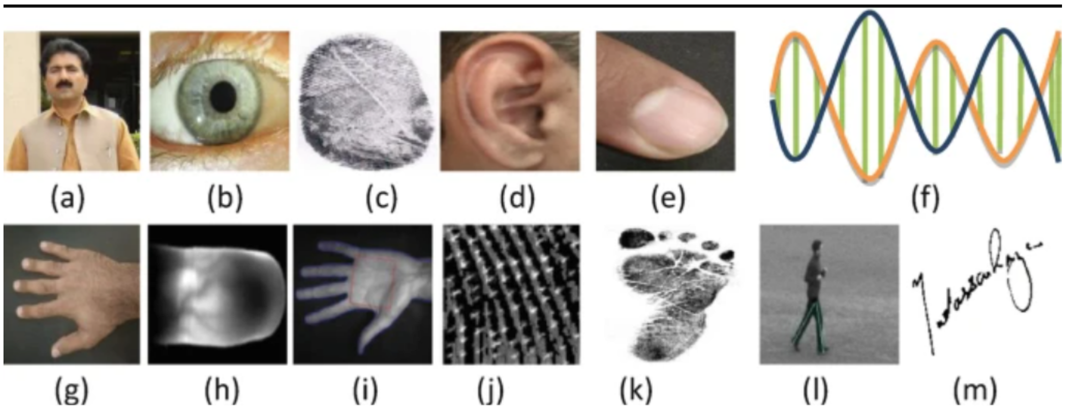


Figure 2.1: Biometric Types (a) face (b) iris (c) finger print (d) ear (e) nail bit (f) DNA (g) hand (h) finger vein (i) Palm vein (j) sweat pores (k) foot print (l) gait (m) signature [13]

input processes, vulnerability to leakage, and limited resistance to attacks. In the face of rapid advancements in information technology, biometric recognition systems have emerged as ubiquitous solutions across diverse authentication scenarios [11]. Biometric recognition revolves around the identification of individuals based on their unique physical and behavioral attributes, including facial features [1], vocal characteristics [2], and fingerprint patterns [6].

Biometric recognition technology offers enhanced efficiency compared to traditional identification methods, owing to its convenience and steadfast security. As the demand for digital security systems continues to surge, biometric recognition systems offer several compelling advantages. For instance, fingerprint-based biometric systems have found widespread adoption in home security, liberating individuals from the need to remember passwords or carry physical keys [8]. Furthermore, biometric recognition technologies, such as facial recognition, handwriting analysis, and voice recognition, play pivotal roles in criminal investigations [9]. Additionally, in the realm of financial services, biometric features, including finger vein patterns, facial recognition, and fingerprint scans, ensure that only authorized users gain access to sensitive financial data [10]. Beyond the aforementioned biometric modalities, iris scans, retinal scans, and gait analysis [12] also find extensive applications in various security contexts.

2.2 Operation of a Biometric System

A biometric system is fundamentally a pattern-recognition (pattern-matching) system. The process of identifying a user within a biometric system comprises two primary stages: enrollment and recognition. In the initial stage, enrollment involves collecting biometric data from an individual and storing it in a database alongside their identity. Typically, only the extracted feature set from the biometric data is retained in the database, while the raw biometric data is discarded. The second stage, recognition, entails re-collecting biometric data from the individual and comparing it against the feature set(s) stored in the database during enrollment to ascertain the user's identity. Consequently, a biometric system can be deconstructed into five fundamental modules: (a) sensor module, (b) quality assessment and feature extraction module, (c) database module, (d) matching module, and (e) decision module. Each of these key modules is elaborated upon below [2].

2.2.1 Sensor Module

To capture or measure the raw biometric data of an individual, a suitable biometric sensor is essential. For example, an optical sensor may be employed to acquire fingerprint images. In order to attain high-quality raw biometric data, the interface between the sensor and the user (human-machine interface) should be user-friendly and ergonomic. Furthermore, the choice of sensor characteristics plays a pivotal role in ensuring the acquisition of high-quality biometric samples.

2.2.2 Feature Extractor and Quality Assessment Module

The acquired biometric data typically undergoes further pre-processing before feature extraction. The feature extraction process aims to generate an informative digital representation, referred to as a template, from the input biometric sample. This template is expected to contain salient discriminatory information crucial for identifying or verifying the individual. During the enrollment stage, the template is registered either in the central database of the system or stored on a token, such as a smart card, issued to the individual. Given that the quality of the query biometric data (input data) is not always sufficient, a quality assessment algorithm is integrated into the biometric system to evaluate the suitability of the query data for subsequent processing. If the quality assessment deems the acquired biometric data unsuitable, the raw data is rejected, and re-acquisition from the user is initiated. In the absence of a quality assessment algorithm, the quality of the input data can be improved by subjecting it to signal enhancement techniques.

2.2.3 Database Module

The extracted features derived from the raw biometric data are stored in the system database (i.e., the template), along with some biographical information about the user (e.g., Personal Identification Number [PIN], name, address, etc.) that serves to distinguish them. To ensure secure storage of biometric templates, these templates should be housed in a centralized database, safeguarded through physical isolation and robust access control measures. This level of protection is essential to preserve the privacy of innocent users and guard against malicious individuals who may seek to exploit the biometric information stored in the database.

2.2.4 Matcher Module

The primary function of the biometric matcher module is to generate match scores by comparing the information extracted from the collected traits (query features) with their corresponding templates created during enrollment. The match score quantifies the degree of similarity between the two sets of features and can take the form of either a similarity or a distance metric. In cases where the matching module produces a similarity score, a higher matching score indicates greater similarity between the stored template and the input biometric sample. Conversely, a smaller distance matching score signifies a greater dissimilarity between the two feature sets. The matching module can perform two types of comparisons: one-to-one for verification purposes and one-to-many for

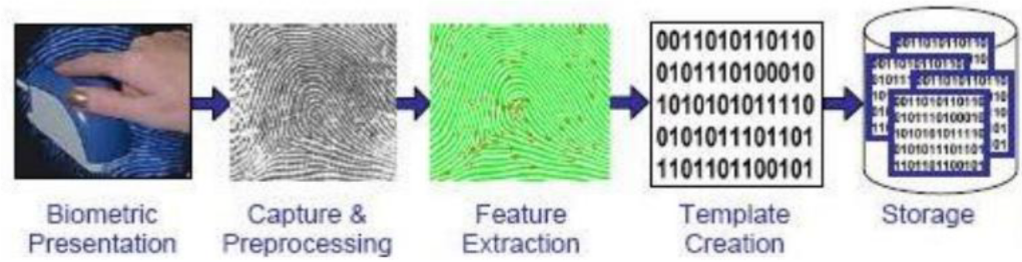


Figure 2.2: Biometric enrollment stage

identification.

2.2.5 Decision Module

In the decision module, the match scores generated by the matcher module are utilized to either validate the claimed individual's identity in the verification task or to rank the enrolled identities in order to identify a user in the identification task. Typically, the match score is compared with a predefined threshold, denoted as τ. If the match score falls within the threshold ($S < \tau$ in the case of similarity scores), the user is authenticated as genuine; otherwise, they are deemed an imposter. This threshold-based approach serves to distinguish between legitimate users and impostors based on the level of similarity or dissimilarity between their biometric data and the stored templates.

2.3 Functionalities of a Biometric System

A biometric system offers two primary modes of recognition: identification and verification (authentication is used synonymously with verification). Figure 2.2 illustrates the enrollment stage, where an individual presents their biometric traits (e.g., fingerprint, face, and iris) to the sensor for conversion into a reference template stored in the system database. The biometric system provides the following two modes, each of which is discussed below [1].

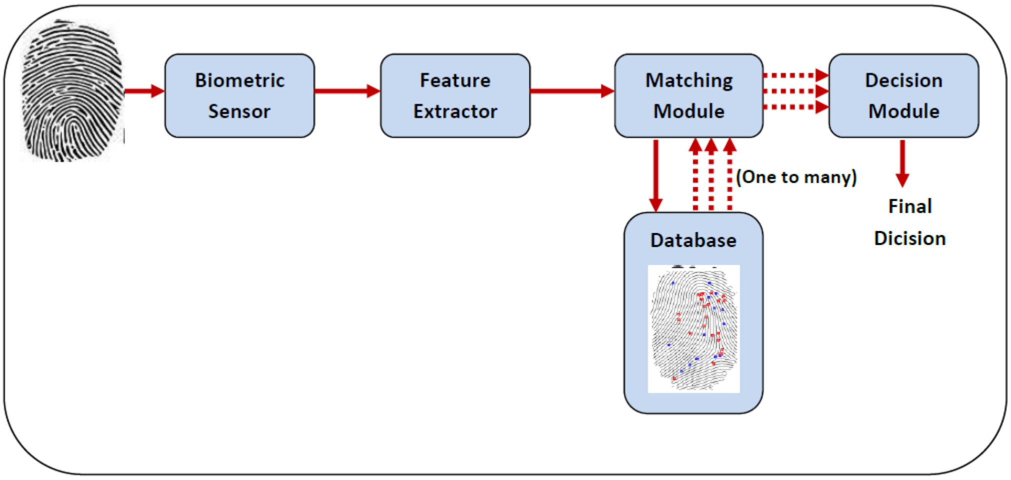


Figure 2.3: Biometric identifcation mode

2.3.1 Identification

In the identification mode, the biometric system conducts a comparison between the individual's biometric inputs and the templates of all users enrolled in the database (i.e., a one-to-many match) to determine the user's identity (refer to Figure 2.3). The system's output in this mode can either be the identity of the individual whose template exhibits the highest degree of similarity with the input sample provided by the user, or it may indicate that the person is not enrolled in the database. Several biometric systems operate in identification mode, such as the US-VISIT IDENT program and the FBI-IAFIS. Due to the substantial number of enrolled users, identification is notably more challenging than verification. Identification mode can be further categorized into two classes:

Positive Identification: In this class, the system determines the identity of the individual from a known set of identities. Essentially, the system answers the question, "Is this person someone who is known to the system?"

Negative Identification: In this class, the user is suspected of concealing their true identity, either explicitly or implicitly, from the system. This type of identification system is also known as screening, and its objective is to ascertain, "Is this person who they claim not to be?"

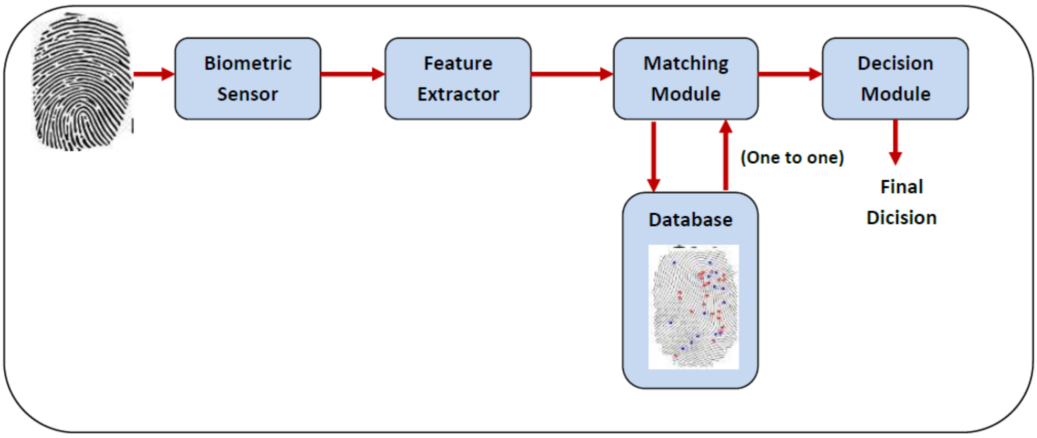


Figure 2.4: Biometric verification mode

2.3.2 Verification

In the verification mode, the biometric system performs a comparison solely between the individual's query input and their own biometric template stored in the database (i.e., a one-to-one match) to authenticate the user's claimed identity (see Figure 2.4). Typically, the identity claim is made using a username, a token (e.g., smart card), or a Personal Identification Number (PIN). The user is accepted as genuine if the query input matches their template with a high degree of similarity, and the degree of similarity exceeds a predefined threshold. This mode ensures that the user's claimed identity aligns with their biometric data on file, offering a heightened level of security for identity verification.

2.4 Selection of Biometric Modality

While biometrics finds applications in various daily scenarios, such as border crossing, mobile user authentication, and forensics, it's important to note that no single biometric trait satisfies all the requirements for these applications. However, several biometric traits are considered suitable. Table 2.1 provides a description of various biometric traits in terms of attributes like collectability, performance, distinctiveness, universality, and permanence. For instance, the fingerprint trait is characterized by medium universality, high distinctiveness, high permanence, medium collectability, high performance, and

Biometric Trait	Universal	Uniqueness	Permanence	Collectability	Performance	Acceptability
Face	H	L	M	H	L	H
Fingerprint	M	H	H	M	H	M
Ear	H	L	M	H	L	H
Iris	H	H	H	M	M	L
Gait	M	L	L	H	L	H
Hand vein	M	M	M	M	M	H
Hand Geometry	M	M	M	H	M	M
Retina	H	H	M	L	H	L
Signature	L	L	L	H	L	H
Voice	M	L	L	M	L	H
DNA	H	H	H	L	L	L

Table 2.1: Comparison of Biometric Traits [2]

Keys:
H: high, M: medium, L: low.

medium acceptability.

Multiple biometric traits have been employed for human identity verification, including fingerprint, face, voice, and palmprint, among others. Each biometric trait comes with its own strengths and weaknesses, and the choice of a biometric modality depends on various factors beyond accuracy performance [2]. The factors influencing the suitability of a biometric trait for a particular application can be summarized as follows [1, 8, 9]:

2.4.1 Universality

Universality implies that each user must possess the required biometric trait for successful enrollment. It's important to note that the universality factor directly impacts the failure to enroll rate.

2.4.2 Uniqueness

To minimize the false match rate (FAR) of a biometric system, the chosen biometric trait should effectively distinguish between different users.

2.4.3 Permanence

In order to achieve a high recognition rate in a biometric system, the user's biometric trait should remain sufficiently invariant over time. Failure to meet this criterion can result in a high false non-match rate (FRR).

2.4.4 Collectability

Collectability or measurability refers to the biometric modality's suitability for capture and its comfort for the individual to present to the biometric sensor.

2.4.5 Performance

This factor encompasses the accuracy, speed, and robustness of the system. The accuracy of biometric systems is typically defined by their false acceptance and false rejection rates. Accuracy can be influenced during the data collection process by environmental factors such as lighting, shadows, and background noise.

2.4.6 Resistance to Circumvention

This factor assesses the degree of resistance a biometric modality offers against spoofing attacks [12, 14]. Spoofing is the fraudulent process by which a user attempts to subvert or attack a biometric system by impersonating a registered user, gaining illegitimate access, and reaping advantages.

2.4.7 Acceptability

Acceptability gauges the level of public acceptance and approval for a given biometric trait. It is crucial that individuals are willing to present their biometric trait to the system, as user acceptance is a vital factor determining the success of any biometric implementation.

2.5 Comparison of Biometric Traits and Their Applications

Biometrics has established itself as a vital security tool in numerous applications within our interconnected society, as depicted in Table 2.2. Questions such as "Is she really who she claims to be?", "Is this person authorized to use this facility?", or "Is he on the government's watch list?" are routinely posed in a wide array of scenarios, spanning from issuing driver's licenses to controlling entry into a country (refer to Figure 2.5). The applications of biometrics can be broadly categorized into three main sectors: the Government sector, the Commercial sector, and the Forensic sector [1, 9, 10].

For a comprehensive understanding, the behavioral and psychological biometric traits, along with their characteristics, are detailed in Appendix A.1. Additionally, Table 2.2 provides an overview of the various applications associated with each biometric trait. These tables serve as valuable references for assessing the suitability of specific biometric modalities in different scenarios.

2.6 Unimodal And Multimodal Biometric Systems

In recent decades, biometric systems have seen significant advancements and have become increasingly vital in addressing security concerns across various domains, including access control, identity verification, and financial transactions. Among the noteworthy developments in biometrics, the debate surrounding unimodal and multimodal biometric systems has garnered substantial attention. This chapter provides an extensive comparative analysis of unimodal and multimodal biometric systems, delving into their respective strengths, weaknesses, and practical applications.

2.6.1 Unimodal Biometric Systems

Unimodal biometric systems rely on a single physiological or behavioral trait for identity verification. Common unimodal biometric modalities include fingerprint recogni-

Figure 2.5: Government and commercial applications that employ biometrics to recognize person (a) The US-VISIT program (b) the Schiphol Privium program, (c) Unique Identity (UID) Card project , and (d) a product by Fujitsu captures the palm vein pattern [1].

Biometric Trait	**Applications**
Face	• Criminal Identification • Access Control Verification • Human-Computer Interaction • Surveillance
Fingerprint	• License and Visa Authentication • Access Control Verification • Human Computer Interaction • Law Enforcement Forensics
Retina	• Security agencies such as FBI, CIA, and NASA
Ear	• Law Enforcement • Forensics Surveillance
Iris	• Identification as Aadhaar card in India • Access Control
Voice	• Web-based transactions • Voice Response-based health and banking systems
Gait	• Chiropractic • Medical diagnosis
Vein Pattern	• Financial systems and banks • Door security systems • Travel and transportation
Palmprint	• Personal Identification • Blood relation Identification • Medical Diagnosis • Selection of athletes
Signature	• Banking system

Table 2.2: Applications of Biometric Traits [1, 9]

tion, iris scanning, facial recognition, voice recognition, and hand geometry. These systems capture, process, and compare the unique features of the chosen modality to verify an individual's identity. The biometric sample is acquired through a sensor, digitized into a biometric template, and compared against an enrolled template for matching. Unimodal systems employ pattern recognition and machine learning algorithms to extract distinctive features and accurately match templates.

2.6.1.1 Sensor Technologies in Unimodal Systems

Various sensor technologies are employed in unimodal biometric systems to acquire raw biometric data. For instance, fingerprint recognition utilizes optical, capacitive, ultrasonic, and thermal sensors to capture ridge patterns. Iris scanning relies on high-resolution near-infrared cameras to photograph intricate iris textures. Facial recognition systems use standard digital cameras and sophisticated 3D sensors to capture facial images. Voice recognition records and digitizes vocal characteristics. Advances in sensor technologies have enabled more reliable and convenient biometric trait capture.

2.6.1.2 Strengths of Unimodal Biometric Systems

Simplicity and Ease of Deployment: Unimodal biometric systems offer the primary advantage of simplicity and ease of deployment. They typically require minimal hardware and software resources, making them cost-effective for a wide range of applications. Moreover, unimodal systems are often more user-friendly since they involve capturing a single biometric trait, reducing the complexity of the authentication process. Unimodal systems are, therefore, well-suited for integration into existing infrastructure and for consumer applications where usability is critical.

High Recognition Accuracy: Unimodal biometric systems can achieve high recognition accuracy when the selected modality is well-suited to the application, and the enrollment process is carefully controlled. For example, fingerprint recognition is known for its exceptional accuracy due to the uniqueness and permanence of friction ridge patterns. With high-resolution sensors, advanced image processing, and machine learning algorithms, fingerprint verification systems can operate with false match rates as low as 0.00008%. Such exceptional accuracy makes unimodal fingerprint systems ideal for stringent security applications.

2.6.1.3 Limitations of Unimodal Biometric Systems

Vulnerability to Spoofing: Unimodal biometric systems are susceptible to spoofing attacks, where malicious actors attempt to deceive the system by presenting counterfeit biometric traits. For instance, fingerprint sensors can be tricked with artificially fabricated fingerprints made using materials like gelatin and silicone. Iris spoofing can be done using high-quality printed images or prosthetic contact lenses. Without liveness detection capabilities, unimodal systems remain vulnerable to such spoofing threats.

Limited Robustness: Unimodal biometric systems may struggle in scenarios where the selected modality is affected by environmental factors or changes in the user's condition. For example, facial recognition systems can be less accurate in low-light conditions or when users wear masks, heavy makeup, or eyeglasses. Fingerprint systems perform poorly if the users' fingers are wet, dirty, or injured. Such variability can degrade the performance of unimodal systems.

Lack of Population Coverage: Certain unimodal biometric traits may not be viable for all individuals in the target population. For instance, fingerprint and iris recognition exclude individuals with damaged friction ridges or iris occlusion. Unimodal voice recognition is affected by voice disorders. By relying on a single modality, these systems fail to provide universal coverage.

2.6.2 Multimodal Biometric Systems

2.6.2.1 Definition and Characteristics

Multimodal biometric systems integrate two or more biometric modalities to enhance security and accuracy. These modalities can be either physiological (e.g., fingerprint, face, iris) or behavioral (e.g., voice, gait, signature). Multimodal systems combine information from multiple sources to create a more comprehensive identity assertion.

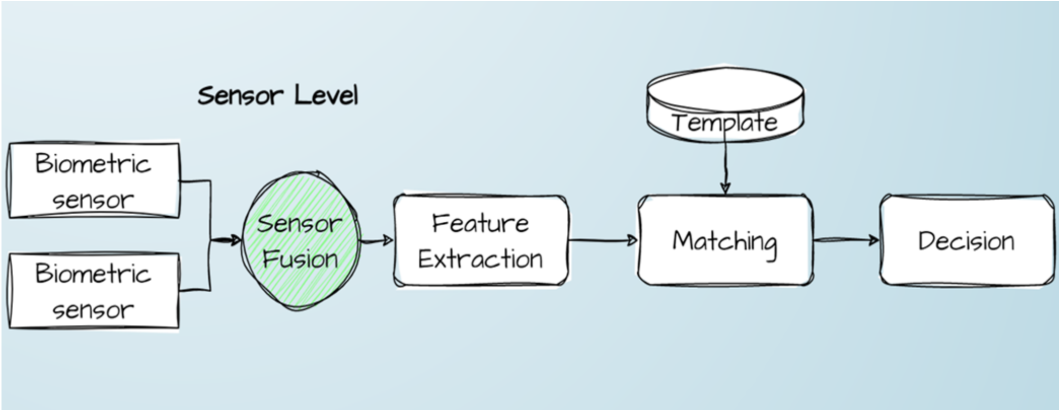


Figure 2.6: Block diagram of sensor level fusion in Multimodal Systems

The fusion of modalities can occur at the sensor, feature, matching score, or decision level.

2.6.2.2 Sensor Fusion in Multimodal Systems

Sensor fusion in multimodal systems involves using multiple biometric sensors to capture different modalities simultaneously (refer to Figure 2.6). For example, mobile devices can integrate fingerprint, face, and iris recognition sensors to create a multimodal system. Sensor fusion provides convenience to the user while also enabling liveness detection. However, employing multiple sensors increases the cost and form factor of devices.

2.6.2.3 Feature Fusion in Multimodal Systems

Feature fusion consolidates the feature sets extracted from multiple modalities into a single feature vector (refer to Figure 2.7). This enables complementary feature information to be combined for better discrimination between individuals. For instance, local binary patterns from face images can be combined with minutiae points from fingerprints. Efficient feature selection and weighting schemes are required to optimize the fused feature set.

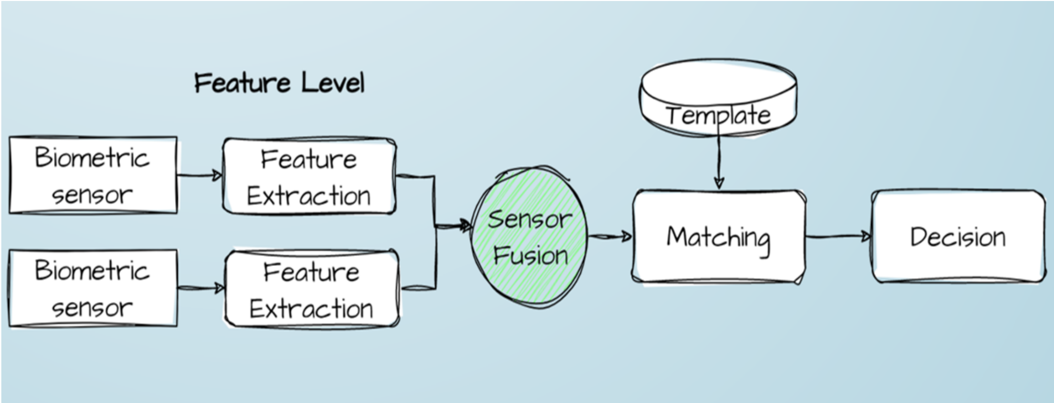


Figure 2.7: Block diagram of Feature level fusion in Multimodal Systems

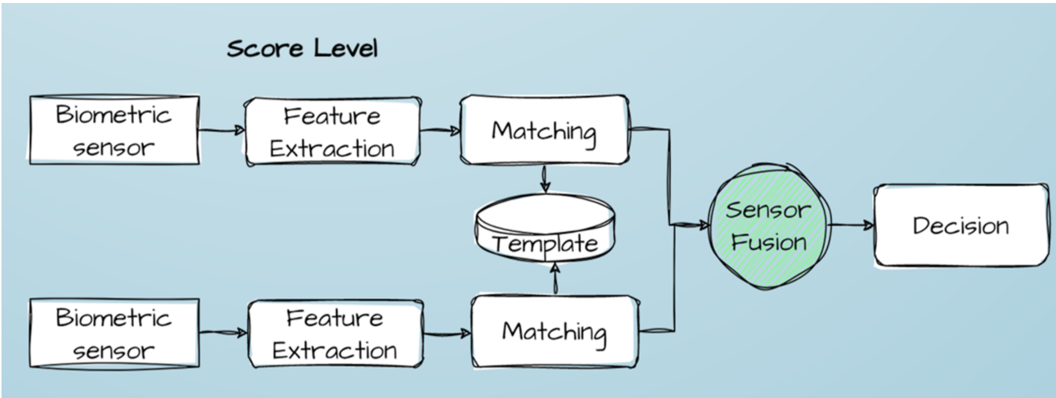


Figure 2.8: Block diagram of Matching Score Fusion in Multimodal Systems

2.6.2.4 Matching Score Fusion in Multimodal Systems

Matching score fusion aggregates the similarity scores obtained by comparing templates from each modality to the corresponding enrolled template (refer to Figure 2.8). The individual scores are normalized and combined using methods like the sum rule, weighted sum rule, or SVM classification. Matching score fusion provides the flexibility to adjust fusion rules for optimal performance.

2.6.2.5 Decision Fusion in Multimodal Systems

In decision fusion, each modality makes an independent authentication decision, which is then consolidated using techniques like majority voting, AND/OR rules, or meta-classification. By fusing final decisions, this method allows deploying matchers with their own optimized thresholds and parameters (refer to Figure 2.9). However, useful matching score information is lost prior to fusion.

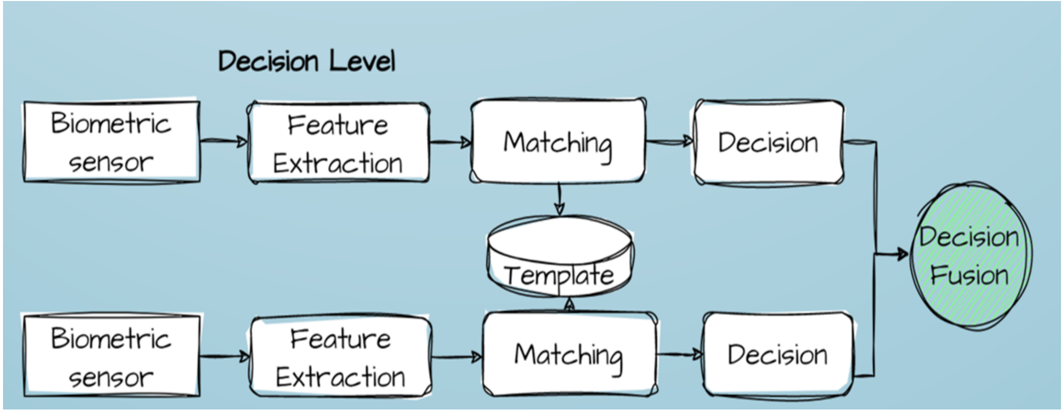


Figure 2.9: Block diagram of Decision Fusion in Multimodal Systems

2.6.2.6 Strengths of Multimodal Biometric Systems

Enhanced Security: Multimodal biometric systems offer significantly enhanced security against spoofing compared to unimodal systems. Performing a successful spoof attack requires compromising multiple modalities simultaneously, exponentially increasing the difficulty for imposters. Liveness detection capabilities can also be readily incorporated.

Increased Accuracy: By utilizing complementary biometric traits, multimodal systems can achieve higher accuracy. For instance, fusing fingerprints and iris can reduce false accept rates to extremely low levels. Multimodal systems also experience lower failure to enroll rates as multiple modalities provide redundancy.

Improved User Convenience: Multimodal systems can use non-intrusive modalities like face and voice along with highly accurate modalities like fingerprints for user convenience as well as accuracy. The ability to authenticate passively and continuously provides a seamless user experience.

2.6.2.7 Limitations of Multimodal Biometric Systems

Increased Cost and Complexity: The need for multiple sensors and computational mechanisms for fusion increases hardware and software costs. Enrollment and authentication processes require additional time to capture and process multiple modalities. System integration and maintenance complexity also increase.

User Acceptance: Collecting and integrating biometric data from multiple sources raises user concerns regarding privacy invasion. Seamless unobtrusive authentication can alleviate such concerns. Gradual introduction and education about benefits over unimodal systems can also foster user acceptance.

Optimization Difficulties: Determining optimal fusion rules, feature selection, matching algorithms, and their parameters for different modalities poses design challenges. Suboptimal fusion can potentially degrade overall accuracy. Adaptive optimization techniques are required for robust fusion.

2.7 Biometric Systems Performance Evaluation

The evaluation of biometric systems' performance represents a pivotal and indispensable facet in the design and architecture of biometric recognition systems. This section delves into the techniques for analyzing biometric systems and elucidates various metrics and graphical representations that shed light on the intricacies of biometric system operations. As previously alluded to, biometric systems can be categorized into two primary modes: verification and identification. It is imperative to differentiate between these two modes, as they exert substantial influence on the evaluation of performance.

The field of biometrics offers an array of solutions for addressing image classification problems [15]. These methods are adaptable to classification problems involving two or more classes, and the performance of classifiers is contingent upon the number of samples per class and their composition. Consequently, the choice of the most suitable method hinges on the specific requirements of the targeted application. A pragmatic approach involves initial method selection, followed by rigorous testing and subsequent evaluations.

In data analysis, the initial step typically involves the construction of an array representation known as a "confusion matrix." This table (Table 2.3) quantifies the number of predictions, denoted as $X_{i,j}$ (or $X_{\text{class,prediction}}$), representing samples of class i assigned to class j among a set of C classes. The number of samples constituting class i is denoted as K_i, and the total number of predictions attributed to this class is referred to as

		Prediction			Total /Classes
		$Class_1$	$Class_i$	$Class_c$	
Real Class	$Class_1$	$X_{1,1}$	$X_{1,i}$	$X_{1,c}$	K_1
	$Class_2$	$X_{i,1}$	$X_{i,i}$	$X_{i,c}$	K_i
	$Class_c$	$X_{c,1}$	$X_{c,i}$	$X_{c,c}$	K_c
Total Predictions		M_1	M_i	M_c	$\sum$

Table 2.3: Prediction Confusion Matrix of a C-Class Classifier

		Prediction		Total /Classes
		Positive Class	Negative Class	
Real Class	Positive Class	Tp	Fn	P
	Negative Class	Fp	Tn	N
Total Predictions		P_{pos}	P_{neg}	$\sum$

Table 2.4: Prediction Confusion Matrix of a C-Class Classifier

M_i. The sums of K_i and M_i collectively amount to the total number of samples ($\sum$).

With this context, for each class i, treated as a binary problem (Class i as positive, all other classes $i \neq j$ as negative), or directly for a two-class problem, the predictions can be classified into four principal categories:

1. **True Positive (Tp)**: Samples of the positive class (i) correctly classified ($X_{i,i}$).

2. **False Negative (Fn)**: Samples of the positive class (i) incorrectly classified (($X_{i,j}$,$\forall\ i \neq j$).

3. **True Negative (Tn)**:Samples of the negative class (j) correctly classified ($X_{i,t}$,$\forall$ $t \in [1,C] \neq i$).

4. **False Positive (Fp)**: Samples of the negative class (j) incorrectly classified ($X_{j,i}$,$\forall\ j \neq i$).

In the case of a problem with N classes, treated individually as binary problems, confusion matrices are constructed for each class i. The confusion matrix for a two-class problem establishes a connection between the total number of samples (P) from the positive class, the total number of samples (N) from the negative class, and the four aforementioned categories, which in turn determine the total number of samples classified as positive (P_{pos}) and negative (P_{neg}).

Various measures can be derived from a confusion matrix, from the problem with Two-classes we can describe the following metrics:

– **False Acceptance Rate (FAR)**: Defined as the probability that the biometric security system mistakenly accepts an access attempt by an unauthorized user.

$$\mathrm{F}AR = \frac{Fp}{Tn+Fp} = \frac{Fn}{N} \tag{2.1}$$

– **False Rejection Rate (FRR)**: Defined as the probability that the biometric security system mistakenly reject an access attempt by an authorized user name.

$$\mathrm{F}RR = \frac{Fn}{Tp+Fn} = \frac{Fn}{P} \tag{2.2}$$

– **Sensitivity**: is calculated as the number of correct positive predictions divided by the total number of positives. It is also called recall or True Positive Rate or Genuine Acceptance Rate (GAR) witch is given by $GAR = 1 - FRR$.

$$\mathrm{S}ensitivity = \frac{Tp}{Tp+Fn} = \frac{Tp}{P} \tag{2.3}$$

– **Specificity**: is calculated as the number of correct negative predictions divided by the total number of negatives. It is also called true negative rate. It can also be calculated by $(1 - specificity = FAR)$.

$$\mathrm{S}pecificity = \frac{Tn}{Tn+Fp} = \frac{Tn}{N} \tag{2.4}$$

– **Precision**: is calculated as the number of correct positive predictions divided by the total number of positive predictions. It is also called positive predictive value.

$$\mathrm{P}recision = \frac{Tp}{Tp+Fp} = \frac{Tp}{Ppos} \tag{2.5}$$

– **Equal Error Rate** (EER): is calculated as the number of all incorrect predictions divided by the total number of the classes. EER defined also as the best compromise between FAR and FRR . The best error rate is 0.0, whereas the worst is 1.0.

$$\mathrm{E}ER = \frac{Fp+Fn}{Tp+Tn+Fp+Fn} = \frac{Fp+Fn}{P+N} \tag{2.6}$$

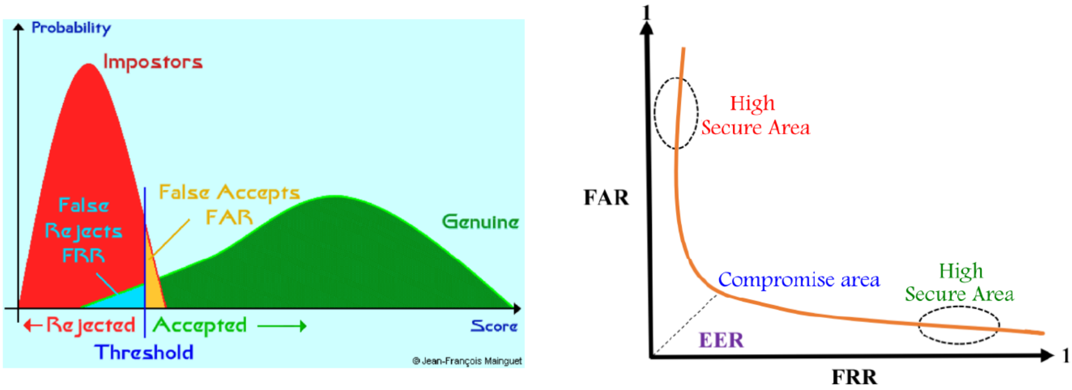
– **Accuracy (ACC)**: is calculated as the number of all correct predictions divided by the total number of the dataset. The best accuracy is 100%, whereas the worst is 0.0.

$$\mathrm{ACC} = \frac{Tp+Tn}{Tp+Tn+Fp+Fn} \tag{2.7}$$

Each of this metrics has a percentage describing a certain capability of the model. The higher the percentage value, the better the model. Sensitivity and specificity only take into account samples from the same test class (positive class for sensitivity and negative class for specificity). Thus, variations in the number of test images per class have no influence on this metrics. However, this is not the case for precision and accuracy. Indeed, the precision takes into account the test samples determined as positive for both classes and the accuracy is a "global" evaluation of the model, considering all the prediction results (the whole confusion matrix). The exploitation of previously described metrics are basic biometric performance measures such as the FRR/FAR, sensitivity/specificity and precision/sensitivity (or recall) pairs. Figure 2.10.a illustrate match score distributions for FRR/FAR by the use of different thresholds. These thresholds, applied to the prediction scores, make it possible to adjust these metrics by considering a prediction as just if its associated score is higher than this threshold.

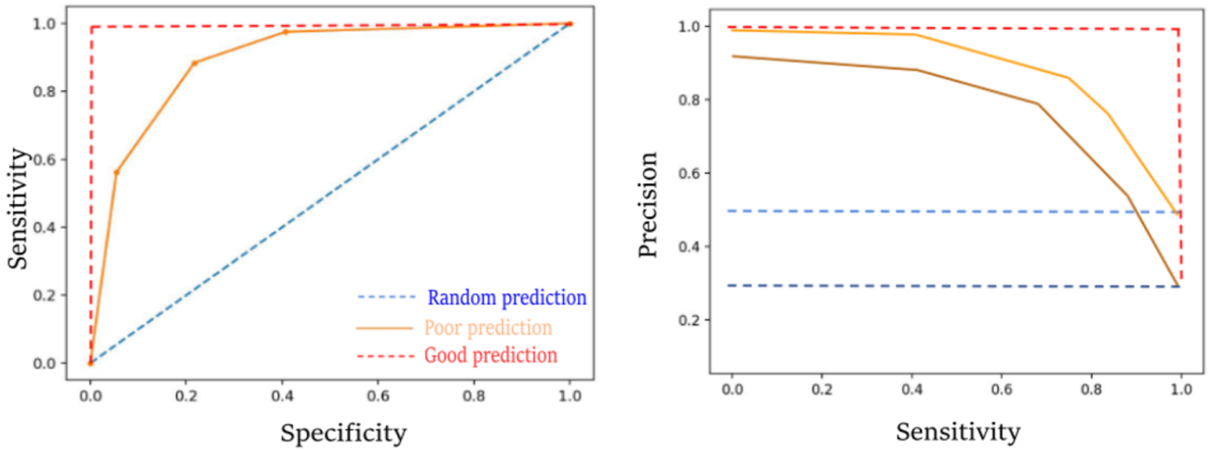
Figure 2.10.b illustrates the Detection error tradeoff (DET) curve witch present the relationship between the FRR and the FAR. It is obtained by varying the decision threshold and each time calculating the two FRR and the FAR values [16].

Figure 2.11 represent the utilization of the sensitivity/specificity and precision/sensitivity pairs. Figure 2.11.a shown the Receiver Operating Characteristic (ROC) curve, which is a popular measure for evaluating classifier performance [16]. The ROC curve is a model-wide evaluation measure that is based on two basic evaluation measures sensitivity/specificity. Similarly, Precision-Recall (PR) curve [16] shows what happens to



(a) (b)

Figure 2.10: Example of FRR/FAR Illustration; (a) Match Score Distributions,(b) Example of DET Curve.



(a) (b)

Figure 2.11: Example of ROC and PR curves illustration; (a) ROC curve, (b) PR curve.

precision and recall as we vary the decision threshold (see Figure 2.11.b).

A metric known as the Area Under the Curve (AUC) score is obtained from the ROC or PR curves. As the name suggests, it quantifies the area under the curve in the ROC or PR space. The AUC score can be calculated using the trapezoidal rule, which involves summing the areas of the trapezoids under the curve.

2.8 Conclusion

In conclusion, this chapter provides a comprehensive overview of biometric systems and their performance evaluation. Biometric systems offer automated person recognition through physiological and behavioral traits. Their accuracy is quantified through metrics like false acceptance rate, false rejection rate, and equal error rate.

Key points covered in this chapter include:

- The fundamentals of biometric system operation including enrollment, matching, and decision modules.

- The two primary functionalities of biometric systems - identification and verification - along with their objectives and differences.

- An elaboration of various biometric traits, their characteristics, strengths, limitations and suitable applications. Factors like uniqueness, collectability and spoof resistance guide trait selection.

- A comparison between unimodal systems relying on a single biometric modality versus multimodal systems fusing multiple traits.

- Techniques for biometric performance evaluation using confusion matrices, FAR, FRR, ROC, DET and other metrics. These quantify accuracy, error rates and tradeoffs.

- The diverse applications of biometrics in government, commercial and forensic domains for surveillance, access control and investigations.

In summary, this chapter provides a foundation for understanding biometric systems, their operation, performance evaluation and applications. With ongoing advances in sensing, algorithms and security, biometrics continue to gain prominence as ubiquitous and reliable solutions for person recognition across critical domains.

In the forthcoming chapter, we will delve into the intricacies of finger vein identification, tracing its historical development, exploring various techniques for feature extraction, and emphasizing the significance of databases. This chapter will lay the groundwork for comprehending the anatomical and operational aspects of finger vein identification, thereby setting the stage for further exploration.

CHAPTER 3

Finger Vein-Based Biometrics

3.1 Introduction

In the pursuit of achieving heightened levels of security, biometric identification methods have emerged as a fundamental component of authentication and verification systems. Among these methods, finger vein identification stands out as a promising and sophisticated technology that capitalizes on the intricate and unique patterns of blood vessels beneath the skin's surface in the human finger. This chapter delves into the realm of finger vein identification, a subject matter that has garnered substantial attention in the field of biometric recognition. We embark on a comprehensive journey to explore the intricacies of finger vein identification, its historical development, various techniques employed in feature extraction, matching, and the pivotal role of databases in the realization of this cutting-edge technology. The chapter begins with a chronological review of related studies in finger vein identification, presenting a timeline of its evolution and the methodologies adopted over time. We explore the properties and advantages that render finger vein identification an intriguing research area, emphasizing its resistance to spoofing, superior accuracy, and non-invasiveness. Through a visual exploration, we provide insights into the underlying anatomy and structure of finger veins, highlighting the significance of Near Infra-Red (NIR) light in the imaging process. Subsequently, we unravel the general model of finger vein identification, delineating the two primary phases: registration and identification. In the registration phase, individual finger vein images undergo preprocessing and feature extraction to create templates stored in the database. In the identification phase, a user's input image is preprocessed and compared against the stored templates to confirm or verify their identity. The chapter delves into the crucial aspect of finger vein image acquisition, discussing various methods, including light reflection, light transmission, side lighting, and bottom light transmission. The advantages and disadvantages of these methods are meticulously examined, with a focus on factors that impact image quality. Image preprocessing, a critical stage in finger vein identification, is explored in detail. Preprocessing involves techniques such as image restoration, segmentation of the region of interest (ROI), and image enhancement. We evaluate the significance of image quality and its role in successful vein feature extraction. Feature extraction, the process of transforming raw images into distinctive feature sets, is a pivotal element of finger vein identification. This chapter elucidates the various feature extraction methods, categorizing them into vein pattern-based, dimensionality reduction-based, local binary pattern-based, and texture-based methods. Each approach is scrutinized for its suitability in capturing the rich details of finger vein patterns. Matching, the decision-making phase, is the focal point of the chapter, as it determines the authenticity of the input image. We differentiate between classifier-based and distance-based matching methods, showcasing their applications and considerations. The chapter elaborates on classifiers, distance metrics, and their suitability in the context of finger vein identification. Lastly, we explore the significance of databases in the realm of finger vein identification. The chapter provides an overview of publicly available databases, discussing their size, diversity, and creation methods. We highlight the importance of constructing representative databases to facilitate robust and reliable identification systems. As we embark on this journey through the intricacies of finger vein identification and databases, we aim to provide a comprehensive understanding of this innovative biometric technology and its critical role in enhancing security and authentication systems.

3.2 Finger Vein Identification

Finger vein identification possesses distinctive attributes, rendering it an intriguing research domain in the biometrics recognition field, employing pattern identification techniques for individual identification and identity verification. In their study [17], the authors harnessed Manifold Learning for finger vein identification. This approach boasts a commendable recognition rate, primarily attributed to the reduced feature dimensions, which transform the images from higher to smaller dimensions. However, this method does exhibit some limitations, as global features are notably sensitive to factors such as location, occlusion, distortion, and lighting, making it unsuitable for extracting vein-based finger features.

Notably, neural networks and Support Vector Machines (SVM) have been employed for finger vein identification in studies [18, 19]. In the work by [18], the authors utilized an Adaptive Neuro-Fuzzy Inference System (ANFIS), while [19] leveraged Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and SVM, along with an Adaptive Neuro Fuzzy Inference System (ANFIS). It is imperative to mention that these methods function effectively in environments with controlled background noise, but they may encounter challenges with images affected by poor lighting, varying observation angles, and other parameters. In [20], conventional recognition techniques, including multi-instance and Local Binary Pattern (LBP), were applied to finger vein identification. It is worth noting that in 2012, Harsha and Subashini [21] presented a novel finger-vein recognition system for authentication in automated teller machines.

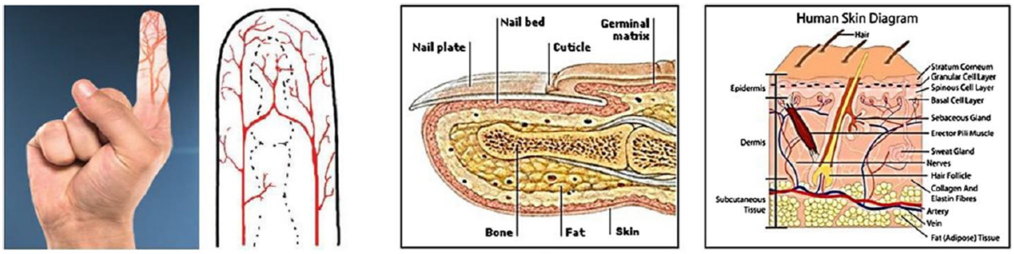
Khellat-Kihel & al. [22] adopted a comprehensive approach, utilizing information capacity, spatial gradient, image entropy, contrast, and Gabor features, coupled with Support Vector Regression (SVR) for finger vein recognition, a machine learning technique. However, this approach focuses on integrating and creating Regions of Interest

(ROI) within the venous system, showcasing a multimodal nature common to many biometric systems.

In [23], a Convolutional Neural Network (CNN) was presented for finger-vein-based biometric identification. This model incorporates five convolutional layers, three max-pooling layers, one SoftMax layer, and one ReLu layer, along with contrast-limited adaptive histogram equalization, forming the Convolutional Neural Network (CNN) Model (CLAH). Nevertheless, this approach is limited to photos of trained classes' finger veins. Zhao & al. [24] implemented a CNN model (FCL) for finger vein recognition, consisting of three convolutional layers, three max-pooling layers, and two fully connected layers. It's crucial to note that the suggested system requires further robustness enhancement to improve performance accuracy. Moreover, providing more detailed information about the model and the loss functions employed in trials would enable a more comprehensive comparison of their performance and the assessment of the advantages and disadvantages of each loss function. Lastly, Rosdi & al. [25] conducted an analysis of principal components and introduced an Adaptive K-Nearest Centroid Neighbor Classifier for finger vein recognition. As an improvement, they introduced an Adaptive Centroid Closer Neighbor (akNCN). In two experiments, akNCN.v1 and akNCN.v2, the accuracy was 85.64, with no improvement in accuracy, but a substantial time difference of 5153 seconds for v2 compared to 6321 for v1. This proposed classifier demonstrates slightly lower classification accuracy compared to the original kNCN, and it appears to discard a significant amount of information. Consequently, this method reduces the size of the training data and eliminates templates.

3.2.1 Finger Vein Features

The term "finger vein" refers to the intricate patterns of blood vessels and capillaries located beneath the skin of the finger. Research has unveiled that these vein patterns are remarkably unique, even among identical twins. In every finger, there exist tissues and organs capable of absorbing Near Infra-Red light (NIR) at various absorptivity levels. Veins carrying deoxygenated blood possess the ability to absorb NIR light, resulting



(a) (b)

Figure 3.1: Cross-section anatomy of finger vein under near Infra-red light [26]; (a) Human finger (b) Finger cross-section under human skin

in their darker appearance when compared to the surrounding tissue in vein maps [26]. Consequently, finger vein images are captured under NIR light to highlight these patterns. Finger vein patterns offer distinct advantages over other biometric traits. Figure 3.1 illustrates the finger vein when exposed to Near Infra-Red light, showcasing its various components.

One of the standout features of finger vein biometrics is its resistance to spoofing [27]. For instance, fingerprint systems can be relatively easily spoofed when a user molds a fake finger image using easily moldable materials, such as wax or dental impression material, to create a replica of a fingerprint impression. In the case of face recognition, there is the possibility of "copy attacks," where a face recognition system can be deceived through the acquisition and use of social media images or photos stolen from social networking websites by an attacker [28]. Iris spoofing attacks, on the other hand, can involve photo attacks, contact-lens attacks, or artificial-eye attacks.

In addition to its resistance to spoofing, finger vein-based identification offers several key advantages compared to other biometric recognition technologies. It is known for its accuracy, characterized by a low False Rejection Rate (FRR) and a low False Acceptance Rate (FAR). Furthermore, it is less invasive as it does not require the subject to make direct contact with the scanning surface of the machine. This absence of physical contact eliminates any hygiene-related concerns associated with finger vein scanning. Another noteworthy advantage is that it doesn't involve the subject leaving latent prints behind on the scanning device. Moreover, finger vein identification is not affected by weather conditions, be it wet or dry, since it operates at a sub-dermal level. Additionally,

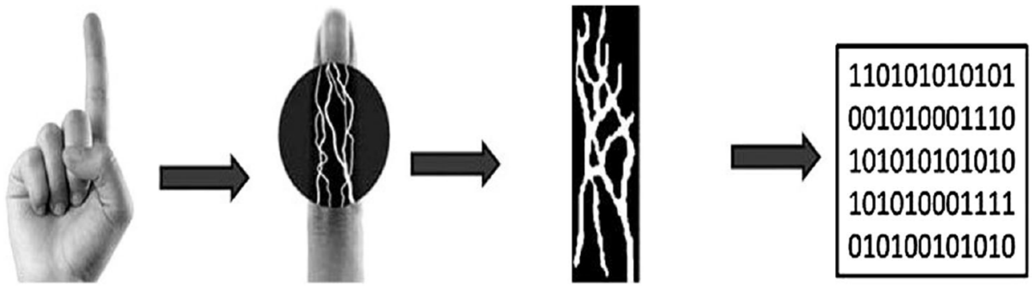


Figure 3.2: Structure of finger vein pattern

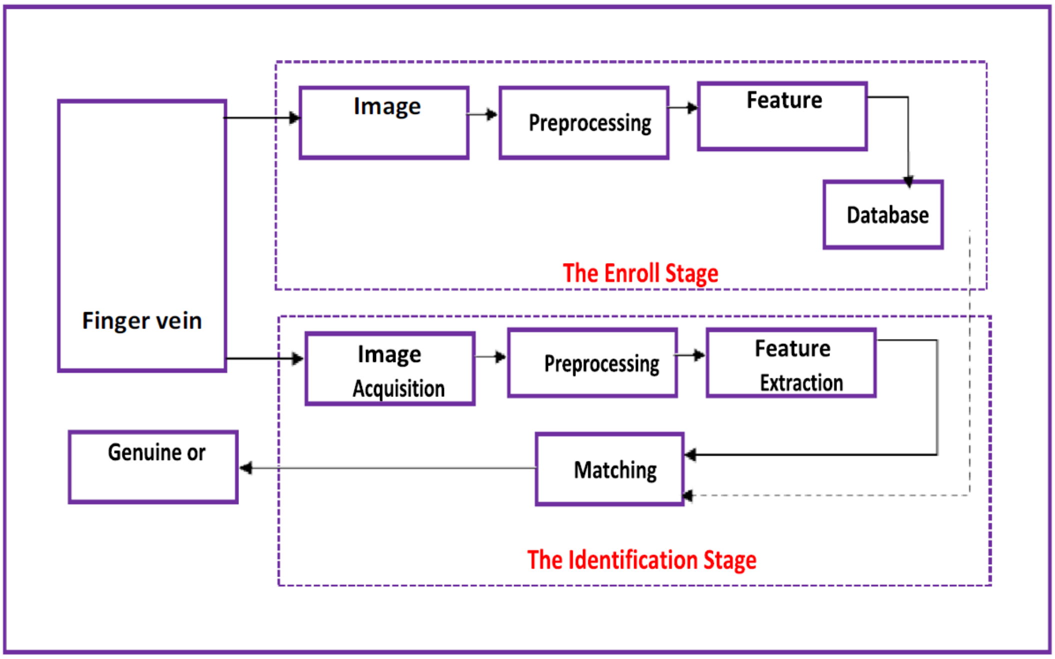


Figure 3.3: The model of finger vein identification

it remains consistent with age, allowing for the use of the enrolled data throughout the subject's lifetime [29]. Figure 3.2 illustrates the vein pattern of the finger, showcasing the distinct features of this biometric trait.

3.2.2 General Model of Finger Vein Identification

A comprehensive finger vein identification system comprises two primary phases, as illustrated in Figure 3.3. The initial phase is registration, also commonly referred to as enrollment, while the subsequent phase is identification, often termed as matching.

During the registration phase, individual finger vein images are captured, and they subsequently undergo preprocessing and feature extraction stages. These processed images are then saved as templates within a database. The registration phase serves the purpose of creating a reference database of the user's unique finger vein patterns. In the identification phase, a user's finger vein image is captured and subjected to preprocessing procedures. This preprocessing step plays a pivotal role in identifying the region of interest within the image that will be used in subsequent procedures. It encompasses tasks such as image alignment and enhancement. Following preprocessing, the distinguishing features extracted from the image are compared against the templates stored in the database. This comparison process can serve to identify the user in an identification mode or to verify the user's identity in the verification mode [30]. It is important to note that the acquisition of finger vein images typically employs a Charged Coupled Device (CCD) camera as the imaging device.

Subsequent sections will delve into the specific procedures and methods employed in each of these phases, highlighting the advantages and disadvantages associated with each method. We will commence with an exploration of the operation of finger vein image acquisition.

3.3 Finger Vein Image Acquisition

Initially, Hashimoto [31] introduced three distinct approaches to acquire finger vein pattern images: light reflection, light transmission, and side lighting. All of these methods utilize infrared (IR) light, but they differ in terms of how the finger and IR light interact. In 2012, an additional approach, bottom light transmission, was introduced to complement the existing methods for finger vein acquisition. Below, we provide detailed descriptions of each of these methods:

3.3.1 Light Reflection Method

The light reflection method captures vein patterns using infrared (IR) light. In this approach, the IR source is placed sideways to the CCD image sensor, and the finger is positioned in front of the sensor, as depicted in Figure 3.4. Typically, this method is not

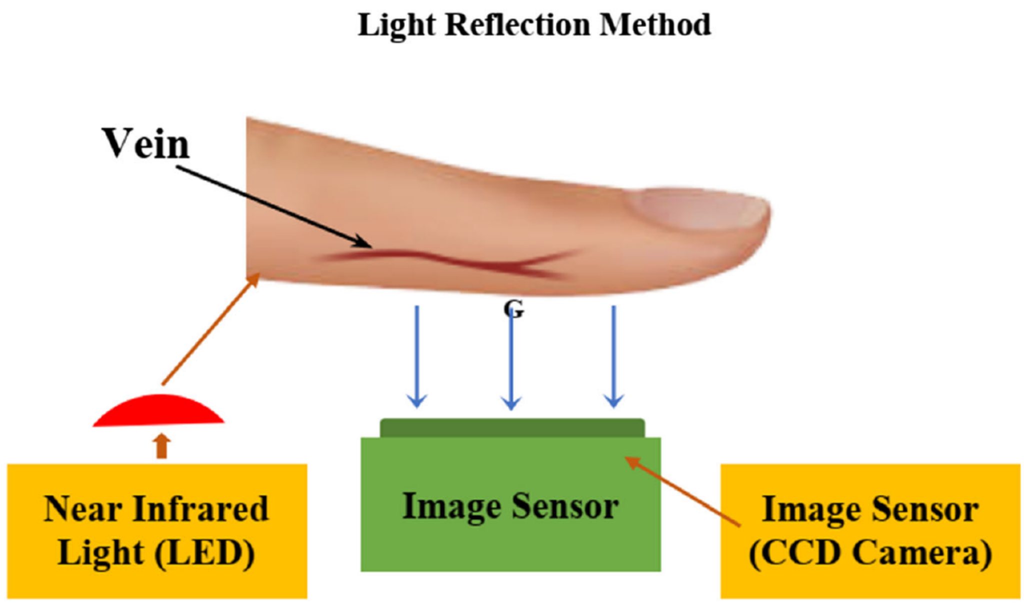


Figure 3.4: Light reflection method

employed for capturing vein patterns in areas such as the palm, palm-dorsa, or wrist, where the tissue thickness is greater, and IR penetration is limited. Light reflection is particularly advantageous in terms of its compact form factor, making it suitable for use in small devices, which is essential for the final product. However, it does suffer from low contrast as the IR light can only penetrate up to a depth of approximately 1 mm in the skin. Therefore, advanced image processing methods are required to enhance image quality, especially for the fine, smaller veins [31].

3.3.2 Light Transmission Method

In the light transmission method, the IR light source is positioned opposite to the CCD sensor, with the finger placed in between, as illustrated in Figure 3.5. This method relies on capturing the IR light transmitted through the finger. While it may seem more reliable for capturing vein patterns, not all parts of the finger can be used, as only areas with the appropriate thickness allow for effective IR light transmission [31].

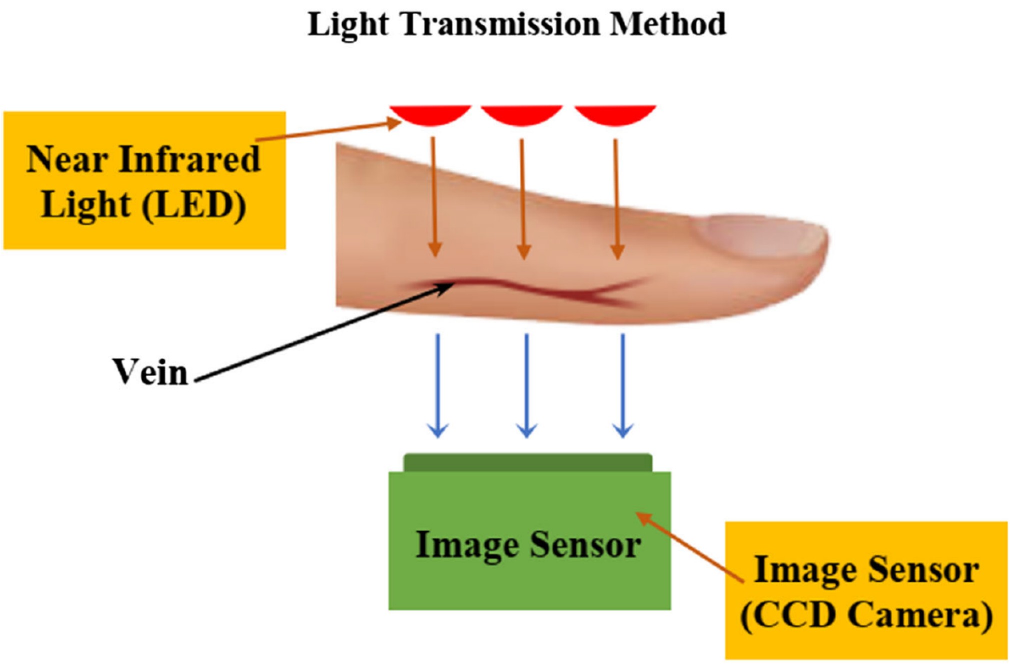


Figure 3.5: Light transmission method

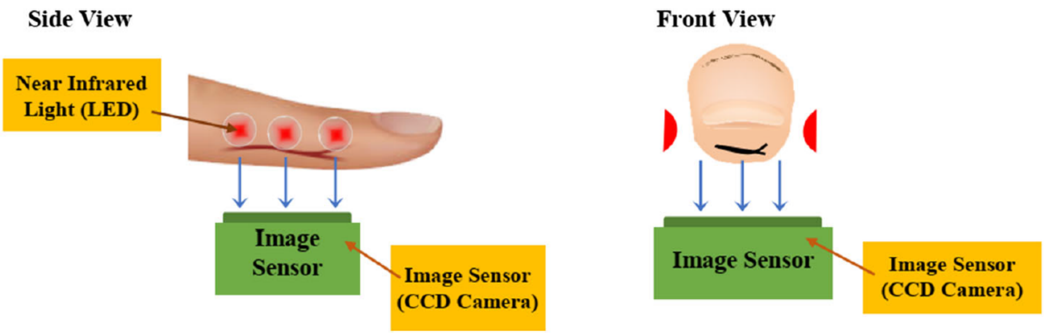


Figure 3.6: Side lighting method

3.3.3 Side Lighting

The side lighting method was introduced by Hashimoto [31]. In this approach, IR light sources are positioned on both sides of the finger, as shown in Figure 3.6. The idea is that the IR light will penetrate through the finger, scatter, and be detected by the sensor to capture the vein image. This method has been proven to provide better and sharper image contrast compared to other methods. The resulting end devices may be larger than those based on light reflection but smaller than those using light transmission.

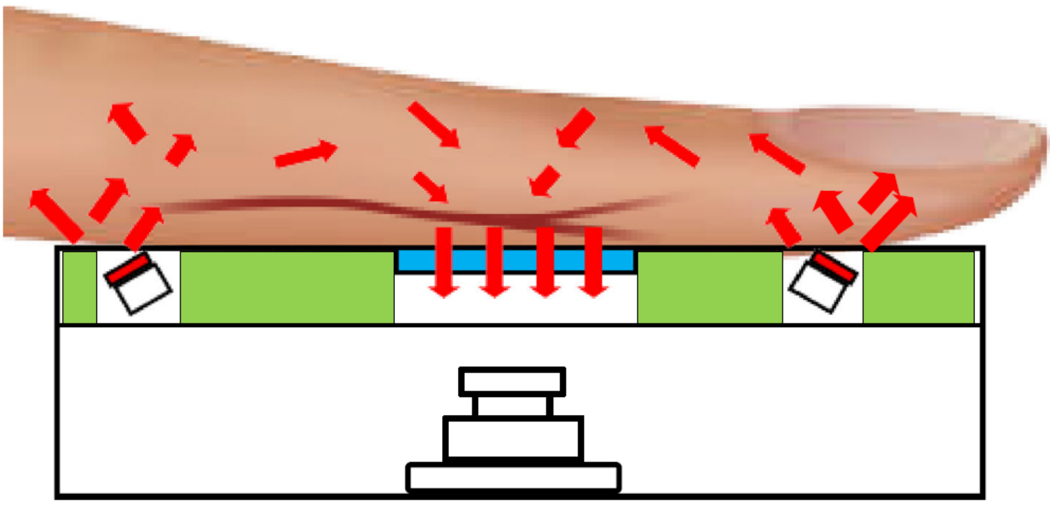


Figure 3.7: Bottom light transmission

3.3.4 Bottom Light Transmission

The fourth approach, bottom light transmission, was reviewed by Vallah [32]. This method seeks to address mobility limitations by positioning the camera and the sensor (IR-LED light) at the base of the device. The approach is similar to the light reflection method, but it necessitates the finger to make contact with the LED screen. Once the sensor detects the finger, the IR light is projected and propagates within the finger, resulting in vein pattern capture similar to that of the side lighting approach, as illustrated in Figure 3.7. A comparative analysis of the advantages and disadvantages of each of these acquisition methods is presented in Table 3.1 for reference.

The quality of captured finger vein images can be affected by various factors, including the thickness of subcutaneous fat, skin color, finger positioning, image background, and the efficiency of the image capture device [33]. However, there is currently no standardized measure for image quality control, resulting in the presence of low-quality images. These low-quality images can be categorized into four distinct forms, as outlined in Table 3.2.

Low-quality finger vein images can lead to challenges in identification, potentially causing delays in pre-processing and feature extraction stages. The FV reader, designed for capturing finger vein patterns as biometric features, utilizes at least one optical imaging unit and a digital signal processor. This scanner is displayed in Figure3.8.

Methods	**Advantages**	**Disadvantages**
Light Reflection	• Sharp contrast after optimizing • Low cost • Low power	High quality requirement for NIR sources and components
Light Transmission	Sharp contrast	Inhomogeneity in images with contrast differences between regions
Side Lighting (Hybrid)	Higher definition and contrast than images obtained by reflection	Complicated and High cost
Bottom Light Transmission	Higher definition like side lighting Can be made to mobile Low cost	

Table 3.1: Comparison of finger vein capturing methods for image acquisition

Problem Category	**Description**
Blurry image	The vein patterns that contain little contrast
Askew image	The vein images with a definite grade of deformation
Dim image	The captured images with a dim or black portion
Bright image	The existence of a sunny portion in the images

Table 3.2: Categories of low-quality finger vein images

Figure 3.8: Typical finger vein readers

3.4 Image Preprocessing

Preprocessing in image processing involves necessary actions or preparations performed before the main data analysis and information extraction. It aims to rectify deficiencies such as low contrast and noise in the image, and in the case of finger vein image processing, includes image enhancement steps. The preprocessing steps consist of image restoration, region of interest (ROI) cropping, and image enhancement. Various algorithms are employed for producing and aligning the ROI, with one commonly used method being the Lee-Region detection [34]. The quality of an image is a critical aspect in image processing, viewed from three perspectives: quality control systems, benchmarking image processing systems and algorithms, and optimization of algorithms and parameter settings within image processing systems [35]. The importance of maintaining high-quality finger vein images becomes evident, as the performance of a finger vein image largely depends on its quality [36].

3.4.1 Image Restoration

Image restoration aims to eliminate or reduce known degradations within an image. This includes correcting distortions due to reader machine limitations or background noise through noise filtering and improving geometric distortions or non-linearity introduced by the sensor. However, before image enhancement, it is essential to crop the finger vein image to remove unwanted areas, as illustrated in Figure 3.9.

3.4.2 Image Segmentation of ROI

Segmenting the Region of Interest (ROI) plays a crucial role in preprocessing finger vein identification systems. In the context of finger veins, ROI represents the area containing the network of vein patterns within the finger. Extracting the ROI is crucial to determine the image portion suitable for vein feature extraction while removing non-useful information from the image. Correctly extracting the ROI can significantly reduce computational complexity, thus enhancing the efficiency of the finger vein recognition system.

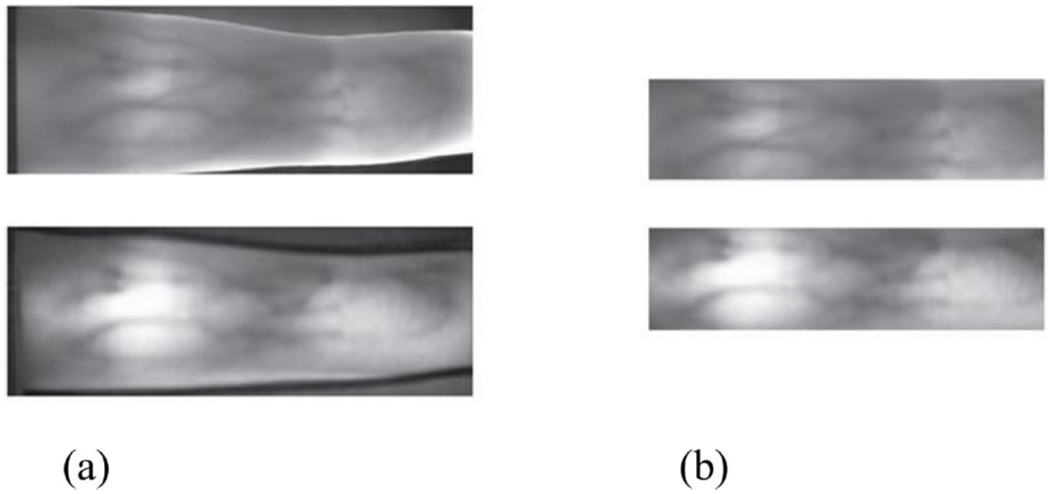


Figure 3.9: Finger Vein Image Cropping: (a). Image before Cropping (b). Image after Cropping

Thus, ROI extraction is a critical operation in finger vein identification systems.

Few algorithms exist for the extraction of ROI vein in the finger. Rosdi and his coresearchers [37] made use of the fixed-size window base to crop out a certain portion of the finger in the finger vein image. The method is sensitive to displacement of the finger and is not accessible to be used by askew finger vein images. In their own cases, Yang and Shi [38] offered an ROI localization method that was based on the physiological structure of human fingers. Though, the issue of displacement of the finger can be resolved, the method is not accessible for askew finger images. Hence, before using the available method of ROI extraction, it is very necessary that askew finger vein images must be corrected at the first stage. ROI extracted through an edge detector was done by Kumar and Zhou [39] after he performed the rotational alignment. However, the method refused to harvest the ROI area from finger vein images, causing the operation of an image to have more background than the vein portion. Figure 3.10 show The measurement of the ROI in [40].

3.4.3 Image Enhancement

Image enhancement is a crucial aspect of image processing, as it improves the visibility of specific parts of an image for further analysis by operators or systems [41]. Various

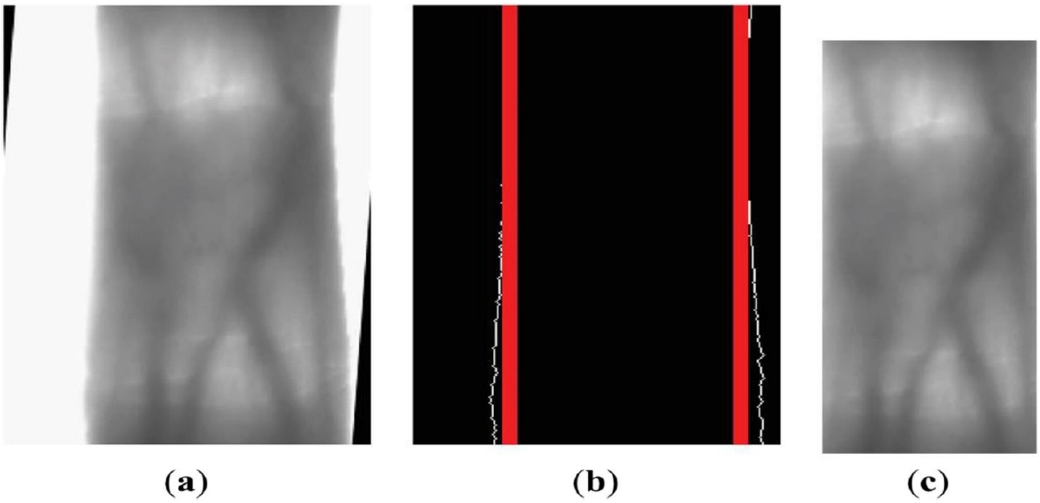


Figure 3.10: The measurement of the ROI [40]; (a) ROI in actual stature. (b) Finger image edge using internal lines. (c) ROI of a finger vein image

digital image enhancement methods are available, and the choice of method depends on the imaging modality, task at hand, and environmental conditions. Image enhancement often involves improving image contrast and reducing noise, leading to better image quality. Proper enhancement significantly contributes to improving the value of degraded images. Many approaches to image enhancement have been proposed. For instance, Arun and colleagues [42] suggested that Adaptive Histogram Equalization provides better results for image enhancement, although it may still leave some blurriness in the images. Agaian & al. [43] proposed global histogram equalization, which, while common, has limitations, as it treats the entire image globally and may result in over-enhancement. Furthermore, the presence of noise in images is common and can lead to degradation of image quality. Noise removal, or denoising, is a crucial part of image enhancement. Median filters and Wiener filters are effective in removing Salt-and-pepper and Gaussian noise, respectively. The application of Median filters followed by Wiener filters has become a standard procedure for enhancing vein images.

Figure 3.10 shows the measurement of the ROI in a study [40]. However, many existing enhancement methods in the field of finger vein identification have made progress but still require further improvement, as the effectiveness of vein feature extraction is highly dependent on the quality of the enhanced vein images.

Method	Descriptions	Ref
Repeated Line Tracking	The vein in the image is traced to randomly select seed (directions chosen from predefined probability). The process is repeatedly done until	[44]
Maximum Curvature	Image extraction by detecting its center line	[45]
Gabor	A linear filter used for edge detection by transforming the image into the frequency domain	[39]
Mean Curvature	Image segmentation using the mean of the surface curvatures in all directions. It can quantify the degree of likeness to a ridge or valley	[46]
Region Growth	This is running the region growing operator on the different seeds with emphasizes continuity and symmetry of valleys in the cross-sectional profile	[47]
Modified Repeated Line Tracking	Find the image locus based on the revised parameters	[48]

Table 3.3: Descriptions of some typical vein pattern-based feature extraction methods

3.5 Feature Extraction

3.5.1 Vein Pattern-Based Methods

Vein pattern-based methods focus on vein pattern segmentation and utilize geometrical shape or topological structure of vein patterns for matching. Various techniques such as Repeated Line Tracking, Maximum Curvature, Gabor, Mean Curvature, Region Growth, and Modified Repeated Line Tracking are employed in this category. Table 3.3 provides an overview of the common methods used in vein pattern-based feature extraction.

3.5.2 Dimensionality Reduction-Based Methods

Dimensionality reduction-based methods employ techniques like Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), (2D)2PCA, and manifold learning. These methods extract either global or local features from finger vein images. While global features are obtained through methods like PCA and LDA, local line features are extracted using Local Projection Pattern (LPP). Balancing local and global features is crucial for accurate recognition. However, large-scale applications might find dimensionality reduction methods challenging due to complexities related to transformation matrix learning [47].

3.5.3 Local Binary Pattern-Based Methods

Local binary pattern-based methods focus on local areas and extract features in binary form. Methods like Local Binary Pattern (LBP), Local Line Binary Pattern (LLBP), Personalized Best Bit Maps (PBBM), Personalized Weight Maps (PWM), and Local Directional Code (LDC) are part of this category. These methods derive binary codes by comparing the gray levels of pixels and their neighbors. Hamming distance (HD) is commonly used to measure similarity between enrolled and input binary vein features.

3.5.4 Texture-Shape Descriptor Methods

Shape descriptors are grouped into contour-based and region-based methods 3.11. This grouping considers whether shape features are removed from the contour or from the entire shape section. Shape descriptors are additionally grouped into structural (local) and global descriptors. If the shape is characterized by bits or regions, it is structural and if the shape is characterized by the whole region, it is global. Another grouping arranges the shape description into spatial and transforms domain methods, which is dependent on the use of coordinate estimations or applying a transformation of the shape. Figure 16 shows the shape representation and description methods.

Texture-shape descriptors include contour-based and region-based methods. Contour-based methods extract boundary information and are sensitive to noise, while region-based methods consider all pixels within the shape, making them more robust in general applications. Contour-based descriptors include Fourier descriptor [50], wavelet descriptors [49], and curvature scale space (CSS). Region-based descriptors include moment invariants and Zernike moments [51].

The distinctive line patterns of finger veins have encouraged researchers to treat them as texture images. Texture features are often extracted using methods like wavelet transform and Gabor filter. Studies have combined local and global features for en-

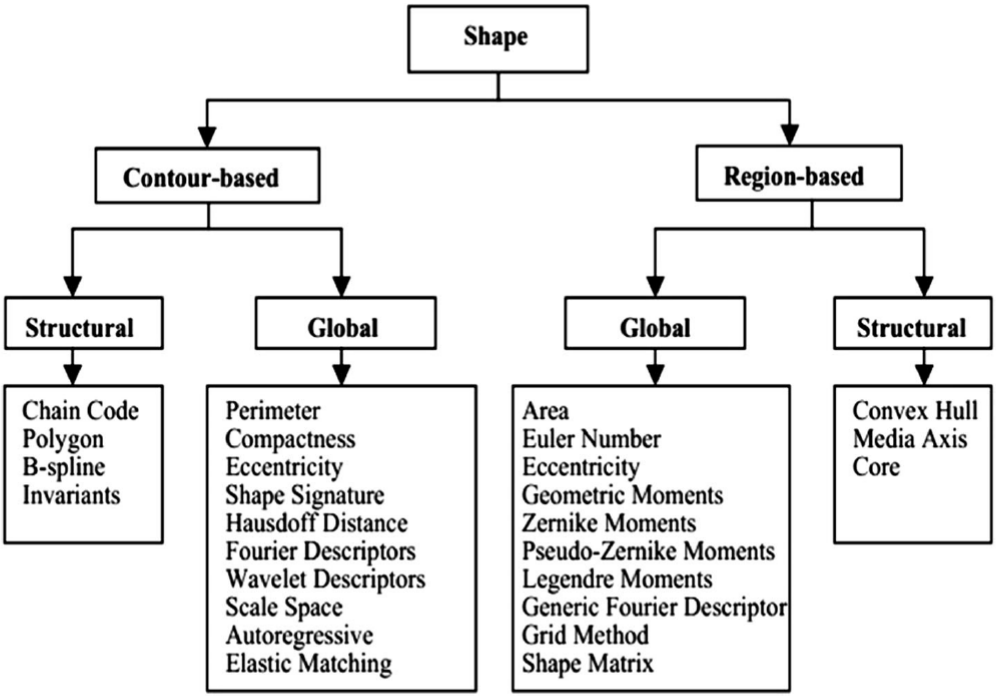


Figure 3.11: Shape representation and description methods [49]

hanced recognition. For example, Park combined Local Binary Pattern (LBP) and Wavelet transform, achieving a recognition rate of 98.9% on a custom database [52]. Gayathri and Ramamoorthy fused correlation, energy, and homogeneity features, achieving a recognition rate of98.4% [53].

3.6 Matching

The decision-making stage in the finger vein identification process is the matching stage. In this stage, the features extracted from a pattern are comparable to those of the enrolment set. This decides if the entry image is original or fake for registered image to produce a matching score (the similarity between the registered template and the entry image). There are twofold categories of matching methods, namely; classifier-based matching and distance-based matching [54]. The distance-based matching method is exploited by conventional finger vein identification approach, and while classifier-based matching method is use for machine learning finger vein identification. Thus, classifier-based matching will try to categorise the pattern that will lead to the generation of hypotheses, and not as a unique solution [55]. Generally, classification is achieved based on features such as minutiae [56], local line binary pattern [37], SIFT [37], soft biometrics [47], statistical measures, machine learning [17], correlation (or template) based methods [39], and hybrid algorithms [52]. The uses of minutiae feature for classification commonly indicates finger vein images low-quality performance, which consists of several fake and limited genuine minutiae. Likewise, fewer amount of accurate and typical SIFT key-points can damage the enactment of classification. Also, because of the pose variation of the finger, using such width of the phalangeal joint soft biometric trait [47] or finger geometry [57] is not productive for classification. In addition, the statistical measures of feature extraction such as local moments for classification, is unproductive due to the less discrimination of statistical features. Classification via machine learning methods needs a massive quantity of training data that can reflect some of the likely distortions, but it is always impossible [17]. Even if the genuine veins are lost or the fake veins are presented, the use of correlation or template based matching can give an accurate result. It removes strong distortion within image registration; hence, it can be described as classification based on strong registration. The similarity score is computed by using registered images. Thus, features for registration are such as vein structure [39]. However, finger poses [40] can decline the correlation or template-based classification performance. K-nearest neighbor (kNN) classifier is one of the most well-known supervised learning algorithms in pattern classification, which employed by some researchers. Many researchers claimed that using K-nearest neighbor (kNN) classifier has several benefits such as intuitiveness, effectiveness, competitive, and simplicity performance of classification in several domains [58]. However, KNN works with a distance-based metric for the evaluation of the comparison level between the feature vector of the input pattern and the tested template(s). Gongping Yang & al. [38] proposed feature extraction of finger veins using a 2D PCA method and KNN classifier for classification of everyone. Furthermore, they adopted to solve the class-imbalance problem using the SMOTE technology. A custom database of 80 individuals' index fingers of the right hand from 18 finger vein images was used.

3.7 Review of Existing Databases

In finger vein identification methods, the preparation of a suitable database is a crucial step before feature extraction and matching. The size and diversity of the database can have a significant impact on the performance and generalizability of the identification system. A well-constructed database should include samples from a range of participants, including both males and females, diverse ethnic groups, adults, and children spanning various age groups. It is typically recommended that a minimum of two samples of vein images be captured for each individual participant, with one sample used for enrolment and the second for algorithm evaluation [59]. Table 3.4 provides an overview of publicly available finger vein databases, including the size of the images and the year in which each database was created. These databases have been used in both research and industry for finger vein identification. It's worth noting that most of the databases in the table were created using the transmission method, except for the CFVD database, which was obtained using the reflection method [60]. The human hand consists of five fingers: thumb, index, middle, ring, and pinky. This totals ten fingers when considering both right and left hands. However, some databases only include the three fingers on each hand that are more suitable for capturing vein images. The thumb and pinky fingers are typically excluded because they are thicker and shorter than the other three fingers [61]. Additionally, near-infrared light may have difficulty penetrating the thick skin of the thumb finger, making it challenging to capture vein patterns. The instability of the capturing device's structure can also affect the ability to capture high-quality images of the thumb. Therefore, the thumb and pinky fingers are often omitted from the database. In the table, the fingers are represented as follows: - Right finger index, middle, and ring: Ri, Rm, Rr - Left finger index, middle, and ring: Li, Lm, Lr Creating a diverse and representative database is essential to ensure the effectiveness and reliability of finger vein identification systems.

Database	Public Available	Subjects	Fingers Per Subject	Session	Acq. Per Session	Images	Image Size	Device	Year
Hitachi Res. Lab. [62]	No	2673	Li,m Ri,m	1	11	117, 612	Unknown	TS-EE3F1	2004
Hitachi-Kyushu [63]	No	506	Ri	1	2	1012	Unknown	TS-EE3F1	2007
PKU v.2,3,4 [64]	Yes	5208	Li,l Ri,l	1	5	50,700	512x384	Proto PKU	2008
GUC45 [65]	No	45	10	12	2	10,800	512x240	Proto GUC	2009
SDUMLA-HMT [66]	Yes	106	Li,m,r,Ri,m,r	1	6	3816	320x240	Proto Wuhan Univ	2011
HKPU [39]	Yes	156	Li,m Ri,m	2	6	6264	513x256	Proto HKPU	2011
UTFVP [67]	Yes	60	Li,m,r,Ri,m,r	1	4	1440	672x380	Proto Twente Univ	2013
MMCBNU6000 [68]	Yes	100	Li,m,r,Ri,m,r	1	10	6000	640x480	Proto Chonobuk Univ	2013
CFVD [69]	Yes	13	Li,m,r,Ri,m,r	2	51	1345	640x480	Proto Shandong Univ	2013
Shandong Univ [70]	No	34	Li,mRi,m	2	20-10	4080	320x240	Proto Whuan Univ	2013
FV-USM [71]	Yes	123	Li,m,r,Ri,m,r	2	6	5904	640x480	Proto Sains Uni	-

Table 3.4: Review of Existing Finger Vein Databases

Keys:
R – Right Hand, L – Left Hand, i – Index finger, m – middle finger, r – ring finger, l – little finger.
Proto – Laboratory-Made Prototype.

3.8 Conclusion

The chapter has embarked on a comprehensive exploration of the captivating world of finger vein identification and the pivotal role of databases in this cutting-edge technology. With its remarkable resistance to spoofing, superior accuracy, and non-invasive nature, finger vein identification has emerged as a promising and sophisticated biometric recognition method. Our journey began with a historical perspective, tracing the evolution of finger vein identification and the methodologies adopted over time. From its nascent stages to its contemporary applications, finger vein identification has evolved into a sophisticated and reliable technology with immense potential. The anatomical intricacies of the human finger vein system were laid bare, emphasizing the significance of Near Infra-Red (NIR) light in capturing these unique biometric patterns. This natural authentication method, rooted in the distinctive vein patterns of each individual, showcases remarkable advantages and has become the subject of intense research and development. The general model of finger vein identification, divided into registration and identification phases, provided a structural framework for the technology's practical application. In the registration phase, individuals' vein images undergo preprocessing and feature extraction, creating templates that find their place in databases. The identification phase involves the matching of an input image with these templates, either confirming or verifying the user's identity. Image acquisition methods, including light reflection, light transmission, side lighting, and bottom light transmission, were scrutinized, emphasizing their advantages and limitations. The importance of image quality in the acquisition process became evident, setting the stage for subsequent stages of preprocessing and feature extraction. Preprocessing emerged as a critical step in enhancing image quality and paving the way for effective feature extraction. The various aspects of preprocessing, such as image restoration, region of interest (ROI) segmentation, and image enhancement, were dissected to underscore their role in capturing the rich details of finger vein patterns. Feature extraction methods, classified into vein pattern-based, dimensionality reduction-based, local binary pattern-based, and texture-based categories, were explored in depth. The chapter shed light on the diversity of techniques used to transform raw images into distinctive feature sets, setting the stage for accurate and reliable identification. Matching, the decision-making phase, came into focus, with distinctions made between classifier-based and distance-based matching methods. The chapter underscored the significance of the various classifiers, distance metrics, and their suitability for finger vein identification. The chapter also underlined the pivotal role of databases in the realm of finger vein identification. The creation of representative databases is essential to support robust and reliable identification systems. The chapter presented a selection of publicly available databases, highlighting their size, diversity, and the methods employed in their construction. In conclusion, the world of finger vein identification is a remarkable intersection of biometric technology and database management. As it continues to evolve, it promises to offer even more secure and reliable authentication solutions. With further research, development, and the creation of comprehensive databases, finger vein identification is poised to play a crucial role in enhancing security and authentication systems across a wide array of applications. This chapter has provided a solid foundation for understanding this exciting and innovative field, setting the stage for future advancements and applications.

In Chapter 4, the attention will pivot toward developing a deep-learning model tailored for finger vein identification, utilizing the InceptionResnet-V2 architecture. This customized InceptionResnet-V2 model exhibited superior performance when compared to current state-of-the-art methods in finger vein identification. The research conducted in this chapter vividly illustrated the potential of deep learning models to significantly enhance both the security and accuracy of finger vein identification systems.

CHAPTER 4

Deep learning model based on inceptionResnet-v2 for Finger vein recognition

4.1 Introduction

The burgeoning need for enhancing security in personal identification systems, driven by the escalating threats of identity theft and cybercrime, has fueled the evolution of biometric identification technologies. Traditional methods such as fingerprints, face recognition, and palm prints, while effective, have found a formidable ally in finger vein (FV) recognition. FV technology, distinguished by its unparalleled accuracy and security, is rapidly gaining prominence in the realm of biometric identification. Unlike other biometric methods, FV patterns are intricate and challenging to counterfeit, making them a reliable choice for automated personal identification systems.

Finger vein identification operates by analyzing the unique patterns of blood vessels in an individual's fingers using near-infrared light. First pioneered by Hitachi's R&D department [4], this technology has found diverse applications in healthcare, finance, automobile security, and confidential systems like automated teller machines (ATMs).

Given the need for swift identification processes and the complexity of FV patterns, there is a demand for cost-effective single-chip designs. Moreover, the field has witnessed a paradigm shift with the advent of deep learning-based models, leveraging the vast expanse of data and computational power available today.

Several notable advancements have been made in this domain. [72] introduced a convolution neural network (CNN) model based on the VGG-16 architecture, showcasing excellent performance, especially concerning finger vein misalignment and environmental factors. [73] employed the Densenet-161 architecture, implementing orientation adjustments for vein images, ensuring robust results. [74] proposed a multi-layer neural network classifier using back propagation neural networks, coupled with Principal Component Analysis (PCA) for extracting low-level details, thereby enhancing FV identification accuracy and robustness. Furthermore, [24] devised a lightweight CNN model integrating central loss function and dynamic regularization, emphasizing computational efficiency without compromising accuracy.

In this study, we present a novel model to FV biometric identification by harnessing the power of deep learning. Specifically, we adopt the pre-trained InceptionResnet-V2 architecture, as detailed in [75], customizing it with additional embedded layers. This modification enhances the model's ability to discern intricate FV patterns across various datasets.

This chapter is structured as follows:The first section serves as an introduction, providing an overview of the chapter's content. Section II provides an in-depth exploration of the proposed InceptionResnet-V2 architecture for finger vein biometric identification. Section III delves into the description of the three datasets utilized for training our model, shedding light on the diversity and complexity of the data. Subsequently, in Section IV, we present a comprehensive analysis of the proposed model's performance, benchmarking it against the State-of-the-Art (SOTA) FV identification methods. Finally, the conclusions drawn from our study are presented in Section V, encapsulating the key findings and implications of our research.

4.2 Methodology

In this section, we present a detailed exposition of the components integral to our proposed InceptionResnet-V2 based architecture, as depicted in Figure 4.1. This architectural framework has demonstrated a high degree of accuracy in the identification of FV images. The process begins with the preprocessing of images from our datasets, incorporating operations such as augmentations and resizing. Subsequently, low-level features are extracted from these FV images through the utilization of a pre-trained InceptionResnet V2 based model. Finally, the extracted features are subjected to a Dropout layer (20% rate) to mitigate the issue of overfitting. Following this, two fully connected layers, equipped with the SOFTMAX activation function, are employed to produce the final identification outcomes.

In this study, we delve into a comprehensive exploration of the InceptionResNet architecture [75], with a specific focus on its key constituent modules: the Inception and the ResNet blocks [76]. This tandem combination yields a significant enhancement in the architectural performance, as visually depicted in Figure 4.2.

The ResNet module plays a pivotal role by introducing residual connections, which significantly facilitate the training of deep architectures and maintain accuracy even in considerably increased network depths. On the other hand, the Inception block empowers the extraction of a diverse set of features from input images characterized by varying scales. Notably, in the Inception-ResNet module, the residual scaling factor is judiciously applied to scale the Inception block. It is recommended to select a residual scaling factor value within the range of 0.1 to 0.3 to ensure network stability during training.

The hallmark of inception models lies in their multi-branch structures, constructed from an ensemble of convolution filters featuring various kernels (1×1, 3×3, 5×5, etc.). Within each branch, these filters are concatenated and meticulously combined. This split-transform-merge architecture of the Inception module imparts a robust representational capability to its dense layers. The hybrid InceptionResNet-V2 network adeptly harnesses residual connections, fostering highly effective training. To

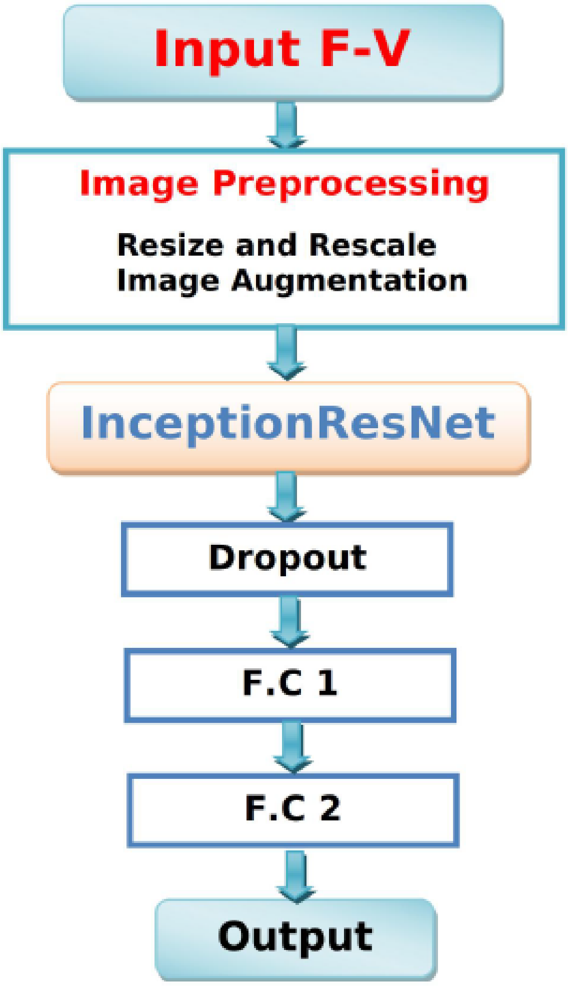
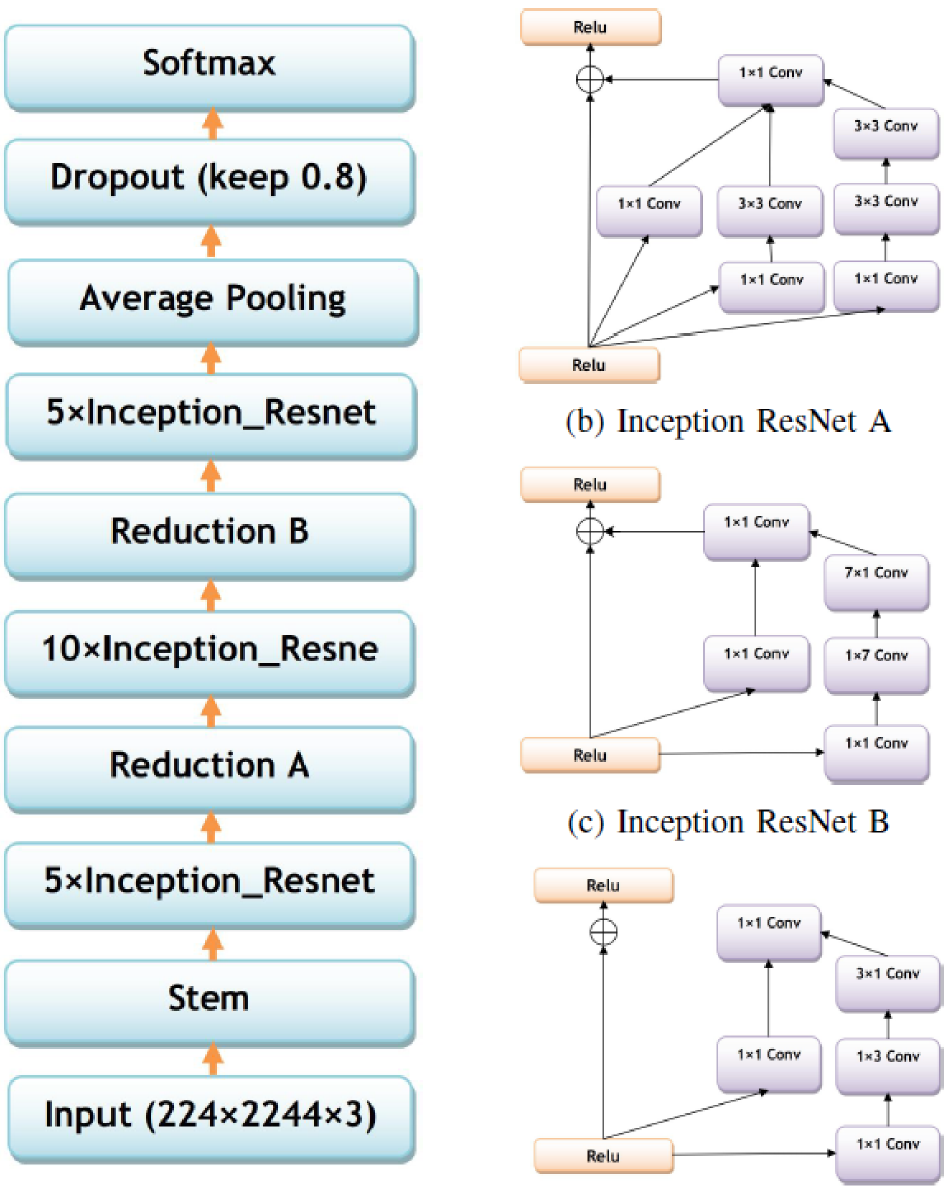


Figure 4.1: Block diagram of the proposed model



(a) Overall structure

(b) Inception ResNet A

(c) Inception ResNet B

(d) Inception ResNet C

Figure 4.2: Overall and module structure of Inception ResNet v2

further enhance architectural stability, residuals are scaled prior to their addition to the preceding layer. Figure 4.2.a visually illustrates the internal architecture of this design, comprising Inception-Resnet-A, Inception-Resnet-B, and Inception-Resnet-C blocks(see Figure 4.2.b, 4.2.c, and 4.2.d, respectively).

4.2.1 Dataset Description

In this section, we provide a detailed overview of the three distinct Finger Vein (FV) datasets used to evaluate the performance of our proposed model. These datasets have been instrumental in gauging the robustness and effectiveness of our InceptionResnet-V2-based model, enabling a comprehensive analysis of its capabilities (see Table 5.1 for descriptions of the datasets). Additionally, Figure 4.3 illustrates samples of Finger Vein images, showcasing images from SDUMLA, MMCBNU, and FV-USM.

SDUMLA-HMT Dataset

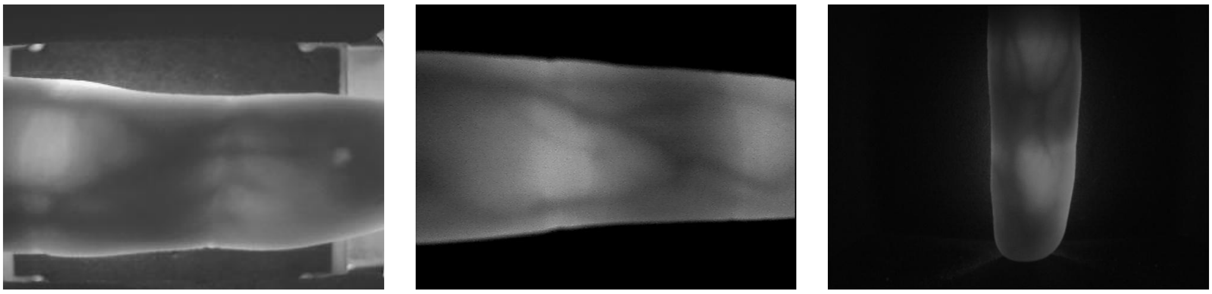
The SDUMLA-HMT dataset [66] comprises data collected from 106 participants, each contributing 36 finger vein images. These images are captured six times from three fingers on both hands, amounting to a total of 3,816 photographs. This rich dataset provides a diverse range of finger vein patterns, enabling a rigorous assessment of our model's ability to discern unique characteristics.

USM-FV Dataset

For the USM-FV dataset [71], we enlisted the participation of 123 individuals, each contributing images of the index and middle fingers on both hands. Each finger is photographed six times across two sessions, culminating in a comprehensive dataset of 5,904 photos. The diversity of participants and the substantial number of images offer a robust benchmark for our model's performance.

MMCBNU-FV Dataset

The MMCBNU-FV dataset [68]features images from 100 individuals, encompassing the index, middle, and ring fingers from both hands. Each participant's fingers are photographed ten times during six sessions, resulting in a dataset of 6,000 finger vein images. This dataset offers a sizable and diverse set of samples for our model to learn



(a) (b) (c)

Figure 4.3: Finger Vein Image Samples: (a) SDUMLA, (b) MMCBNU, (c) FV-USM.

Dataset	Years	# of subjects	# of sessions	#of images/subject	Total images
SDUMLA -HTM	2011	106	1	6	3816
MMCBNU 6000	2013	100	2	10 (10 images per session)	6000
FV-USM	2014	123	2	6	5904

Table 4.1: DESCRIPTIONS OF THE THREE PUBLIC FV DATASETS.

from.

4.2.2 Data Preprocessing and Augmentation

To ensure consistency and optimal performance, all images from these datasets are resized to a standardized dimension of 224×224 pixels and normalized from the range [0, 255] to [0, 1]. Furthermore, the datasets are partitioned into three subsets: 70% for training, 15% for validation, and 15% for testing. Recognizing the inherent challenges posed by small datasets, we employ data augmentation techniques to counteract overfitting during model training. Techniques such as rotation, random flipping, and adjustments in brightness are applied to diversify the training samples, enriching the model's ability to generalize across the datasets. This meticulous approach to dataset preparation ensures that our InceptionResnet-V2-based model is exposed to a comprehensive and diverse array of finger vein images, enabling it to achieve robust and reliable identification results. In the subsequent section, we delve into the experimental setup and performance evaluation, shedding light on the rigor employed to assess the model's capabilities.

4.3 Results and Discussion

In this section, we present the results of our experiments and engage in a comprehensive discussion of the outcomes of the proposed model, which has been implemented using the TensorFlow framework.

4.3.1 Experimental Setup

Our experiments were conducted on a system equipped with an Intel CORE i7 10510U processor, 16GB of RAM, and an Nvidia GTX 1650 GPU. The model was trained using the Adam optimizer [77], employing an initial learning rate of 10e-4. The loss function used was Sparse Categorical Cross-Entropy, and a batch size of 32 was employed for training.

4.3.2 Evaluation Metric

Equal Error Rate (EER)

To assess the performance of the proposed model, we employed the Equal Error Rate (EER), a widely recognized metric in biometric systems. EER is the point at which the False Acceptance Rate (FAR) equals the False Rejection Rate (FRR). This metric is instrumental for comparing the accuracy of different models with various Receiver Operating Characteristic (ROC) curves in image recognition tasks.

The formulas for calculating FAR and FRR are as follows:

$$\mathrm{FAR} = \frac{\text{Number of matching scores in false acceptance}}{\text{Total number of matching scores}} \tag{4.1}$$

$$\mathrm{FRR} = \frac{\text{Number of matching scores in false rejection}}{\text{Total number of matching scores}} \tag{4.2}$$

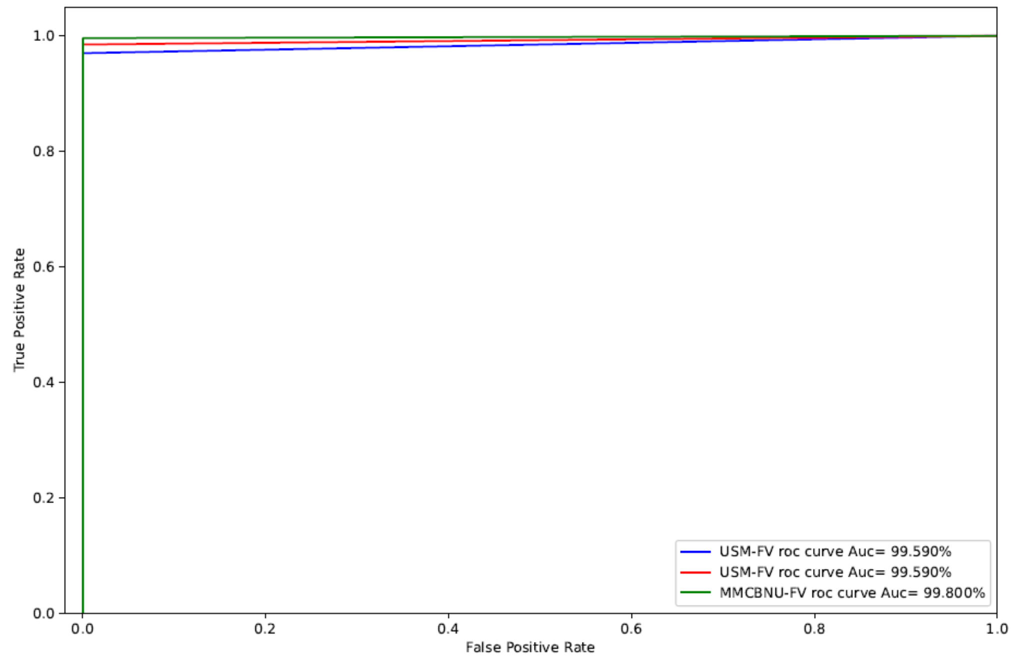


Figure 4.4: ROC curves obtained on SDUMLA, USM-FV, and MMCBNU datasets

Model	Accuracy	Precision	Recall	Sensitivity	Specificity
SDUMLA	98.980%	98.983%	98.980%	98.484%	97.926%
FV-USM	99.097%	99.544%	99.095%	99.193%	98.747%
MMCBNU 6000	99.660%	99.823%	99.651%	99.892%	99.751%

Table 4.2: Results of the InceptionResNet-V2 Model

4.3.3 Results

The results of our experiments are presented in Tables 4.2, and Figure 4.4 illustrates the ROC curves of the proposed model on the three databases. A glance at Table 4.2 demonstrates that the fusion of Inception and ResNet blocks within the InceptionResnet-V2 model yields robust and efficient results on all three datasets.

4.3.4 Comparison with State-of-the-Art (SOTA)

We further evaluate the performance of our proposed model by comparing it with the State-of-the-Art (SOTA) FV identification methods, including VGG-16 [72], FV-GAN [74], CNN with center loss and dynamic regularization [24], and CNN with large margin and softmax loss [78] on the three databases.

Ref	**Name of Method**	**Equal Error Rate (EER)**
[72]	VGG-16	1.37%
[74]	FV-GAN	0.94%
[79]	Deep Generalized Label Algorithm	2.23%
Our [80]	Proposed method	0.725%

Table 4.3: Comparisons of EER on SDUMLA-DB for the proposed model and recent models

Ref	**Method**	**EER**
[78]	CNN (Large Margin Softmax Loss)	0.76%
[24]	CNN combining center loss and dynamic regularization	1.07%
Our [80]	InceptionResNet-V2 pretrained model	0.41%

Table 4.4: Comparisons of EER on FV-USM-DB for the proposed model and recent models

The results of these comparisons are detailed in Tables 4.3, 4.4, and 4.5. Notably, in Table 4.3, our proposed model achieves the lowest EER value of 0.725%, outperforming Alexnet (EER of 0.8%) and other methods [72], [74],[79], which report EER values exceeding 0.9%. This observation underscores the superior accuracy of our proposed model.

Tables 4.4, and 4.5 demonstrate that our model significantly outperforms CNN-based models. In the FV-USM dataset, our model attains an EER value of 0.41%, while CNN-based models exhibit EER values greater than 0.7%. Similarly, in the MMCBNU dataset, our model achieves an EER of 0.2%, while CNN-based models report EER values exceeding 0.3%. These results further emphasize the superior performance of our proposed model across all three databases.

In summary, the experimental results validate the high performance of our proposed model on all three datasets, establishing its superiority over the existing State-of-the-Art methods.

Ref	**Method**	**EER**
[78]	CNN (Large Margin Softmax Loss)	0.30%
[24]	CNN combining center loss and dynamic regularization	0.503%
[81]	Shallow Convolutional Neural Network	0.47%
Our [80]	InceptionResNet-V2 pretrained model	0.20%

Table 4.5: Comparisons of EER on MMCBNU 6000-DB for the proposed model and recent models

4.4 Conclusion

In this study, we introduced a novel InceptionResnet-V2 deep learning model for the purpose of finger vein identification, leveraging the power of transfer learning. The proposed model, was designed to enhance the field of finger-vein-based biometrics. Through rigorous experimentation and comparisons with State-of-the-Art (SOTA) methods on three publicly available databases - SDUMLA, MMCBNU, and FV-USM - we have provided compelling evidence that our approach surpasses existing methods in terms of robustness, security, and accuracy. Our contributions and key findings can be summarized as follows:

1. We harnessed the advanced capabilities of the InceptionResnet-V2 architecture, optimizing it for finger vein identification through transfer learning.

2. Experimental results demonstrated the superiority of our proposed model in comparison to existing SOTA methods. Our model outperformed them across all three datasets, achieving lower Equal Error Rates (EERs) and showcasing its robustness and reliability.

3. The potential for future work is promising, with a particular focus on the exploration of advanced deep learning models, such as Capsule Networks (Caps-nets) and Vision Transformers. These explorations aim to further enhance the recognition performance of finger vein identification systems, ensuring ongoing innovation and improvement in this critical area of biometrics security.

In conclusion, our work contributes to the evolving landscape of biometric security, offering a highly accurate and robust solution for finger vein identification. The results presented in this study underscore the potential for utilizing deep learning to push the boundaries of biometric identification systems, with an unwavering commitment to enhancing both security and performance.

Chapter 5 introduces the FVCT model, a novel hybrid Convolutional Transformer-based approach for finger vein identification. This model, blending the power of convolutional neural networks (CNNs) and transformers, captured intricate local-to-global relationships exceptionally. FVCT surpassed current transformer and hybrid models, establishing a new performance benchmark. This chapter's research underscores the promise of advancing finger vein identification by integrating CNNs and transformers.

CHAPTER 5

A hybrid convolutional Transformer-Based Network Model for Finger Vein Identification

5.1 Introduction

In the realm of biometric recognition, the identification of individuals is primarily based on their distinctive physical characteristics, such as fingerprints, voice patterns, or iris characteristics [3]. With the increasing demand for digital security in sectors like online finance and security, biometric recognition plays a vital role in monitoring and verifying identities. Compared to traditional identification methods, biometric recognition technologies offer enhanced effectiveness, simplicity, and consistent security.

Among the emerging technologies for biometric identification, finger vein (FV) recognition stands out as a unique feature located within the hypodermic layer, making it more challenging to steal or replicate compared to other biometric features on the body surface [5] . FV identification boasts several advantages over alternative biometric methods:

1) Enhanced stability: The dispersed nature of FV under the finger skin results

in less variation depending on the individual's age and weight. Additionally, the FV is shielded directly by the human skin, preventing pollution by external factors and reducing susceptibility to damage.

2) Inherent difficulty of usurpation: Due to the specific distribution and imaging circumstances of the FV, obtaining FV images without the owner's consent is significantly challenging.

3) User-friendly operation: FV authentication entails a straightforward process where users need only place one of their fingers on a finger vein device to perform fast and effortless identification.

4) Liveliness detection capability: FV imaging exhibits a distinct distribution of gray levels due to the veins' ability to absorb near-infrared light at a different rate than other finger tissues, enabling the detection of liveliness.

5) Portability: VF identification devices are designed to be compact, slightly larger than the size of a finger, making them easily portable and convenient.

In recent years, the surge in graphics processing units (GPUs) and publicly available FV databases has led to a surge of deep learning-based FV identification algorithms. These algorithms surpass traditional methods by enabling deep neural networks to automatically learn hierarchical features, obviating the need for manual feature extraction. Convolutional neural networks (CNNs) [82] have been the primary deep learning choice in FV biometric identification, with a particular emphasis on enhancing CNN-based architectures. For example, Das & al. [23] proposed a CNN-based FV biometric recognition system, while Li & al. [83] compared the performance of CNN, AlexNet, and VGG-16 for FV identification. Additionally, Lu & al. [84] introduced a pre-trained CNN model and a CNN-based local descriptor for FV identification.

However, CNNs may struggle with capturing spatial dependencies among underlying target features, leading to suboptimal performance. Furthermore, CNN pooling layers often result in information loss, limiting improvements in FV identification accuracy.

In contrast, the Transformer architecture, initially introduced for Natural Language

Processing (NLP) tasks [85], has recently emerged in computer vision applications [86]. Transformer models, empowered by self-attention mechanisms, excel at capturing intrinsic properties. Dosovitsky & al. [87] introduced the Vision Transformer (ViT) model, which excelled in various image classification benchmarks. Unlike CNN models, ViT integrates global contextual information using self-attention mechanisms, allowing the extraction of robust features that account for long-range dependencies. Consequently, Transformer-based models have gained ground in computer vision applications [86]. However, ViT's performance falls behind CNNs in low-data scenarios [88], despite remarkable results with large JFT 300M [89] datasets.

On the contrary, Convolutional Neural Networks (CNNs) possess inherent priors such as translation invariance (through shared convolutional weights) and scale invariance (via pooling). These priors enable CNNs to learn effectively even from smaller datasets [88]. However, when compared to Transformers, CNNs struggle to capture long-range dependencies, necessitating deeper networks with multiple layers to increase the receptive field. Recent studies have sought to combine the strengths of CNNs and Vision Transformers (ViTs) by leveraging their complementary advantages [90, 91, 92, 93]. This fusion has given rise to more potent computer vision models that harmonize both approaches, resulting in improved performance across various computer vision tasks [86].

Motivated by this development, this work introduces a hybrid Convolutional Transformer-based model for finger vein identification. By harnessing feature extraction capabilities from both CNNs and ViTs, we devise a specialized classifier for finger vein identification. Our empirical analysis demonstrates the superior performance of our proposed technique against several robust baselines.

5.1.1 Contributions

The primary contributions of this work are as follows:

1. Baseline Experimental Results: This chapter presents baseline experimental results assessing the performance of hybrid Conv-Transformer and ViT models in

finger vein (FV) identification. The study focuses on four state-of-the-art (SOTA) Transformer and hybrid Conv-Transformer models, highlighting their efficacy in accurately identifying finger vein patterns. Nonetheless, there remains room for further performance improvement.

2. Hybrid Conv-Transformer Model: We introduce a hybrid Conv-Transformer FV identification model, named FVCT. Comparative analysis reveals its superiority over existing approaches like DeiT [94], Cait [94], Coatnet [95], and ConvMixer [96] in the context of FV identification. Furthermore, FVCT exhibits competitive performance with state-of-the-art models.

3. Comprehensive Experiments: We conduct extensive experiments on three widely recognized finger vein datasets. The experimental protocol, setup, and evaluation metrics are meticulously described to ensure a fair and comprehensive comparison.

5.1.2 Structure of the Chapter

This chapter is structured as follows:

Section I - Introduction: provides an overview of the chapter's content, outlining the key topics and research questions addressed.

Section II - Literature Review: A comprehensive review of related works in the field. This section explores existing research and advancements in finger vein identification, transformers, vision in transformers, and hybrid transformers.

Section III - Transformers in Computer Vision: An in-depth overview of transformers, including their architecture and functioning. We introduce the concept of vision in transformers and delve into the details of the proposed FVCT model, which combines the strengths of Convolutional Transformers to address the unique characteristics of finger vein information.

Section IV - Experimental Results: We present the experimental results of recent ViT models, hybrid Conv-Transformer models, and the proposed FVCT model. Ablation models are discussed, focusing on their performance in finger vein identification tasks. Additionally, we conduct a comprehensive comparison with state-of-the-art finger vein identification approaches to assess the competitiveness of our proposed model.

Section V - Conclusion and Future Directions: The article concludes by summarizing the key findings and contributions. We also outline potential avenues for future research and development in the field of finger vein identification.

In Chapter 4, we will delve further into the experimental results and provide detailed discussions on the performance of the hybrid Convolutional Transformer-based model and its implications for the field of finger vein identification.

5.2 Literature Review

Finger vein (FV) identification methods can be broadly categorized into two groups: handcrafted feature-based approaches and deep learning-based approaches. In this section, we provide a comprehensive review of key studies associated with these methodologies.

5.2.1 Handcrafted Feature-Based Approaches

The first category encompasses manual feature extraction methods employed for FV recognition, often involving techniques that leverage local grayscale variations within FV images to extract vein pattern features. Some notable studies include:

Miura & al. [97] proposed an algorithm that extracts vein pattern features from FV images based on local grayscale variations. Qin & al. [98] improved feature extraction by incorporating the region growth method, resulting in more precise finger vein pattern extraction. Miura & al. [99] introduced the maximum curvature algorithm, which determines the maximum curvature of local cross-sections within the image, enhancing feature extraction accuracy. Gupta & al. [100] proposed local multi-scale matching filters to address issues related to low-quality FV images, mitigating noise caused by uneven illumination and improving recognition performance. Rosdi & al. [37] introduced a local linear binary pattern feature based on finger veins, allowing the extraction of coding features within linear local regions. Van & al. [101] presented a novel method for obtaining local invariant directional vein features from digital vein data, which was combined with GridPCA to remove redundant data and enhance finger vein recognition accuracy.

5.2.2 Deep Learning-Based Approaches

With the rise of deep learning techniques, researchers have explored methods that leverage neural networks for FV identification. Some significant advancements in this category include:

He & al. [74] proposed a multi-layer classifier based on backpropagation neural networks for FV image identification, enhancing it with Principal Component Analysis (PCA) to extract low-level details and improve accuracy and robustness. Tang & al. [102] developed a simple FV feature extraction method using a pre-trained Convolutional Neural Network (CNN) and distillation learning, achieving high performance and fast inference capability. Yang & al. [103] introduced FV-GAN, utilizing a fully convolutional network for feature extraction and classification of FVs. Another approach presented by Yang & al. [103] involved processing FV image sequences using CNN and Long Short-Term Memory (LSTM) networks. Hou & al. [104] proposed an ECA-Resnet model with channel attention and residual connections, combined with an arccosine center loss function for FV image identification. Zhao & al. [24] introduced a lightweight CNN model with a central loss function and dynamic regularization, achieving low error rates with reduced computational complexity.

5.2.3 Transformer-Based Approaches

More recently, Transformer-based techniques have made significant strides in the field of FV identification. Some notable studies in this emerging area include:

Huang & al. [105] introduced the Finger Vein Transformer (FVT) model for authentication, showcasing competitive results. Li & al. [106] combined the Vision

Transformer (ViT) with a capsule network for finger vein recognition, demonstrating the effectiveness of this fusion in FV identification.

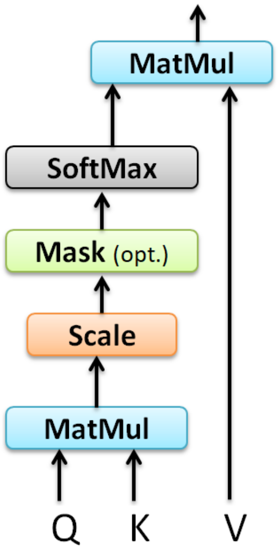
This comprehensive review of related work highlights the evolution of FV identification methodologies, from handcrafted features to deep learning and the recent emergence of Transformer-based approaches. Each category has contributed to the advancement of the field, with deep learning and Transformer models showing promise in achieving higher accuracy and robustness in finger vein identification. In the following sections, we delve deeper into the experimental results and implications of these approaches.

5.3 Methodology

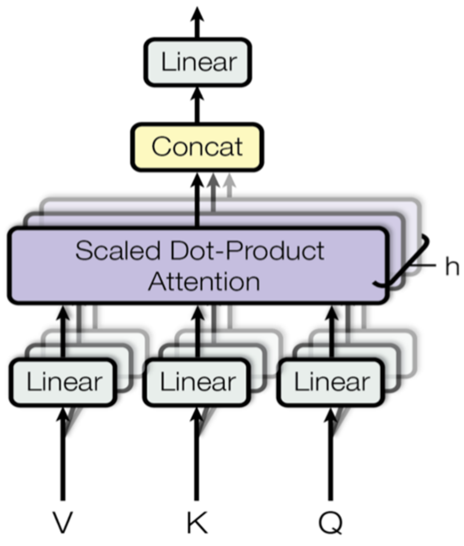
In this section, we present the methodology employed in this study, which includes an overview of transformers, the concept of vision in transformers, and the integration of hybrid convolution transformers. Subsequently, we introduce our proposed model, the Hybrid Conv+Transformers, which leverages this newly developed architecture for accurate identification and classification of finger vein (FV) images.

5.3.1 Transformer

Originally developed for Natural Language Processing (NLP) tasks, transformers have emerged as highly efficient architectures for modeling sequential data, such as sentences or sequences of words. Transformers offer significant advantages over other sequential models, such as Recurrent Neural Networks (RNNs), by addressing various challenges. The core building blocks of transformers consist of stacked transformer blocks, which are multilayer networks comprised of simple linear layers, feed-forward networks, and self-attention layers. The self-attention mechanism, depicted by the "Multi-Head Attention" box in Figure 5.1, is the key innovation of transformers.



(a)



(b)

Figure 5.1: Attention Mechanisms: (a) Self-attention mechanism, (b) Multi-head attention.

5.3.1.1 Self-Attention

The self-attention layer comprises a multi-head self-attention mechanism and a fully connected feed-forward network. The multi-head mechanism consists of h self-attention layers, each performing scaled dot-product attention. Between them, two linear transformations and a ReLU activation function are applied in the fully connected feed-forward network.

The self-attention mechanism begins by multiplying the query vector with the key matrix, resulting in a query matrix Q of dimension model. The self-attention layer then calculates the attention using the query matrix Q, the key matrix K, and the value matrix V. Finally, the output vector is obtained by transforming the self-attention layer. The three matrices Q, K, and V contain vectors from different inputs. The construction of the attention function between these input vectors can be summarized as follows (Figure 5.1.a):

1. Calculate the dot product of different input vectors, denoted as $S = Q \cdot K$.

2. Normalize the resulting values for gradient stability using $S_n = \frac{S}{\sqrt{d_K}}$.

3. Apply the softmax function to obtain probability values, denoted as $P = \text{Softmax}(S_n)$.

4. Compute the weighted sum of the value matrix using $Z = V \cdot P$.

This process can be consolidated into a single function:

$$\text{Attention}(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5.1}$$

where d_K represents the dimensionality of the key matrix. The output vector of the attention layer is obtained by concatenating the vectors $A^{\{1\}} \cdots A^{\{m\}}$, where m denotes the number of attention heads.

5.3.1.2 Multi-Head Attention

The Multi-Head Attention layer extends the capabilities of the self-attention layer by incorporating multiple self-attention layers. Each self-attention layer attends to different parts of the input vectors, and the final output of the multi-head attention layer is a linear combination of the outputs from these individual attention layers. The multi-head attention layer takes a sequence of vectors Q, K, and V as input, all of which have the same dimensionality. The parameter h represents the number of attention heads, and the dimensionality of each self-attention layer is defined as $d_k = \frac{d_m}{h}$. The vectors from the different inputs are organized into separate matrices: Q, K, and V. Then, self-attention is applied to each vector in these matrices:

$$\text{MultiHead}(Q,K,V) = \text{Concat}(head_1, ..., head_h)W^O \tag{5.2}$$

Here, $W^O \in \mathbb{R}^{d_m \times d_m}$ is a learnable matrix. The matrices Q, K, and V are divided into h sub-matrices. The self-attention function is applied to each sub-matrix, and the resulting outputs are concatenated to produce the final output of the multi-head attention layer.

5.3.1.3 Transformers in Vision

Transformer-based architectures have proven to be effective for solving vision problems, as demonstrated by the introduction of Vision Transformers (ViT). In the ViT model, Dosovitskiy & al. [87] employed a patch-based approach where the image is divided into fixed-size patches. Each patch is then linearly embedded, positional embeddings are added, and the resulting sequence of vectors is fed into a standard Transformer encoder. For classification, a "classification token" is added to the sequence, and standard methods are used. To account for the locality and 2D properties of images, the Swin Transformer adopts a hierarchical architecture that utilizes shifted windows [107]. Recent advancements in transformer-based vision models have focused on improving both model and data efficiency. These advancements include techniques such as sparse attention [108, 109, 110], pyramid design [111], enhanced locality [112], and improved training strategies [95, 113], among others. For a comprehensive review, we direct readers to the dedicated survey on Vision Transformers [86].

5.3.1.4 Hybrid Models

Purely Transformer-based vision models have been observed to exhibit poor generalization due to their relatively low induced bias [87, 90]. Additionally, Vision Transformers often face challenges in terms of optimizability, resulting in subpar performance [91]. To address these limitations, a hybrid approach that combines Transformer and Convolution layers has been proposed. This hybrid design involves replacing the coarse patchify stem of the Transformer model with a few convolutional layers [90, 91]. By incorporating convolutional layers, the Transformer model can benefit from their ability to capture local and spatial information, leading to improved performance [90, 91, 92, 93].

5.3.2 FVCT MODEL

The integration of convolution and self-attention mechanisms in the FVCT (Finger Vein Convolution-Transformer) model is crucial for the following reasons:

1) Considering practical applications, both Transformers and CNNs have their own

strengths and weaknesses. In general, transformer models tend to achieve better performance but are challenging to train and come with a high computational cost [94]. On the other hand, CNNs may not match the performance of transformers but offer unique advantages. CNNs are easier to train and benefit from better hardware support. Particularly, when it comes to small models intended for mobile or edge devices, CNNs still dominate [114].

2) CNNs and transformers exhibit distinctive characteristics in information processing. Transformers excel at extracting global information and capturing dependencies across different positions in the input data [115]. On the other hand, CNNs possess inductive biases that provide strong priors for capturing local dependencies [90].

To overcome these challenges and leverage the strengths of both architectures, we propose a novel Finger Vein Convolution-Transformer Network (FVCT) for finger vein identification tasks. Our model builds upon the architecture of Coatnet [90]. In this section, we provide a detailed description of the proposed FVCT model, which combines convolutional layers and transformer layers to effectively capture both local and global dependencies. We compare our FVCT model with the main Coatnet architecture, as illustrated in Table 5.3. We also conduct an ablation study by evaluating it against two other Coatnet architectures. Furthermore, we have made modifications to reduce the number of parameters in our FVCT architecture, as outlined in Table 5.2. These contributions highlight the novelty and advancements of our proposed approach compared to the Coatnet architecture.

5.3.2.1 FVCT Architecture

Our objective is to develop a hybrid network that harnesses the strengths of Convolutional Neural Networks (CNNs) and transformers. Figure 5.2 provides a visual depiction of the FVCT architecture, highlighting its key components and information flow.

The FVCT architecture begins with an input image, which undergoes a series of stages to extract features at multiple levels. These stages are labeled as S0, S1, S2, S3, and S4.

In S0, an initial convolutional stem is applied to the input image. This stem stage includes a 3×3 convolution with a stride of 2, which aids in reducing the spatial size of the input images. It is then followed by two 3×3 convolutions with a stride of 1, enabling efficient extraction of local information.

Moving forward, each successive stage in S1, S2, S3, and S4 involves a spatial size reduction of $2X$ and an increase in the number of channels at the start of each stage.

In S1, a modified MBConv Block incorporating squeeze-excitation (SE) and GELU activations is employed to capture relevant features. The SE block enhances important information, while the GELU activation function introduces non-linearity to the model.

In S2 and S3 stages, Transformer blocks are integrated to leverage the power of self-attention and capture long-range dependencies in the data. These Transformer blocks enhance the network's ability to effectively capture contextual information.

The S4 stage concludes with another MBConv block, further refining the learned representations.

Finally, the FVCT model culminates with a global average pooling layer, which aggregates the spatially extracted features, followed by a fully connected classification layer with a softmax activation function for making the final classification decision.

By combining CNN-based feature extraction with Transformer-based attention mechanisms, the FVCT architecture aims to deliver a potent and efficient model for Finger Vein Identification. Figure 5.2 provides a visual representation of this architecture, illustrating the flow of information and the role of each component in the network.

5.3.2.2 Model Details

a) MBConv

The MBConv block [116] serves as the primary convolution operator in the model. Both the MBConv and Transformer blocks adopt pre-activation structures to ensure consistency in the model architecture [117]. The pre-activation structure is used to consistently promote homogeneity between MBConv and Transformer blocks [95, 118].

Specifically, assuming x is an input feature, the formulation of the MBConv block

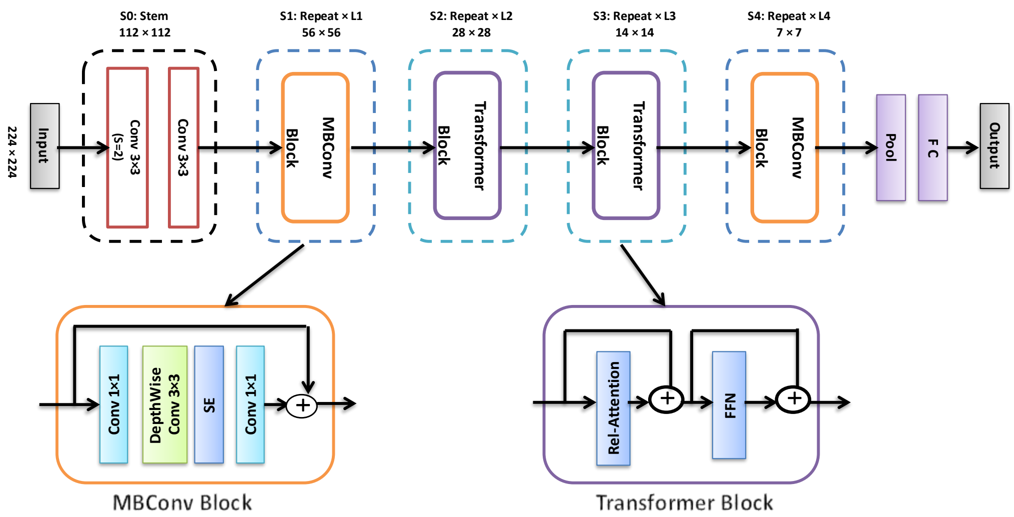


Figure 5.2: Proposed model

without downsampling is as follows:

$$x \leftarrow x + \mathrm{Proj}(SE(DWConv_{3\times3}(Conv(Norm(x))))) \tag{5.3}$$

Here, "Norm" refers to BatchNorm [119], and "Conv" denotes the extended Conv1x1 followed by BatchNorm and GELU activation function [120], which are common choices for Transformer-based models. "DWConv" stands for Depthwise $\mathrm{Conv}_{3\times3}$, followed by BatchNorm and GELU. "SE" represents the Squeeze-Excitation layer [121], and "Proj" is the down-projecting $\mathrm{Conv}_{1\times1}$ to reduce the number of channels.

In each stage, the first MBConv block undergoes downsampling through the application of a Stride-2 Depthwise $\mathrm{Conv}_{3\times3}$, and pooling and channel projection are applied to the shortcut branch:

$$x \leftarrow x + \mathrm{Proj}(\mathrm{Pool2D}(x) + \mathrm{Proj}(SE(DWConv_{3\times3}(Conv(Norm(x)))))) \tag{5.4}$$

This formulation ensures that the MBConv block captures and refines features at various stages of the FVCT architecture.

b) Relative Attention

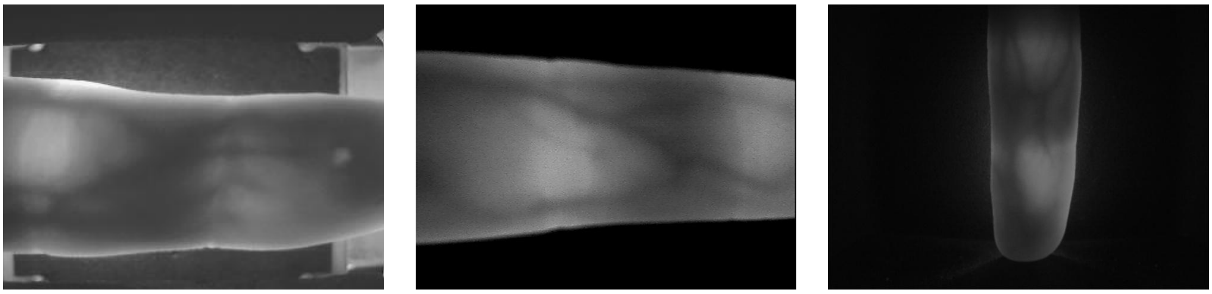
For our model, we select a single head from the multi-head attention. Typically, the same head dimension is used in multi-head attention implementations. The relative attention is defined as follows:

$$\text{RealAttention}(Q,K,V) = \text{Softmax}\left(\frac{Q.K^T}{\sqrt{d_K}} + B\right) \cdot V \tag{5.5}$$

Here, Q, K, and V are query, key, and value matrices of size $\mathbb{R}^{(H\times W)\times C}$, respectively. B represents the learned static location-based matrix. $\frac{Q.K^T}{\sqrt{d_K}}$ denotes the scaled input adaptive attention, where d_K refers to the hidden dimension. The attention weights are jointly determined by B and $\frac{Q.K^T}{\sqrt{d_K}}$. Taking into account the differences in two-dimensional coordinates, the relative positional distortion B is parameterized by a matrix. According to typical practice [90], when fine-tuning is performed at a higher resolution, such as $H' \times W'$, bilinear interpolation is used to calculate B from $\mathbb{R}^{(2H-1)(2W-1)}$ to $\mathbb{R}^{(2H'-1)(2W'-1)}$. This relative attention benefits from input adaptivity, translation equivariance, and global interactions. Therefore, this relative attention mechanism is utilized by default in the attention operators of our model.

5.4 Experiments and Results

This section provides a comprehensive description and analysis of the conducted experiments. Firstly, we introduce three widely recognized Finger Vein (FV) datasets, which serve as the basis for evaluating our proposed protocol in terms of training, validation, and testing. Secondly, we present in-depth details of the experimental setups and metrics employed for the purpose of comparison. The third part showcases the results obtained from our FVCT model. Furthermore, utilizing the same experimental configuration, we conduct experiments on four recently developed Vision Transformer (ViT) and hybrid Convolution-Transformer models, enabling meaningful comparisons with



(a) (b) (c)

Figure 5.3: Finger Vein Image Samples: (a) SDUMLA, (b) MMCBNU, (c) FV-USM.

Dataset	Years	# of subjects	# of sessions	#of images/subject	Total images
SDUMLA -HTM	2011	106	1	6	3816
MMCBNU 6000	2013	100	2	10 (10 images per session)	6000
FV-USM	2014	123	2	6	5904

Table 5.1: DESCRIPTIONS OF THE THREE PUBLIC FV DATASETS.

our proposed FVCT. In the fourth part, we perform ablation experiments by selectively modifying design choices within the FVCT model, aiming to assess their effectiveness. Finally, a comprehensive comparison is conducted with state-of-the-art approaches in Finger Vein Identification to establish the superiority of our proposed method.

5.4.1 Datasets

We use the same datasets described in 4.2.1 for evaluating the proposed FVCT model.

Table 5.1 presents a summary of the Finger Vein (FV) datasets utilized in this study. The datasets were partitioned into three subsets: 70% for training, 20% for testing, and 10% for validation, employing person-level splits to ensure data integrity and prevent leakage. During training, the images were normalized to a size of $224 \times 224 \times 3$. Additionally, the single-channel grayscale images were duplicated twice to transform them into three-channel images. To enhance performance and mitigate overfitting, various image augmentation techniques, such as Horizontal Flip, Vertical Flip, and zoom, were employed to augment the training data and increase the variability of images during each training epoch.

5.4.2 Experimental Setup

5.4.2.1 Evaluation Metrics

The performance of the proposed model is evaluated using the Equal Error Rate (EER) metric. EER is defined as the point on the Receiver Operating Characteristic (ROC) curve where the False Acceptance Rate (FAR) and False Rejection Rate (FRR) are equal. The EER provides a concise measure for comparing the accuracy of different models in image recognition tasks. A lower EER indicates higher accuracy. FAR and FRR are calculated using the following formulas:

$$\text{FAR} = \frac{\text{Number of matching scores in false acceptance}}{\text{Total number of matching scores}} \tag{5.6}$$

$$\text{FRR} = \frac{\text{Number of matching scores in false rejection}}{\text{Total number of matching scores}} \tag{5.7}$$

These metrics help quantify the model's performance in terms of its ability to distinguish between genuine and impostor finger vein samples. By analyzing the FAR and FRR values, the model's accuracy and robustness can be assessed effectively.

5.4.2.2 Implementation Details

The experiments were conducted using the PyTorch [122] and Timm [123] libraries. We utilized pre-trained Cait and Deit models, which had been trained on the Imagenet dataset, as our Transformer models. The AdamW optimizer [124] was employed for training the models.

During training, the batch size was set to 32, momentum was set to 0.9, and weight decay was set to 0.05. We used a cosine learning rate schedule with an initial learning rate of 10^{-4}. The input images were resized to a size of 224, and both the training and testing procedures were executed on an NVIDIA Tesla T4 GPU for efficient computation.

For each dataset, we trained the models for a fixed number of epochs: 20 epochs on the SDUMLA dataset and 30 epochs on the FV-USM and MMCBNU datasets. Al-

Stage	Size	# of blocks	# of channels
S0- Conv-stem	1/2	2	64
S1- MBConv	1/4	2	96
S2- Transformer	1/8	5	192
S3- Transformer	1/16	3	384
S4- MBConv	1/32	2	768

Table 5.2: FVCT architecture configurations

though we did not explicitly define a stopping criterion, our observations during the experiments indicated that the chosen number of epochs was sufficient for convergence. Further training did not yield significant improvements in performance, leading us to conclude that the models had reached convergence within the specified epochs.

5.4.2.3 FVCT model configurations

The FVCT model architectural variants are presented in Table 5.2. The number of channels is incrementally increased from Stage S1 to S4 by a factor of two, while ensuring that the width of the Stem S0 is less than or equal to that of S1. Furthermore, for simplicity, only the number of blocks in Stages S2 and S3 is scaled during the expansion of the network's depth. Each attention head in the attention layers is assigned a value of 32. The MBConv blocks maintain an expansion rate of 4 and a shrinking rate of 0.25 in the Squeeze-Excitation (SE) module. The architectural configurations for each stage are as follows:

1. Stage S0 (Conv-stem): 2 blocks with 64 channels.

2. Stage S1 (MBConv): 2 blocks with 96 channels.

3. Stage S2 (Transformer): 5 blocks with 192 channels.

4. Stage S3 (Transformer): 3 blocks with 384 channels.

5. Stage S4 (MBConv): 2 blocks with 768 channels.

Our objective in this study is to evaluate the performance of four deep learning models on three distinct datasets. We utilized the pre-trained Deit [94] and Cait [95]

Model		Pre-trained	Eval Size	#Params
Deit	Deit-S	*	224	22M
Cait	Cait_xs24	*	224	26.6M
Coatnet	CoAtNet-0	-	224	25M
ConvMixer	Convmixer_768_32	*	224	21M
FVCT(our model)	**FVCT**	-	224	15M

Table 5.3: Comparison of different types of models

Keys:
(*) pretrained model
(-) not pretrained model

Dataset	Deit		Cait		Coatnet		ConvMixer		FVCT (Ours)	
	Acc	EER	Acc	EER	Acc	EER	Acc	EER	Acc	EER
SDUMLA -HTM	93.33%	3.34%	92.27%	3.43%	98.17%	0.75%	98.66%	0.47%	99.46%	0.26%
FV-USM	91.43%	4.43%	89.35%	5.46%	97.22%	0.84%	99.09%	0.53%	99.54%	0.22%
MMCBNU 6000	97.00%	1.50%	95.63%	1.89%	99.12%	0.37%	98.97%	0.31%	99.71%	0.17%
Acc/EER weighted mean	93.92%	3.09%	92.41%	3.59%	98.17%	0.65%	98.90%	0.43%	**99.57%**	**0.21%**

Table 5.4: Performance of different models on three databases.

models as Transformer models, while Coatnet [90] and ConvMixer [96] served as hybrid models. To ensure credibility, we selected models with smaller sizes in terms of the number of parameters. As shown in Table 5.3, our FVCT model has the fewest parameters among the four models, highlighting our emphasis on evaluating its performance relative to similarly sized models.

5.4.3 Results and Discussion

The application of Transformer and Conv-Transformer methods in Finger Vein Identification is relatively limited in existing literature. To address this gap, we conducted experiments comparing our proposed model with four state-of-the-art models in the field: two Transformer models (Deit and Cait) and two hybrid Conv-Transformer models (ConvMixer and Coatnet). These models were selected to assess the performance of our proposed model and establish its competitive positioning in the domain.

The experimental results are summarized in Table 5.4, utilizing the experimental setup and metrics described in Section 5.4.2. Figure 5.4 illustrates the ROC curves of the proposed Finger Vein Convolution-Transformer (FVCT) model for the three Finger

Vein datasets.

Firstly, as observed from Table 5.4, the weighted mean Equal Error Rates (EERs) of the Deit and Cait Transformer models are 3.09% and 3.59%, respectively. This suggests that while Transformer networks can exhibit efficiency in Finger Vein recognition, they may encounter challenges in achieving optimal performance when confronted with smaller datasets. This finding aligns with previous studies [88] that have reported the limitations of Transformer models in handling datasets with limited samples. In contrast, the hybrid Convolution-Transformer models, namely Coatnet and ConvMixer, deliver outstanding results with weighted EERs of 0.65% and 0.43%, respectively. Therefore, we conclude that the fusion of CNN and Transformer architectures has yielded remarkable performance for the Finger Vein identification task.

Secondly, the proposed FVCT model surpasses both the Transformer and hybrid Convolution-Transformer models on all Finger Vein datasets, achieving a weighted EER of 0.21%. These results clearly indicate that our FVCT model is not only effective but also exhibits generalizability across different Finger Vein datasets. Additionally, our model boasts the fewest parameters (Params) compared to the other models, as shown in Table 5.2. This benchmark further establishes the efficiency of our model for real-world recognition tasks.

Overall, the experimental results validate the superiority of our proposed FVCT model, which leverages the strengths of both CNN and Transformer architectures, for Finger Vein Identification. The combination of these two paradigms has yielded significant advancements in accuracy and efficiency, making it a promising approach for practical Finger Vein recognition systems.

5.4.4 Ablation Study

In this comprehensive ablation study, our objective is to meticulously investigate the impact of specific design choices within our FVCT model. To achieve this, we conducted a series of experiments on three public finger vein datasets: SDUMLA, FV-USM, and MMCBNU while maintaining consistent training configurations, with the exception of

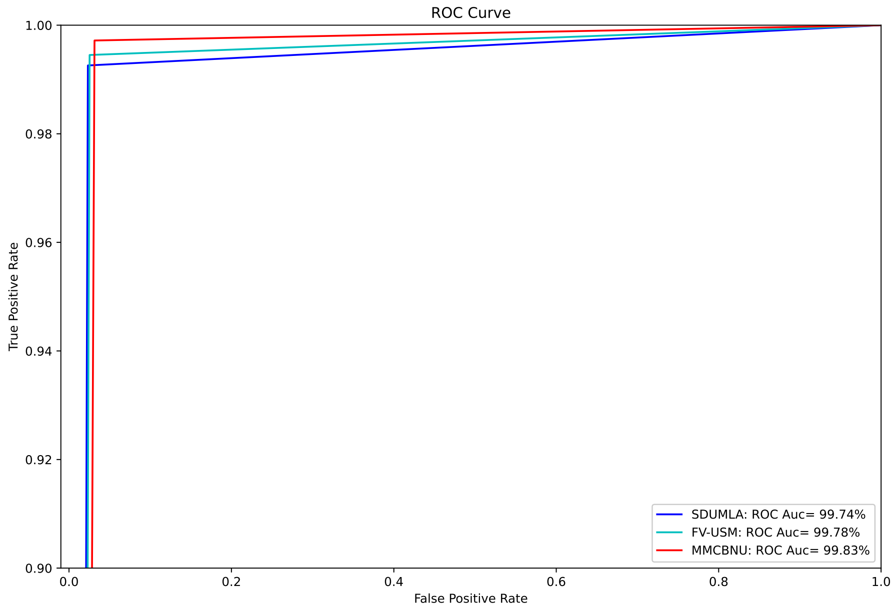


Figure 5.4: (ROC) curves of the FVCT model

the particular design choice under scrutiny.

a) FVCT Design

As outlined in Section 5.3.2, our model is built upon the esteemed Coatnet architecture. Coatnet explores diverse design variants with progressively increasing numbers of Transformer stages denoted as C-C-C-C, C-C-C-T, C-C-T-T, and C-T-T-T, where C represents Convolution and T represents Transformer. In accordance with the findings of Coatnet [90], the C-C-C-T and C-C-C-C architectures have been deemed the most effective, suggesting that models with a greater number of convolution stages tend to yield superior results on datasets of smaller sizes. Furthermore, when comparing architectures on the large-scale JFT dataset, it becomes apparent that the C-C-T-T architecture has the most significant impact on model capacity.

To evaluate the performance of our proposed architecture, namely C-T-T-C, we conducted dedicated experiments on the three finger vein datasets. As demonstrated in Table 5.5, our model consistently outperforms the two alternative architectures, C-C-C-T and C-C-T-T, thereby substantiating the superiority of our design choice.

b) MBConv with Squeeze Excitation

dataset	SDUMLA-HTM		FV-USM		MMCBNU 6000	
	Accuracy	EER	Accuracy	EER	Accuracy	EER
C-C-T-T (Coatnet)	98.17%	0.75%	97.22%	0.84%	99.12%	0.37%
C-C-C-T	99.03%	0.38%	98.72%	0.51%	99.44%	0.24%
C-T-T-C (Our model)	99.46%	0.26%	99.54%	0.22%	99.71%	0.17%

Table 5.5: Ablation on architecture layout.

dataset	SDUMLA-HTM		FV-USM		MMCBNU 6000	
	Accuracy	EER	Accuracy	EER	Accuracy	EER
With SE	99.46%	0.26%	99.54%	0.22%	99.71%	0.17%
Without SE	99.17%	0.31%	99.12%	0.29%	99.56%	0.19%

Table 5.6: Ablation of Squeeze Excitation.

In this segment of the ablation study, we scrutinize the significance of Squeeze Excitation (SE) within the MBConv block. We compare two models: one incorporating SE and the other without SE. The results are displayed in Table 5.6. The model incorporating SE achieved impressive results. This indicates that the inclusion of Squeeze Excitation in the MBConv block exerts a positive impact on the performance of our FVCT model, resulting in improved performance for Finger Vein identification across all datasets.

These meticulous ablation studies offer valuable insights into the design choices incorporated within our FVCT model. The results unequivocally establish the superiority of the C-T-T-C architecture and the indispensability of integrating Squeeze Excitation in the MBConv block, thereby underscoring their substantial contributions to the overall performance of the FVCT model across different datasets.

5.4.5 Comparison with recent Finger Vein identification methods

In this section, we compare the performance of our proposed Finger Vein Convolution-Transformer (FVCT) model with state-of-the-art (SOTA) methods for Finger Vein identification on the SDUMLA, MMCBNU, and USM datasets. Our comparison includes five SOTA methods that have been widely used in previous studies. Specifically, we selected methods proposed by Hou & al. [104], Zhao & al. [24], and Hu & al. [78],

ref	Year	Method	EER (%)		
			SDUMLA	MMCBNU	USM-FV
[78]	2018	Proposed CNN model	0.30	0.76	1.20
[24]	2020	lightweight CNN + center loss and dynamic regularization	-	0.5	1.1
[104]	2021	CNN+ Arccosine center loss (Arcvein)	1.53	-	0.25
[105]	2022	Transformer (FVT)	1.50	0.92	0.44
[106]	2022	Capsule network+ Transformer (ViT-Cap)	1.3	0.63	0.28
Ours [125]	2023	Hybrid Conv-Transformer (FVCT)	**0.26**	**0.17**	**0.22**

Table 5.7: Comparison with recent Finger Vein identification methods.

which utilize CNN algorithms, as well as methods proposed by Huang & al. [105] and Li & al. [106], which utilize Transformer algorithms. By comparing the performance of these methods with our proposed FVCT model, we aim to demonstrate the effectiveness and superiority of our approach in Finger Vein identification.

The results of the comparative analysis between our proposed FVCT model and five state-of-the-art (SOTA) methods for Finger Vein identification are presented in Table 5.7. Our FVCT model outperformed the other SOTA methods in terms of Equal Error Rate (EER) on all three datasets (SDUMLA, MMCBNU, and USM-FV).

Compared to the CNN-based methods proposed by Hu & al. [78] and Zhao & al. [24], our FVCT model achieved significantly lower EER values, demonstrating its superiority in extracting Finger Vein features. The lightweight CNN model with center loss and dynamic regularization by Zhao & al. achieved competitive results, but our FVCT model still outperformed it.

In terms of Transformer-based methods, our FVCT model surpassed the performance of the FVT model proposed by Huang & al. [105] and the Capsule network with Transformer (ViT-Cap) proposed by Li & al. [106]. This indicates that the combination of convolutional and Transformer layers in our FVCT model leads to more effective Finger Vein identification.

Overall, our FVCT model achieved the lowest EER values across all three datasets, highlighting its superiority compared to existing SOTA methods. This demonstrates the effectiveness and comprehensiveness of our FVCT model for Finger Vein identification tasks.

5.5 CONCLUSION

In this study, we conducted a comprehensive investigation into the integration of convolutions into the Vision Transformer (ViT) architecture. Our goal was to leverage the strengths of both convolutional neural networks (CNNs) and transformers for Finger Vein Identification. By combining the CNN's effectiveness in extracting low-level features and capturing local patterns with the ViT's ability to model long-range dependencies, we introduced a hybrid approach that merges the two paradigms.

We introduced the Finger Vein Convolution-Transformer (FVCT) model specifically designed for Finger Vein Identification, setting a new benchmark in the field. Our model outperformed Transformer models and hybrid Conv-Transformer models (Deit, Cait, Coatnet, ConvMixer) on the public finger vein datasets. Moreover, the FVCT model offered a unique representation of local-to-global relationships, enabling it to excel in Finger Vein Identification compared to other state-of-the-art models.

Future research should focus on further reducing the model's parameter count and computational cost of the hybrid Conv-Transformer model. Additionally, improving the Transformer model's capability to extract local information would be valuable for enhancing the model's performance. These advancements would contribute to the development of more efficient and accurate Finger Vein Identification systems.

CHAPTER 6

Conclusion

The field of biometric security systems has witnessed remarkable advancements and a shift toward more secure, efficient, and convenient methods of personal identification. This thesis, titled "Biometric Security System: Unimodal Identification Using Finger Veins," has undertaken an in-depth exploration of the realm of finger vein identification. The journey has led to significant contributions to the field and laid the foundation for future innovations in biometric security systems. In this concluding chapter, we summarize the key findings, reflect on the research process, and provide recommendations for future work in this critical area of biometrics.

The primary research question that guided this thesis was: Can finger vein identification be advanced as a secure and efficient unimodal biometric recognition modality? Through the exploration of various aspects of finger vein identification, including historical development, feature extraction techniques, and the application of deep learning models, we have provided a comprehensive response to this question. The findings of this thesis affirm that finger vein identification holds the potential to be an advanced, secure, and efficient unimodal biometric recognition modality.

6.1 Summarizing the Research Process

The research process embarked with a thorough examination of the overarching field of biometric systems. Chapter 2 provided a solid foundation for understanding the fundamentals of biometric systems, their operation, performance evaluation, and diverse applications. This chapter highlighted the need for advanced, secure, and efficient biometric recognition methods.

Chapter 3 delved into the intricacies of finger vein identification, tracing its historical development, exploring various techniques for feature extraction, and emphasizing the significance of databases. The chapter laid the groundwork for understanding the anatomical and operational aspects of finger vein identification, setting the stage for further exploration.

In Chapter 4, the focus shifted to the development of a deep learning model for finger vein identification based on the InceptionResnet-V2 architecture. The InceptionResnet-V2 model, customized for finger vein identification, showcased superior performance in comparison to existing state-of-the-art methods. The research in this chapter demonstrated the potential of deep learning models in enhancing the security and accuracy of finger vein identification.

Chapter 5 introduced a hybrid Convolutional Transformer-based model for finger vein identification. This model, known as FVCT, harnessed the strengths of both convolutional neural networks (CNNs) and transformers, offering a unique representation of local-to-global relationships. The FVCT model outperformed existing transformer and hybrid models, setting a new benchmark in the field. The research in this chapter highlighted the potential for advancing finger vein identification through the fusion of CNNs and transformers.

6.2 Contributions

The contributions of this thesis to the field of biometric security systems are manifold. Through the exploration of finger vein identification, we have advanced the understanding and application of this modality as a secure and efficient unimodal biometric recognition method. The specific contributions include:

- InceptionResnet-V2 Model: Chapter 3 introduced a novel deep learning model, based on the InceptionResnet-V2 architecture. The proposed model demonstrated superior performance in finger vein identification, surpassing existing state-of-the-art methods. This contribution underlines the potential of transfer learning in enhancing the robustness and accuracy of biometric recognition systems.

- Hybrid Convolutional Transformer Model: Chapter 4 introduced the Finger Vein Convolution-Transformer (FVCT) model, a hybrid approach that merges the strengths of convolutional neural networks (CNNs) and transformers. The FVCT model set a new benchmark in finger vein identification, outperforming existing transformer and hybrid models. This contribution underscores the potential for fusing different deep learning paradigms to advance biometric security.

- Comprehensive Exploration: Throughout this thesis, we conducted a multifaceted exploration of finger vein identification, encompassing historical development, feature extraction techniques, and the application of deep learning models. This comprehensive exploration provides a solid foundation for future research and development in the field of biometric security systems.

6.3 Future Work

The research conducted in this thesis opens the door to several avenues for future work and innovation in the field of biometric security systems. Some key recommendations for future research include:

- Advanced Deep Learning Models: The exploration of advanced deep learning models, such as CNN, Capsule Networks (Caps-nets), and Vision Transformers, holds promise for further enhancing the recognition performance of finger vein identification systems. Future research should focus on developing and optimizing these models for improved accuracy and efficiency.

- Reducing Computational Cost: While the FVCT model demonstrated exceptional performance, future work should aim to reduce the model's parameter count and computational cost. This will make the model more practical for real-world applications, including those with limited computational resources.

- Enhancing Local Information Extraction: Improving the Transformer model's capability to extract local information is essential for enhancing the model's performance, especially in scenarios with complex finger vein patterns. Future research should focus on techniques to enhance local feature extraction within the transformer architecture.

- Diverse Datasets: Expanding the scope of research to include diverse and representative datasets is crucial for the development of robust and reliable finger vein identification systems. Future work should involve the creation and utilization of datasets that encompass a wide range of finger vein patterns.

In conclusion, this thesis has explored the field of biometric security systems, with a specific focus on finger vein identification. The research has provided a comprehensive understanding of this innovative biometric technology and its critical role in enhancing security and authentication systems. The contributions made in the development of advanced deep learning models, such as InceptionResnet-v2-Based and FVCT models, showcase the potential for further advancements in the field.

The findings presented in this thesis underscore the significance of finger vein identification as a secure and efficient unimodal biometric recognition modality. The fusion of different deep learning paradigms and the exploration of advanced models have paved the way for future innovations in biometric security.

As the field of biometric security systems continues to evolve, the research conducted in this thesis contributes to ongoing efforts to enhance the security, accuracy, and convenience of personal identification. With a commitment to advancing biometric security, this thesis provides a strong foundation for future research and development, ultimately leading to more secure and efficient biometric recognition systems across diverse applications.

Bibliography

[1] Anil K. Jain, Arun A. Ross, and Karthik Nandakumar. *Introduction to Biometrics*. Springer US, 2011.

[2] Anil K Jain, Arun Ross, and Salil Prabhakar. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1):4–20, 2004.

[3] Uday Bhanu Ghosh, Rohan Sharma, and Abhishek Kesharwani. Symptoms-based biometric pattern detection and recognition. In *Augmented Intelligence in Healthcare: A Pragmatic and Integrated Analysis*, pages 371–399. Springer, 2022.

[4] Kejun Wang, Hui Ma, Oluwatoyin P Popoola, and Jingyu Li. Finger vein recognition. *Biometrics*, pages 31–53, 2011.

[5] Xinwei Qiu, Wenxiong Kang, Senping Tian, Wei Jia, and Zhixing Huang. Finger vein presentation attack detection using total variation decomposition. *IEEE Transactions on Information Forensics and Security*, 13(2):465–477, 2017.

[6] Changxing Ding and Dacheng Tao. Pose-invariant face recognition with homography-based normalization. *Pattern Recognition*, 66:144–152, 2017.

[7] Maneet Singh, Richa Singh, and Arun Ross. A comprehensive overview of biometric fusion. *Information Fusion*, 52:187–205, 2019.

[8] Chuck Wilson. *Vein pattern recognition: a privacy-enhancing biometric*. CRC press, 2010.

[9] Shaveta Dargan and Munish Kumar. A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities. *Expert Systems with Applications*, 143:113114, 2020.

[10] Bir Bhanu and Venu Govindaraju. *Multibiometrics for human identification*. Cambridge University Press, 2011.

[11] Zahid Akhtar, Abdenour Hadid, Mark S Nixon, Massimo Tistarelli, Jean-Luc Dugelay, and Sebastien Marcel. Biometrics: In search of identity and security (q & a). *IEEE MultiMedia*, 25(3):22–35, 2018.

[12] Luca Ghiani, Abdenour Hadid, Gian Luca Marcialis, and Fabio Roli. Fingerprint liveness detection using local texture features. *IET Biometrics*, 6(3):224–231, 2017.

[13] Muhammad Sharif, Mudassar Raza, Jamal Hussain Shah, Mussarat Yasmin, and Steven Lawrence Fernandes. An overview of biometrics methods. In *Handbook of Multimedia Information Security: Techniques and Applications*, pages 15–35. Springer International Publishing, 2019.

[14] Abdenour Hadid, Nicholas Evans, Sebastien Marcel, and Julian Fierrez. Biometrics systems under spoofing attack: an evaluation methodology and lessons learned. *IEEE Signal Processing Magazine*, 32(5):20–30, 2015.

[15] Dakshina Ranjan Kisku, Phalguni Gupta, and Jamuna Kanta Sing. *Advances in biometrics for secure human authentication and recognition*. CRC Press, 2013.

[16] Yang Liu and Elizabeth Shriberg. Comparing evaluation metrics for sentence

boundary detection. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, volume 4, pages IV–185. IEEE, 2007.

[17] Zhi Liu, Yilong Yin, Hongjun Wang, Shangling Song, and Qingli Li. Finger vein recognition with manifold learning. *Journal of Network and Computer Applications*, 33(3):275–282, 2010.

[18] Jian-Da Wu and Chiung-Tsiung Liu. Finger-vein pattern identification using principal component analysis and the neural network technique. *Expert Systems with Applications*, 38(5):5423–5427, 2011.

[19] Jian-Da Wu and Chiung-Tsiung Liu. Finger-vein pattern identification using svm and neural network technique. *Expert Systems with Applications*, 38(11):14284–14289, 2011.

[20] Md Hossain, Jundong Chen, and Khandaker Rahman. On enhancing serial fusion based multi-biometric verification system. *Applied Intelligence*, 48:4824–4833, 2018.

[21] P Harsha and C Subashini. A real time embedded novel finger-vein recognition system for authenticated on teller machine. In *2012 International Conference on Emerging Trends in Electrical Engineering and Energy Management (ICETEEEM)*, pages 271–275. IEEE, 2012.

[22] Souad Khellat-Kihel, Nuno Cardoso, João Monteiro, Mohamed Benyettou, et al. Finger vein recognition using gabor filter and support vector machine. In *International image processing, applications and systems conference*, pages 1–6. IEEE, 2014.

[23] Rig Das, Emanuela Piciucco, Emanuele Maiorana, and Patrizio Campisi. Convolutional neural network for finger-vein-based biometric identification. *IEEE Transactions on Information Forensics and Security*, 14(2):360–373, 2018.

[24] Dongdong Zhao, Hui Ma, Zedong Yang, Jianian Li, and Wenbo Tian. Finger vein recognition based on lightweight cnn combining center loss and dynamic regularization. *Infrared Physics & Technology*, 105:103221, 2020.

[25] Bakhtiar Affendi Rosdi, Nordiana Mukahar, and Ng Tze Han. Finger vein recognition using principle component analysis and adaptive k-nearest centroid neighbor classifier. *International Journal of Integrated Engineering*, 13(1):177–187, 2021.

[26] Shan Juan Xie, Yu Lu, Sook Yoon, Jucheng Yang, and Dong Sun Park. Intensity variation normalization for finger vein recognition using guided filter based singe scale retinex. *Sensors*, 15(7):17089–17105, 2015.

[27] Bin Qin, Jian-fei Pan, Guang-zhong Cao, and Ge-guo Du. The anti-spoofing study of vein identification system. In *2009 international conference on computational intelligence and security*, volume 2, pages 357–360. IEEE, 2009.

[28] Zhiwei Zhang, Dong Yi, Zhen Lei, and Stan Z Li. Face liveness detection by learning multispectral reflectance distributions. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pages 436–441. IEEE, 2011.

[29] David Mulyono and Horng Shi Jinn. A study of finger vein biometric for personal identification. In *2008 International Symposium on Biometrics and Security Technologies*, pages 1–8. IEEE, 2008.

[30] K Syazana-Itqan, AR Syafeeza, NM Saad, Norihan Abdul Hamid, and WHBM Saad. A review of finger-vein biometrics identification approaches. *Indian J. Sci. Technol*, 9(32):1–9, 2016.

[31] Junichi Hashimoto. Finger vein authentication technology and its future. In *2006 Symposium on VLSI Circuits, 2006. Digest of Technical Papers.*, pages 5–8. IEEE, 2006.

[32] Hemant Vallabh. *Authentication using finger-vein recognition*. University of Johannesburg (South Africa), 2013.

[33] Wei Pi, Jinook Shin, and Dongsun Park. An effective quality improvement approach for low quality finger vein image. In *2010 international conference on electronics and information engineering*, volume 1, pages V1–424. IEEE, 2010.

[34] Eui Chul Lee, Hyeon Chang Lee, and Kang Ryoung Park. Finger vein recognition using minutia-based alignment and local binary pattern-based feature extraction. *International Journal of Imaging Systems and Technology*, 19(3):179–186, 2009.

[35] Pedram Mohammadi, Abbas Ebrahimi-Moghadam, and Shahram Shirani. Subjective and objective quality assessment of image: A survey. *arXiv preprint arXiv:1406.7799*, 2014.

[36] Kuan-Quan Wang, Anne S Khisa, Xiang-Qian Wu, and Qiu-Shi Zhao. Finger vein recognition using lbp variance with global matching. In *2012 international conference on wavelet analysis and pattern recognition*, pages 196–201. IEEE, 2012.

[37] Bakhtiar Affendi Rosdi, Chai Wuh Shing, and Shahrel Azmin Suandi. Finger vein recognition using local line binary pattern. *Sensors*, 11(12):11357–11371, 2011.

[38] Jinfeng Yang and Yihua Shi. Finger–vein roi localization and vein ridge enhancement. *Pattern Recognition Letters*, 33(12):1569–1579, 2012.

[39] Ajay Kumar and Yingbo Zhou. Human identification using finger images. *IEEE Transactions on image processing*, 21(4):2228–2244, 2011.

[40] Gongping Yang, Rongyang Xiao, Yilong Yin, and Lu Yang. Finger vein recognition based on personalized weight maps. *Sensors*, 13(9):12093–12112, 2013.

[41] Suhaimi Yusoff, Abdul Rahman Ramli, Shaiful Jahari Hashim, and Fakhrul Zaman Rokhani. Review on vein enhancement methods for biometric system. *Int. J. Res. Eng. Technol*, 4(04):833–841, 2015.

[42] R Arun, Madhu S Nair, R Vrinthavani, and Rao Tatavarti. An alpha rooting based hybrid technique for image enhancement. *image*, 9(10):1–10, 2011.

[43] Sos S Agaian, Blair Silver, and Karen A Panetta. Transform coefficient histogram-based image enhancement algorithms using contrast entropy. *IEEE transactions on image processing*, 16(3):741–758, 2007.

[44] Manpreet Kaur, Geetanjali Babbar, and CEC Landran. Finger vein detection using repeated line tracking, even gabor and multilinear discriminant analysis (mda). *vol*, 6:3280–3284, 2015.

[45] Munalih Ahmad Syarif, Thian Song Ong, Andrew BJ Teoh, and Connie Tee. Enhanced maximum curvature descriptors for finger vein verification. *Multimedia Tools and Applications*, 76:6859–6887, 2017.

[46] Hoshang Kolivand, Shiva Asadianfam, Kayode Akinlekan Akintoye, and Mohd Shafry Rahim. Finger vein recognition techniques: a comprehensive review. *Multimedia Tools and Applications*, pages 1–35, 2023.

[47] Lu Yang, Gongping Yang, Yilong Yin, and Xiaoming Xi. Exploring soft biometric trait with finger vein recognition. *Neurocomputing*, 135:218–228, 2014.

[48] T Liu, JB Xie, W Yan, PQ Li, and HZ Lu. An algorithm for finger-vein segmentation based on modified repeated line tracking. *The Imaging Science Journal*, 61(6):491–502, 2013.

[49] Dengsheng Zhang and Guojun Lu. Review of shape representation and description techniques. *Pattern recognition*, 37(1):1–19, 2004.

[50] Charles T Zahn and Ralph Z Roskies. Fourier descriptors for plane closed curves. *IEEE Transactions on computers*, 100(3):269–281, 1972.

[51] Yudong Zhang, Shuihua Wang, Ping Sun, and Preetha Phillips. Pathological brain detection based on wavelet entropy and hu moment invariants. *Bio-medical materials and engineering*, 26(s1):S1283–S1290, 2015.

[52] Kang Ryoung Park. Finger vein recognition by combining global and local features based on svm. *Computing and Informatics*, 30(2):295–309, 2011.

[53] R Gayathri and P Ramamoorthy. Multifeature palmprint recognition using feature level fusion. *International Journal of Engineering Research and Application*, 2(2):1048–1054, 2012.

[54] Kashif Shaheed, Aihua Mao, Imran Qureshi, Munish Kumar, Sumaira Hussain, Inam Ullah, and Xingming Zhang. Ds-cnn: A pre-trained xception model based on depth-wise separable convolutional neural network for finger vein recognition. *Expert Systems with Applications*, 191:116288, 2022.

[55] Qais Al-Nuzaili, Siti Z Mohd Hashim, Faisal Saeed, Mohammed Sayim Khalil, and Dzulkifli Bin Mohamad. Enhanced structural perceptual feature extraction model for arabic literal amount recognition. *International Journal of Intelligent Systems Technologies and Applications*, 15(3):240–254, 2016.

[56] Daniel Hartung, Martin Aastrup Olsen, Haiyun Xu, and Christoph Busch. Spectral minutiae for vein pattern recognition. In *2011 international joint conference on biometrics (IJCB)*, pages 1–7. IEEE, 2011.

[57] Byung Jun Kang and Kang Ryoung Park. Multimodal biometric authentication based on the fusion of finger vein and finger geometry. *Optical Engineering*, 48(9):090501–090501, 2009.

[58] Jianping Gou, Lan Du, Yuhong Zhang, Taisong Xiong, et al. A new distance-weighted k-nearest neighbor classifier. *J. Inf. Comput. Sci*, 9(6):1429–1436, 2012.

[59] Miyuki Kono. A new method for the identification of individuals by using of vein pattern matching of a finger. In *Proc. Fifth Symposium on Pattern Measurement, Yamaguchi, Japan, 2000*, pages 9–12, 2000.

[60] Matthias Vanoni, Pedro Tome, Laurent El Shafey, and Sébastien Marcel. Cross-database evaluation using an open finger vein sensor. In *2014 IEEE workshop on biometric measurements and systems for security and medical applications (BIOMS) proceedings*, pages 30–35. IEEE, 2014.

[61] Yu Lu, Shan Juan Xie, Sook Yoon, and Dong Sun Park. Finger vein identification using polydirectional local line binary pattern. In *2013 International Conference on ICT Convergence (ICTC)*. IEEE, October 2013.

[62] Mitsutoshi Himaga and Katsuhiro Kou. Finger vein authentication technology and financial applications. In *Advances in Biometrics: Sensors, Algorithms and Systems*, pages 89–105. Springer, 2008.

[63] Takashi Yanagawa, Satoshi Aoki, and Tetsuji Ohyama. Human finger vein images are diverse and its patterns are useful for personal identification. 2007.

[64] Beining Huang, Yanggang Dai, Rongfeng Li, Darun Tang, and Wenxin Li. Finger-vein authentication based on wide line detector and pattern normalization. In *2010 20th international conference on pattern recognition*, pages 1269–1272. IEEE, 2010.

[65] Daniel Hartung. Vascular pattern recognition: And its application in privacy-preserving biometric online-banking systems. 2012.

[66] Yilong Yin, Lili Liu, and Xiwei Sun. Sdumla-hmt: a multimodal biometric database. In *Biometric Recognition: 6th Chinese Conference, CCBR 2011, Beijing, China, December 3-4, 2011. Proceedings 6*, pages 260–268. Springer, 2011.

[67] Bram T Ton and Raymond NJ Veldhuis. A high quality finger vascular pattern dataset collected using a custom designed capturing device. In *2013 International conference on biometrics (ICB)*, pages 1–5. IEEE, 2013.

[68] Yu Lu, Shan Juan Xie, Sook Yoon, Zhihui Wang, and Dong Sun Park. An available database for the research of finger vein recognition. In *2013 6th International congress on image and signal processing (CISP)*, volume 1, pages 410–415. IEEE, 2013.

[69] Congcong Zhang, Xiaomei Li, Zhi Liu, Qijun Zhao, Hui Xu, and Fangqi Su. The cfvd reflection-type finger-vein image database with evaluation baseline. In *Biometric Recognition: 8th Chinese Conference, CCBR 2013, Jinan, China, November 16-17, 2013. Proceedings*, pages 282–287. Springer, 2013.

[70] Xiaoming Xi, Gongping Yang, Yilong Yin, and Xianjing Meng. Finger vein recognition with personalized feature selection. *Sensors*, 13(9):11243–11259, 2013.

[71] Mohd Shahrimie Mohd Asaari, Shahrel A Suandi, and Bakhtiar Affendi Rosdi. Fusion of band limited phase only correlation and width centroid contour distance for finger based biometrics. *Expert Systems with Applications*, 41(7):3367–3382, 2014.

[72] Hyung Gil Hong, Min Beom Lee, and Kang Ryoung Park. Convolutional neural network-based finger-vein recognition using nir image sensors. *Sensors*, 17(6):1297, 2017.

[73] Jong Min Song, Wan Kim, and Kang Ryoung Park. Finger-vein recognition based on deep densenet using composite image. *Ieee Access*, 7:66845–66863, 2019.

[74] Changlong He, Zuojin Li, Liukui Chen, and Jun Peng. Identification of finger vein using neural network recognition research based on pca. In *2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*, pages 456–460. IEEE, 2017.

[75] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. Inception-v4, inception-resnet and the impact of residual connections on learn-

ing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

[76] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2016.

[77] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[78] Hui Hu, Wenxiong Kang, Yuting Lu, Yuxun Fang, Hongda Liu, Junhong Zhao, and Feiqi Deng. Fv-net: learning a finger-vein feature representation based on a cnn. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 3489–3494. IEEE, 2018.

[79] Zhiyong Tao, Haotong Wang, Yalei Hu, Yueming Han, Sen Lin, and Ying Liu. Dglfv: Deep generalized label algorithm for finger-vein recognition. *IEEE Access*, 9:78594–78606, 2021.

[80] Sif Eddine Boudjellal, Naceur-Eddine Boukezzoula, and Abdelwahhab Boudjelal. Deep learning model based on inceptionresnet-v2 for finger vein recognition. In *2022 International Conference of Advanced Technology in Electronic and Electrical Engineering (ICATEEE)*, pages 1–5. IEEE, 2022.

[81] Jiazhen Liu, Ziyan Chen, Kaiyang Zhao, Minjie Wang, Zhen Hu, Xinwei Wei, Yicheng Zhu, Yuncong Yu, Zhe Feng, Hakil Kim, et al. Finger vein recognition using a shallow convolutional neural network. In *Biometric Recognition: 15th Chinese Conference, CCBR 2021, Shanghai, China, September 10–12, 2021, Proceedings 15*, pages 195–202. Springer, 2021.

[82] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[83] Xiaoxia Li, Di Huang, and Yunhong Wang. Comparative study of deep learning methods on dorsal hand vein recognition. In *Biometric Recognition: 11th Chinese Conference, CCBR 2016, Chengdu, China, October 14-16, 2016, Proceedings 11*, pages 296–306. Springer, 2016.

[84] Yu Lu, Shanjuan Xie, and Shiqian Wu. Exploring competitive features using deep convolutional neural network for finger vein recognition. *IEEE access*, 7:35113–35123, 2019.

[85] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[86] Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. Transformers in vision: A survey. *ACM computing surveys (CSUR)*, 54(10s):1–41, 2022.

[87] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[88] Haoran Zhu, Boyuan Chen, and Carter Yang. Understanding why vit trains badly on small datasets: An intuitive perspective. *arXiv preprint arXiv:2302.03751*, 2023.

[89] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision*, pages 843–852, 2017.

[90] Zihang Dai, Hanxiao Liu, Quoc V Le, and Mingxing Tan. Coatnet: Marrying convolution and attention for all data sizes. *Advances in Neural Information Processing Systems*, 34:3965–3977, 2021.

[91] Tete Xiao, Mannat Singh, Eric Mintun, Trevor Darrell, Piotr Dollár, and Ross Girshick. Early convolutions help transformers see better. *Advances in Neural Information Processing Systems*, 34:30392–30400, 2021.

[92] Haiping Wu, Bin Xiao, Noel Codella, Mengchen Liu, Xiyang Dai, Lu Yuan, and Lei Zhang. Cvt: Introducing convolutions to vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 22–31, 2021.

[93] Weijian Xu, Yifan Xu, Tyler Chang, and Zhuowen Tu. Co-scale conv-attentional image transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9981–9990, 2021.

[94] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International conference on machine learning*, pages 10347–10357. PMLR, 2021.

[95] Hugo Touvron, Matthieu Cord, Alexandre Sablayrolles, Gabriel Synnaeve, and Hervé Jégou. Going deeper with image transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 32–42, 2021.

[96] Asher Trockman and J Zico Kolter. Patches are all you need? *arXiv preprint arXiv:2201.09792*, 2022.

[97] Naoto Miura, Akio Nagasaka, and Takafumi Miyatake. Feature extraction of finger-vein patterns based on repeated line tracking and its application to personal identification. *Machine vision and applications*, 15:194–203, 2004.

[98] Huafeng Qin, Lan Qin, and Chengbo Yu. Region growth–based feature extraction method for finger-vein recognition. *Optical Engineering*, 50(5):057208–057208, 2011.

[99] Naoto Miura, Akio Nagasaka, and Takafumi Miyatake. Extraction of finger-vein patterns using maximum curvature points in image profiles. *IEICE TRANSACTIONS on Information and Systems*, 90(8):1185–1194, 2007.

[100] Puneet Gupta and Phalguni Gupta. An accurate finger vein based verification system. *Digital Signal Processing*, 38:43–52, 2015.

[101] Hoang Thien Van, Thanh Tuan Thai, and Thai Hoang Le. Robust finger vein identification base on discriminant orientation feature. In *2015 Seventh International Conference on Knowledge and Systems Engineering (KSE)*, pages 348–353. IEEE, 2015.

[102] Su Tang, Shan Zhou, Wenxiong Kang, Qiuxia Wu, and Feiqi Deng. Finger vein verification using a siamese cnn. *IET biometrics*, 8(5):306–315, 2019.

[103] Wenming Yang, Changqing Hui, Zhiquan Chen, Jing-Hao Xue, and Qingmin Liao. Fv-gan: Finger vein representation using generative adversarial networks. *IEEE Transactions on Information Forensics and Security*, 14(9):2512–2524, 2019.

[104] Borui Hou and Ruqiang Yan. Arcvein-arccosine center loss for finger vein verification. *IEEE Transactions on Instrumentation and Measurement*, 70:1–11, 2021.

[105] Junduan Huang, Weijian Luo, Weili Yang, An Zheng, Fengzhao Lian, and Wenxiong Kang. Fvt: finger vein transformer for authentication. *IEEE Transactions on Instrumentation and Measurement*, 71:1–13, 2022.

[106] Yupeng Li, Huimin Lu, Yifan Wang, Ruoran Gao, and Chengcheng Zhao. Vitcap: A novel vision transformer-based capsule network model for finger vein recognition. *Applied Sciences*, 12(20):10364, 2022.

[107] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[108] Xiaoyi Dong, Jianmin Bao, Dongdong Chen, Weiming Zhang, Nenghai Yu, Lu Yuan, Dong Chen, and Baining Guo. Cswin transformer: A general vision transformer backbone with cross-shaped windows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12124–12134, 2022.

[109] Yongming Rao, Wenliang Zhao, Benlin Liu, Jiwen Lu, Jie Zhou, and Cho-Jui Hsieh. Dynamicvit: Efficient vision transformers with dynamic token sparsification. *Advances in neural information processing systems*, 34:13937–13949, 2021.

[110] Jianwei Yang, Chunyuan Li, Pengchuan Zhang, Xiyang Dai, Bin Xiao, Lu Yuan, and Jianfeng Gao. Focal self-attention for local-global interactions in vision transformers. *arXiv preprint arXiv:2107.00641*, 2021.

[111] Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 568–578, 2021.

[112] Li Yuan, Yunpeng Chen, Tao Wang, Weihao Yu, Yujun Shi, Zi-Hang Jiang, Francis EH Tay, Jiashi Feng, and Shuicheng Yan. Tokens-to-token vit: Training vision transformers from scratch on imagenet. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 558–567, 2021.

[113] Daquan Zhou, Bingyi Kang, Xiaojie Jin, Linjie Yang, Xiaochen Lian, Zihang Jiang, Qibin Hou, and Jiashi Feng. Deepvit: Towards deeper vision transformer. *arXiv preprint arXiv:2103.11886*, 2021.

[114] Jianyuan Guo, Kai Han, Han Wu, Yehui Tang, Xinghao Chen, Yunhe Wang, and Chang Xu. Cmt: Convolutional neural networks meet vision transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12175–12185, 2022.

[115] Sachin Mehta and Mohammad Rastegari. Mobilevit: light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv preprint arXiv:2110.02178*, 2021.

[116] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[117] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 630–645. Springer, 2016.

[118] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxvit: Multi-axis vision transformer. In *European conference on computer vision*, pages 459–479. Springer, 2022.

[119] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015.

[120] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016.

[121] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[122] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[123] Ross Wightman. Pytorch image models (2019). *DOI: https://doi.org/10.5281/zenodo*, 4414861.

[124] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[125] Sif Eddine Boudjellal, Naceur-Eddine Boukezzoula, and Abdelwahhab Boudjelal. Hybrid convolutional transformer-based network model for finger vein identification. *Journal of Electronic Imaging*, 53005(1), 2023.

APPENDIX A

Biometric modalities and their characteristics

Biometric Modality	**Advantages**	**Limitations**
Face	• physical contact is not required • Convenient, less complex statistics • Fast recognition process • 3D offers increased precision	• For twins, differences may not be clear • With age, facial traits may change • Potential privacy concerns • Lighting and variations in pose can reduce accuracy
Fingerprint	• Generally uses small, low-cost readers • Reliable and highly accurate • Fast matching process • An effective biometric for large-scale systems • Widely accepted forensic tool	• Not considered hygienic • Twists, cuts, or dirt may create obstacles
Iris	• High accuracy and more protective • High stability of characteristics over time • Moderate data storage requirements • Works well with either verification or identification applications	• Small sample size • Diseases may affect accuracy • Challenges at a large distance

Ear	• Identification process is fast • Most stable and less computational complexity • Less computational complexity	• Identification process is fast • Uncomfortable as it requires direct contact
Hand Geometry	• Operates well in challenging environments • Widely used • Less processing	• Not accurate for moderate to large populations • Unhygienic • Injuries and jewels may harm the results
Palmprint	• Large variety of features • High reliability and permanent • Good recognition even with low-resolution scanners	• Unhygienic • Injuries may create obstacles
Retina	• Among the most accurate of biometrics • Moderate storage requirements for templates	• Special hardware is required • Expensive
Vein Pattern	• Highly private • Very accurate • Difficult to circumvent • Near contactless, hygienic	• Not yet widely used • Can be impacted by bright ambient light
Voice	• Easy implementation • Less expensive • Convenient to employ • High public acceptance	• Throat disease can affect accuracy • Generally large storage requirements for templates • Not sufficiently distinctive for identification over large databases
Keystroke Dynamics	• Easy implementation and use • Additional hardware is not required for keyboarding	• Only useful for certain applications

Gait	• Easy to capture the image • Convenient to use • No distance problem	• Computationally expensive • Lack of accuracy
Signature	• More accuracy • Less false acceptance rate • Low storage requirement	• Can be forged • Changes based on the emotional and medical condition of the person
DNA	• Highly unique feature • High performance • Its universality is very high	• More storage required • Not an automatic technique • More informative, so privacy issues

Table A.1: Biometric modalities and their characteristics [1, 9, 8]